Weighting and Standard Error Estimation for ABS Household Surveys

Paper prepared for ABS Methodology Advisory Committee, July 2000

Philip Bell
Australian Bureau of Statistics, e-mail: philip.bell@abs.gov.au

1 Introduction

The Australian Bureau of Statistics (ABS) conducts a variety of household surveys to collect data about households or the persons in them. These surveys typically use a multistage sample design that first selects a set of geographic areas and then a sample of dwellings to be approached by interviewers. Such a design gives little control over the types of households and persons that are selected — so it is important to use estimation techniques that make some correction for any imbalance in the sample. This combination of a multistage design with a possibly complex estimation technique makes variance estimation a nontrivial problem.

Within the ABS, household survey estimation has been undergoing a period of transition. In the 1980s a technique known as post-stratified ratio estimation was typically used. For these estimates a valid variance estimator was available, known in the ABS as split-halves. Standard practice was to use split-halves estimates of standard error as the basis of the standard error models published with ABS data.

Over time the drive to do more with each survey has increased the complexity of sample design and required extensions to estimation. Now weighting may include a separate non-response adjustment phase, and may use a variety of auxiliary data — for example, population counts for households as well as persons. The enabling technology for some of these extensions was the CALMAR macro of Deville, Särndal and Sautory (1993), which allows weighting using the generalised regression method and related calibration methods..

In the ABS these changes to estimation methods moved faster than the ability to estimate variances. Up until 1997 split-halves was the only tool used for variance estimation, and any bias in the resulting standard error models was not measured. In late 1997 this lack was identified as a key problem, and it was intensively researched over a two year period. As a result the ABS is moving to a new standard for weighting and for variance estimation.

This paper describes the methodological principles behind the generalised regression weighting approach used in the ABS for estimation in household surveys. It then presents a number of approaches to estimating variance for such estimates, and recommends the group jackknife approach for standard use in the ABS. These methods are available via SAS macros written within the ABS. They will be implemented as components of the Household Survey Facilities (HSF) processing system. It is also planned to make the group jackknife approach available through the SUPERCROSS tabulation and aggregation facility so that variance estimates are available for ad hoc requests.

Section 2 describes the development of weighting techniques and in particular the generalised regression weighting approach. Section 3 presents a variety of methods for estimating the variance of these estimates. Section 4 focuses specifically on the group jackknife approach to variance estimation, and presents a theoretical justification. Section 5 compares the variance estimators in the context of a simulated population with systematic sampling of clusters. Section 6 describes an actual survey, the Australian Labour Force Survey, and presents a comparison of different variance estimators. It also shows how the new methods apply to estimating variance for complex estimates such as the trend. Section 7 summarises results from a number of evaluations conducted on more complex surveys. These evaluations focus on the value of the estimates for deriving variance models. Section 8 concludes with some comments on the application of the techniques in generalised facilities in the ABS.

2 Weighting for household surveys

2.1 The household survey sample

For the purposes of selecting a household survey, the ABS maintains a sampling frame which covers all dwellings in Australia. Each state is divided into zones (two for each state and one for each territory), then into "frame strata" (based on dissemination region, area type and population density) and into finer geographic groupings e.g. collectors districts (CDs), blocks and clusters. The physical process of sampling involves ordering the CDs within state by frame stratum and systematically selecting CDs, taking account of their expected size. The selected CDs are then inspected to identify blocks and clusters, and finally a single cluster is selected. The selected CDs are used for a five year period in a variety of surveys, to save the expense of inspecting more CDs than is necessary. Different surveys, of course, select different clusters of dwellings.

This sampling process has a simple theoretical description. Except in a few "sparsely settled" strata, it is equivalent to listing all the clusters in a state in a set order (given by the sampling frame), choosing a random start and then systematically selecting every K-th cluster in the list. The value of K used for a state is known as the state skip. The sample usually consists of all the dwellings in the selected clusters, although smaller surveys may select a subset of dwellings from each cluster.

This sample design gives each in-scope dwelling in Australia a predefined probability of selection, normally constant in each state. Typically a survey will select all in-scope units in selected dwellings. In the rare case where a survey only selects some of the persons in a household, this must be done at random, and appropriate information collected to ensure that the respondent's probability of selection is known.

2.2 Linear weighting and selection weighted estimates

This paper looks at weighted estimates in which a weight w_i^A is associated with each unit i in the sample. The estimate of a population total Y based on reported values y_i is then obtained by weighted aggregation; that is,

$$\hat{y}^A = \Sigma_i w_i^A y_i \qquad (2.1)$$

A superscript (A in this case) will be used to identify different weights and the corresponding estimates. Unless otherwise stated, the summation in this paper is over all units in the sample.

Horvitz and Thompson (1952) proposed weighting each unit by the inverse of the selection probability. Let the selection probability of unit i be given by π_i. The inverse of this selection probability is known as the selection weight and denoted $w_i^\pi = 1/\pi_i$. Typically the selection weight equals the state skip. The resulting selection-weighted estimator

$$\hat{y}^\pi = \Sigma_i w_i^\pi y_i \qquad (2.2)$$

would be unbiased if there was no non-response to the survey.

2.3 Post-stratified ratio estimates

Suppose that the units in the sample can be classified to post-strata p for which the population counts N_p are known. The post-stratified ratio estimator uses these population totals as auxiliary data to improve on an existing estimate. The new estimate is given by

$$\begin{aligned} \hat{y}^{\mathrm{P}} &= \Sigma_p(\Sigma_{i\in p}\, w_i^{\pi} y_i) N_p / \Sigma_{i\in p}\, w_i^{\pi} \\ &= \Sigma_i\, w_i^{\mathrm{P}} y_i \end{aligned} \qquad (2.3)$$

for $\quad w_i^{\mathrm{P}} = w_i^{\pi} N_p / \Sigma_{i\in p}\, w_i^{\pi} \qquad$ for $i \in p$. $\qquad (2.4)$

The new weights are a pro-rata adjustment of the selection weights w_i^{π} so that the total weight in each post-stratum matches the appropriate population total. This technique was described in Cochran (1963).

The post-stratified ratio estimate $\hat{y}^{\mathrm{P}}$ will have lower sampling error than $\hat{y}^{\pi}$ if the y_i values are sufficiently different between post-strata. In the presence of non-response at rates that differ between post-strata, $\hat{y}^{\mathrm{P}}$ should also have lower bias.

2.4 Multi-step weighting

A post-stratified ratio estimation step can be applied to a set of input weights other than the selection weights, by substituting those weights for the selection weights w_i^{π} in (2.4). This opens the possibility of successive steps of weighting, with the output weights from one step being used as input weights to the next.

An example could be where the first step performs a non-response adjustment, with the post-stratum categories chosen for similar likelihood of response. The population counts for these non-response categories could even be estimates rather than known totals. A second step could then adjust the weights to demographic benchmarks.

Another example arises in a two-phase survey, where certain questions are only asked of a random subset of the sample. It seems logical to weight the full sample, and then to use these weights as input weights for weighting the subsample. This may involve a non-response adjustment step and a final adjustment to demographic benchmarks.

2.5 Generalised regression estimates

Post-stratified ratio weighting could be used to adjust successively to two or more sets of benchmarks. While this has some benefits, the resulting weights will sum to the last set of benchmarks only. The generalised regression method allows adjusting weights to add to multiple sets of benchmarks in a single step.

This approach was discussed in Bethlehem and Keller (1987), and is a particular case of a more general class of calibration estimates introduced by Deville and Särndal (1992). The methods were introduced to the ABS using the SAS macro CALMAR of Deville, Särndal and Sautory (1993). More recently the ABS has produced an internally written program GREGWT which performs similar computations but much more quickly. The method uses the generalised regression estimator and variants; the calculations are described in Singh and Mohl (1996).

Suppose that x_i is a row vector of auxiliary variables, and X is a corresponding vector of benchmark values. Let w_i^A be input weights which give estimates $\hat{y}^{\mathrm{A}}$ and $\hat{x}^{\mathrm{A}} = \Sigma_i\, w_i^{\mathrm{A}} x_i$. The generalised regression, GREG, or GR estimator is given by:

$$\hat{y}^{\mathrm{GR}} = \hat{y}^{\mathrm{A}} + (X - x^{\mathrm{A}})\hat{\beta}^{\mathrm{GR}} \qquad \text{, where} \qquad (2.5)$$

$$\hat{\beta}^{\mathrm{GR}} = (\Sigma_i (w_i^{\mathrm{A}}/c_i) x_i' x_i)^{-1}\, \Sigma_i (w_i^{\mathrm{A}}/c_i) x_i' y_i \qquad (2.6)$$

Here $\hat{\beta}^{GR}$ is a sample estimate of the parameter describing the regression of y_i against x_i in the whole population. Typically $c_i = 1$ is used; larger values for c_i increase the penalty for changing the weight of unit i. In what follows c_i will often be dropped to save space.

The GR estimator can be written in the weighted form

$$\hat{y}^{GR} = \Sigma_i w_i^{GR} y_i \quad \text{, for weights given by} \tag{2.7}$$

$$w_i^{GR} = w_i^A g_i \quad \text{, for} \tag{2.8}$$

$$g_i = 1 + (X - x^A)(\Sigma_i (w_i^A / c_i) x_i' x_i)^{-1} x_i' / c_i \tag{2.9}$$

Note that the weight adjustments g_i do not depend on the variable being estimated. Thus the same weights are used for any item we estimate. Note also that $\Sigma_i (w_i^A / c_i) x_i' x_i$ may be a singular matrix - in this case, we can use any generalised inverse of this matrix in place of the matrix inverse at (2.6) or (2.9).

2.6 Examples of generalised regression

The post-stratified ratio estimator

In the post-stratified case the benchmarks $X = (N_1, ..., N_P)$ are population counts for P non-overlapping post-strata. Thus each unit contributes to a single post-stratum, and the auxiliary variable is an indicator vector $x_i = (0, 0, ..., 0, 1, 0, ..., 0)$, where the 1 indicates the unit's post-stratum. Writing $\hat{n}_p^A = \Sigma_{i \in p} w_i^A$ for the total weight from units in post-stratum p, the estimator simplifies to

$$\begin{aligned} \hat{y}^{GR} &= \hat{y}^A + \Sigma_p (N_p - \hat{n}_p^A)(\hat{n}_p^A)^{-1} (\Sigma_{i \in p} w_i^A y_i) \\ &= \Sigma_p \frac{N_p}{\hat{n}_p^A} (\Sigma_{i \in p} w_i^A y_i) \end{aligned}$$

which is just the post-stratified ratio estimator given at (2.3).

Calibrating to two sets of benchmarks

Given two sets of benchmarks X^A and X^B corresponding to different classifications, we set up x_i^A and x_i^B to indicate the class unit i belongs to under each classification. Then we can use $X = (X^A, X^B)$ and $x_i = (x_i^A, x_i^B)$ in the generalised regression formulae. The formulae do not simplify as in the first example, since the matrix to be inverted is no longer diagonal. The resulting weights will add as required to both the sets of benchmarks.

Integrated weighting

It is also possible for the auxiliary variables x_i to be vectors of numbers rather than of indicator (zero-one) variables. A common situation here is when the units i correspond to households, but weighting is to be performed so that the estimated numbers of persons in various categories match known demographic benchmarks. Often we also impose benchmark constraints on numbers of households in some other categories. This is an example of what we call integrated weighting (because it integrates the household weighting with person counts). The method was introduced by Lemaître & Dufour (1986).

In integrated weighting, $x_{i,k}$ (the k-th element of the vector x_i) represents the contribution of household i to the k-th benchmark $X_{i,k}$. If $X_{i,k}$ is a person count then $x_{i,k}$ gives the number of that type of person in household i. If $X_{i,k}$ is a household count then $x_{i,k}$ will be a zero-one indicator variable.

2.7 Generalised regression as a constrained minimisation

The GR estimator can be derived as follows. We want to adjust the input weights w_i^A to get new weights $w_i^{GR} = g_i w_i^A$ which meet our *benchmark constraints*

$$\Sigma_i w_i^{GR} x_i = X \tag{2.10}$$

In choosing the new weights, we want to make them as close to the initial weights as possible. The GR estimator is the set of new weights that meet the constraints (2.10) while minimising the generalised least-squares distance function given by

$$F^{GLS} = \Sigma_i \frac{c_i (w_i - w_i^A)^2}{w_i^A} = \Sigma_i c_i w_i^A (g_i - 1)^2 \tag{2.11}$$

2.8 Restricting the weights or changing distance function

Looking at the estimator in this way opens up a range of alternative estimators. For instance, we could put *range restrictions* on the weights:

$$L_i \le w_i \le U_i \tag{2.12}$$

Alternatively we could use a different distance function. To penalise large proportional changes in the weights (rather than simple difference) we can minimise the exponential distance function given by

$$F^{EXP} = \Sigma_i c_i [w_i \log(w_i / w_i^A) - w_i + w_i^A] \tag{2.13}$$

This distance function ensures that the weights are non-negative, but they could be very high, which is not desirable. This distance function gives the same weights as a process called iterative proportional fitting.

To compute the new weights under the extra constraints (2.12) or the exponential distance function (2.13) requires an iterative algorithm. The algorithms implemented by the macro GREGWT are described in an Appendix. They implement methods 5 and 6 of Singh and Mohl (1996). These are methods that strictly meet any range restrictions like those in (2.12), even if by doing so they fail to add to one or more of the benchmark constraints (2.11).

It is possible to aim for some other compromise between range restrictions and benchmark constraints, as in Rao and Singh (1997). We preferred to go with the most computationally practical methods. A failure to meet the benchmarks is likely to signify a poor weighting approach, and it seems best to flag this for human intervention rather than make an automatic compromise.

3 Variance estimation methods

3.1 Why variance estimation is a difficult problem

Two aspects of ABS household surveys make variance estimation a difficult problem. First, they use a multistage sample design, resulting in a sample in which units are clustered geographically. It is therefore important to use methods that account for this clustering. Variance estimators suitable for a simple random sample of units are not appropriate for this clustered data. Furthermore, the clusters themselves are not selected by simple random sampling but by a systematic sampling approach.

Second, the estimates are fairly complex, being based on one or more stages of weighting using the techniques described in section 2. Some variance estimation methods are appropriate for a single stage of weighting but do not measure the effect of multiple stages of weighting.

This section will outline some of the variance estimators that are used in household surveys. Section 4 will then go on to describe in more detail the group jackknife variance estimator and how it applies in ABS surveys.

3.2 The weighted residual variance estimator

Suppose that the first stage of selection is performed within strata. The unit at this first stage of selection is referred to as the "variance group" - we could define it to be a grouping of clusters rather than a single cluster if required. The weighted residual variance estimator computes variance at stratum level by comparing "weighted residuals" computed from the different variance groups. The method was discussed by Särndal, Swensson and Wretman (1989).

Let stratum h contain G_h variance groups. Suppose unit i has initial weight w_i^{A} and final weight w_i^{GR} after a single stage of generalised regression weighting. The weighted residuals variance estimator for $\hat{y}^{\mathrm{GR}} = \Sigma_i w_i^{\mathrm{GR}} y_i$ is given by

$$\mathrm{var}(\hat{y}^{\mathrm{GR}}) = \sum_h \frac{G_h}{G_h - 1} \Sigma_{g \in h} (\hat{e}_{hg} - \hat{e}_h)^2 \quad \text{, where} \tag{3.1}$$

$$\hat{e}_{hg} = \sum_{i \in hg} w_i^{\mathrm{GR}} (y_i - x_i \hat{\beta}^{\mathrm{GR}}) \quad \text{, and} \tag{3.2}$$

$$\hat{e}_h = \frac{1}{G_h} \sum_{g \in h} \hat{e}_{hg} \tag{3.3}$$

Here $\hat{\beta}^{\mathrm{GR}}$ is the regression parameter used in the generalised regression method. The values $\hat{e}_{hg}$ are known as the weighted residuals - they are a weighted aggregation of the difference between a unit's observed value and its prediction using the regression parameters.

This approach is presented in Stukel, Hidiroglou and Särndal (1996), who quote its use in the SUPERCARP package (Hidiroglou, Fuller and Hickman, 1980). Yung and Rao (1996) obtain (3.1) as an approximation to the Jackknife variance estimator; they call it the " Jackknife Linearisation variance estimator". In the ABS it has been used since at least the early 1980s for the case of a post-stratified ratio estimator. This ABS application has been called split-halves because is uses two variance groups in each stratum.

3.3 Problems with using the weighted residuals method

Single stage of weighting

The method assumes a single step of weighting - it cannot fully reflect the effects of multiple steps of weighting. It is possible to calculate a variance estimate using (3.1) to (3.3) based on the last step of the weighting, but such a variance estimate may be subject to bias. This use of the

method may not reflect the variation in previous steps of weighting, as only the residuals from the last step are calculated.

Biased for fine calibration

If the calibration being performed is very fine, some of the regression parameters may be very strongly influenced by a few individuals. In such a case, these individuals will have small residuals that do not fully reflect the instability in the estimation of the regression parameters. As a result, the weighted residual variance estimator will tend to underestimate variance when the calibration is very fine. It is good practice to avoid such fine calibrations in weighting.

Applies directly only to GR estimates

The method applies directly only to linear combinations of GR estimates. To apply to the ratio of two estimates requires an approximation based on linearising the ratio using a Taylor Series expansion. To apply to one of the variants of the GR estimate (such as a calibration estimate using the exponential distance function), Stukel, Hidiroglou and Särndal (1996) suggest it is sufficient to use formula (3.1) above but using the GR weights. Our implementation uses the final weights in formula (3.1) but computes residuals at (3.2) based on the GR regression parameters (which are available at the first iteration of the weighting algorithm). This appears to provide estimates of satisfactory reliability.

Computations require detailed data from the weighting process

For the post-stratified ratio estimator the computations are straightforward. For more general GR estimators the regression parameters $\hat{\beta}^{GR}$ need to be computed for each estimate of interest. The macro CALMAR does not output the required information for these calculations. The newer SAS macro GREGWT provides the facility to request tables as an output of the weighting process, and for these tables to include weighted residuals standard errors. This proved more convenient than attempting to store the required information for later variance calculations.

A separate issue is that information on cluster membership is needed in order to perform the weighted residuals variance calculations. This information is considered too sensitive for release on confidentialised unit record files. So external users of the survey data cannot perform weighted residuals variance calculations to go with their analyses.

Assumes stratified random sampling of variance groups

In the ABS the selection of clusters is not done by simple random sampling within strata, but systematically in each state. So in the ABS the "strata" on the sample frame are not actually treated as strata in sampling - they are used as an ordering variable for a systematic selection of clusters within each state.

Historically, the sample frame "strata" have been treated as strata for variance estimation. This gives a much more stable variance estimator, but produces a variance conditional on the number of clusters actually selected in each frame stratum - this could be biased downwards for the unconditional variance, but any bias seems to be small in practice.

The systematic sampling leads to difficulties even assuming stratified sampling. The usual form for the weighted residuals estimator treats each cluster as a separate variance group - this is the form this paper will refer to as weighted residuals (although in the ABS it has been termed the ultimate cluster estimator). But these variance groups do not reflect the systematic sampling of units.

Split-halves - dealing with stratified systematic sampling

To get around this we can form variance groups systematically and treat them as the first stage units. The variance groups will be more like a simple random sample than the clusters themselves. In the ABS this approach has been applied by dividing the clusters systematically into two variance groups - this is the technique called split-halves. In split halves, the formulae (3.1) to (3.3) are used, but the number of variance groups G_b is constant for each stratum at two. The formula (3.1) in this case reduces to

$$\text{var}^{\text{SH}}(\hat{y}^{\text{GR}}) = \sum_b (\hat{e}_{b1} - \hat{e}_{b2})^2 \qquad (3.4)$$

The use of two groups can lead to fairly unstable variance estimates, even after aggregating across strata. This may not be of as much concern as a marked bias, particularly as an input to variance modelling (discussed in 7.1).

Using the split-halves approach was considered worthwhile in the ABS to ensure that the standard error estimates included any benefit of our ordering of clusters within strata. Previous measures of this benefit have suggested a five percent decrease in standard errors compared to using a random ordering within strata. The study presented in section 6 suggests that the bias from using the weighted residual form rather than split-halves is small for key Labour Force Survey estimates - perhaps two percent. This seems inconsistent with there being a marked benefit from the ordering of clusters within strata.

3.4 An approximate split-halves variance estimator

Given the computational difficulties of the weighted residuals estimator (and recognising that it is only an approximate variance estimator anyhow) an approximation may be appropriate. In the ABS, a SAS macro has been available since 1985 for split-halves variance estimation for tables of estimates assuming a post-stratified ratio estimator. Until the GREGWT macro became available, this macro was used for approximate variance calculation in situations involving calibrated weights.

The approximation requires assigning each unit to a post-stratum, choosing a post-stratification which approximates the effect of the weighting process without allowing any post-stratum to contain only a few individuals. For example, calibration to two sets of benchmarks as described in section 2.7 could be approximated by a post-stratification to only one of the two sets of benchmarks.

Residuals are now calculated as a difference between a unit's response and the mean response for its post-stratum. Thus, if the mean for post-stratum p is given by

$$\bar{y}_p^{\text{A}} = \frac{\sum_i w_i^{\text{A}} y_i}{\sum_i w_i^{\text{A}}} \qquad (3.5)$$

we replace the weighted residual calculation at (3.2) by the following:

$$\hat{e}_{bg}^{\text{POST}} = \sum_p \sum_{i \in bgp} w_i^{\text{GR}} (y_i - \bar{y}_p^{\text{A}}) \qquad (3.6)$$

This method seems to give similar variance estimates to the full split-halves approach in many practical situations, provided that a reasonable choice of benchmarks is made. Some comparisons will be provided in section 7.

It may be useful to use this approximate approach when it is difficult or inconvenient to reproduce the weighting process when calculating variances. A weighted residuals version of the approximation could be used rather than split-halves if desired.

3.5 Other widely-used variance estimators

A number of other variance estimators are in use in other statistical agencies, including balanced repeated replications, the drop-out-one jackknife, and methods using detailed knowledge of selection probabilities and joint probabilities. A good introduction to the range of methods available is given in Wolter (1985).

These variance estimators typically assume a probability sample of a fixed number of variance groups from each stratum at the first stage of selection. In this they are similar to the weighted residuals estimator examined above. The variance calculation formulae resemble (3.1) in that variance calculations are performed for each stratum and then summed across strata.

These approaches are not a direct match to the ABS sample design, where state plays the role of stratum. The "frame stratum" on the ABS sampling frame is not used as a stratum, but to order the clusters for systematic selection. This systematic selection is therefore a critical part of the ABS sampling scheme, and cannot be well approximated as a probability sample of clusters.

To apply these methods in the ABS setting, one reasonable approach is to treat the sample as though the frame strata were actual strata, and to condition on the number of clusters selected. This is the approach that the ABS applies in its use of weighted residual method. Within these frame strata it may be reasonable to treat the clusters as a probability sample.

An alternative approach is to form variance groups systematically, in such a way that it is reasonable to treat the sample of variance groups as a probability sample. This is done within strata in the split-halves approach described above (using two variance groups within each stratum). It is also the basis for the *group jackknife* approach.

The group jackknife corresponds to a standard drop-out-one jackknife variance estimator, but it assumes no stratification. It is applied to variance groups formed in a systematic fashion across strata. This turns out to give a good variance estimator with a number of practical advantages. The group jackknife and its theoretical justification will be described in section 4.

4 Group jackknife variance estimation

4.1 Household survey sample selection

A household survey sample can be viewed as a systematic sample of clusters within each state, taken with probability proportional to size (PPS). (Various non-standard practices can be accommodated within this view by appropriate definition of cluster and by choice of the cluster size measure).

Within each state h the clusters are numbered purposively $I = 1, ..., N_h$ (in the current design they are sorted by stratum and then serpentine order). A sample is taken by choosing for each state a skip K_h and a random start $k_h \in (0, K_h]$. If the cumulative size measure is C_{hI} (with $C_{h0} = 0$), the ith selected cluster in state h is that cluster I for which $C_{h,I-1} < k_h + (i-1)K_h \leq C_{h,I}$. This gives for each state h a random number n_h of selected clusters, with total number of clusters $n = \Sigma_h n_h$.

4.2 Forming replicate groups

The group jackknife depends on dividing the selected clusters into G variance groups or *replicate groups*. This is done by going through the selected clusters ordered by state h and selection i and assigning clusters to replicate groups in cyclical order. This gives G systematic samples of clusters. The state skips in each replicate group are given by GK_h, but each replicate group has different start points taken from the set $k_h, k_h + K_h, \ldots, k_h + (G-1)K_h$ for each state h.

4.3 Justifying the group jackknife for systematic selection

Jackknife variance estimators are usually justified in a context of a stratified sample and assuming probability proportional to size (PPS) or simple random sampling (SRS) of clusters within strata. For the group jackknife this justification is given in Kott (1998). This paper will look at the ABS situation where the sample is obtained systematically within state.

In the ABS the sample of clusters within each state is systematic. It would be unreasonable to ignore this and treat the clusters as a PPS random sample selected within each state. We have ordered the clusters purposively, and selected systematically, in order to make the sample more representative. If this has been effective it would reduce the actual variance, but it would increase a variance estimate produced under the assumption of simple random sampling of clusters.

To make progress on measuring the variance of an estimate we proceed as follows. Any given replicate group that could have arisen by our selection process can be specified by a set of random starts $d = \{d_1, ..., d_8\}$ where $d_h \in (0, GK_h]$, being the random starts used for systematic sampling in the eight states. The selections in the replicate group are then given by the clusters I for which $C_{h,I-1} < d_h + iGK_h \leq C_{h,I}$ for some integer i.

Let D be the set of all possible replicate groups corresponding to all the random start sets d that could have arisen by our selection process. By construction, any element of D could have arisen as a replicate group. But there are elements of D that could not have arisen *together* in a set of G replicate groups. In fact, any set of G replicate groups that we select are obtained in a systematic manner from the set D.

To obtain a variance estimate, we have to assume that the G groups were generated at random (with replacement) from D. This assumption leads to variance estimates with a slight upward bias. First, because we assume with replacement selection, two random groups with starts selected from D may include the same clusters, whereas the actual selections cannot select a

cluster twice. This bias will be small if the proportion of clusters selected within any state is small. Second, there is a bias since the actual replicate groups were chosen systematically.

So we end up having to treat a systematic sample (of replicate groups) as a random sample in order to get a variance. How then are we better off than just treating the systematic sample of clusters as a random sample? In fact, we are much better off. Because each replicate group in D is itself a systematic selection of clusters, we can expect the major benefits of our systematic sampling of clusters to be represented in each replicate group. So the replicate groups are not nearly as different to each other as the clusters were, and any bias from treating them as a random sample is therefore much smaller. This argument applies as long as the replicates contain sufficiently many clusters i.e. provided that G is not too large.

4.4 Group jackknife variance estimator

Suppose that $\bar{y}_1, ..., \bar{y}_G$ are estimates of true mean $\bar{Y}$ from each of G replicate groups selected at random from the set D. Let $\bar{y} = \frac{1}{G}\Sigma_{g=1}^{G}\bar{y}_g$. Define $\bar{y}_{(g)}$ to be the *jackknife estimate* obtained by deleting replicate group g, given by

$$\bar{y}_{(g)} = \tfrac{1}{G-1}(G\bar{y} - \bar{y}_g) \tag{4.1}$$

Then $\mathrm{E_D}(\bar{y}) = E_D(\bar{y}_g) = E_D(\bar{y}_{(g)}) = \bar{Y}$, where the subscript D denotes expectation over the possible drawings of replicate groups from D. Then by simple calculation

$$\mathrm{E_D}(\tfrac{G-1}{G}\sum_{g=1}^{G}(\bar{y}_{(g)} - \bar{y})^2) = \mathrm{E_D}(\tfrac{1}{G(G-1)}\sum_{g=1}^{G}(\bar{y}_g - \bar{y})^2) = \mathrm{var_D}(\bar{y}) \tag{4.2}$$

The formula

$$\mathrm{v}^{GJ}(\bar{y}) = \tfrac{G-1}{G}\sum_{g=1}^{G}(\bar{y}_{(g)} - \bar{y})^2 \tag{4.3}$$

is called the *group jackknife variance estimator* for $\bar{y}$. Note that a stratified form of the jackknife variance estimator is not necessary. Thus we have shown:

Theorem 1: Group jackknife estimator of variance is unbiased for variance of simple mean

4.5 Group jackknife for more complex estimators

For the simple estimator of mean given above the group jackknife variance estimator is unbiased (given the assumption of randomly sampled replicate groups). The group jackknife variance estimator can be defined for a more general estimator as follows.

First, divide the sampled clusters into G replicate groups. Produce the overall estimate $\hat{y}$ using all the data. For each replicate group g produce the jackknife estimate $\hat{y}_{(g)}$ by applying exactly the same estimation process to the data excluding replicate group g. The group jackknife variance estimator is then given by

$$\mathrm{v}^{\mathrm{GJ}}(\hat{y}) = \tfrac{G-1}{G}\sum_{g=1}^{G}(\hat{y}_{(g)} - \hat{y})^2 \tag{4.4}$$

For complex estimators this may be biased. The bias for a generalised regression estimator is explored below.

4.6 Group jackknife for generalised regression estimators

The generalised regression estimator can be written in the form

$$\hat{y}^{\mathrm{GR}} = X(\Sigma_i a_i x_i' x_i)^{-1}\Sigma_i a_i x_i' y_i \tag{4.5}$$

where X is a row vector of benchmarks and x_i is the corresponding row vector of benchmark values for unit i. (This form is equivalent to the form given previously provided that there is a vector a such that $x_i a = 1$ for all i.)

The jackknife estimates are given by

$$\hat{y}^{\mathrm{GR}}_{(g)} = X(\Sigma_{i\notin g} a_i x_i' x_i)^{-1} \Sigma_{i\notin g} a_i x_i' y_i \qquad (4.6)$$

where $\Sigma_{i\notin g}$ denotes summation over all units not in replicate group g. The group jackknife variance estimator (4.4) is then applied to these estimates.

Theorem 2: Expectation of jackknife variance estimator for generalised regression

Under the assumption that the replicate groups are randomly drawn from a set D

$$\mathrm{E_D}(\mathrm{v}^{GJ}(\hat{y}^{GR})) = \mathrm{var_D}(\hat{y}^{GR}) + O(n^{-2}) \qquad (4.7)$$

Proof:

Write $\hat{\Omega}_g = \Sigma_{i\in g} a_i x_i' x_i$, $\hat{\Omega} = \frac{1}{G}\Sigma_j \hat{\Omega}_j$ and $\hat{\Omega}_{(g)} = \frac{1}{G-1}\Sigma_{j\neq g}\hat{\Omega}_j$,

and $\hat{\omega}_g = \Sigma_{i\in g} a_i x_i' y_i$, $\hat{\omega} = \frac{1}{G}\Sigma_j \hat{\omega}_g$ and $\hat{\omega}_{(g)} = \frac{1}{G-1}\Sigma_{j\neq g}\hat{\omega}_j$

so that $\hat{y}^{\mathrm{GR}} = X\hat{\Omega}^{-1}\hat{\omega}$ and $\hat{y}^{\mathrm{GR}}_{(g)} = X\hat{\Omega}^{-1}_{(g)}\hat{\omega}_{(g)}$.

Also let $\Omega = \mathrm{E_D}(\hat{\Omega}_g) = E_D(\hat{\Omega}) = E_D(\hat{\Omega}_{(g)})$ and $\omega = \mathrm{E_D}(\hat{\omega}_g) = E_D(\hat{\omega}) = E_D(\hat{\omega}_{(g)})$.

Then we can write

$$\begin{aligned}
\hat{y}^{\mathrm{GR}}_{(g)} - \hat{y}^{\mathrm{GR}} &= (X\hat{\Omega}^{-1}_{(g)}\hat{\omega}_{(g)} - X\Omega^{-1}\omega) - (X\hat{\Omega}^{-1}\hat{\omega} - X\Omega^{-1}\omega) \\
&= X\big((\hat{\Omega}^{-1}_{(g)} - \Omega^{-1})\omega + \hat{\Omega}^{-1}_{(g)}(\hat{\omega}_{(g)} - \omega) - (\hat{\Omega}^{-1} - \Omega^{-1})\omega + \hat{\Omega}^{-1}(\hat{\omega} - \omega)\big) \\
&\doteq X\big((\hat{\Omega}^{-1}_{(g)} - \hat{\Omega}^{-1})\omega + \Omega^{-1}(\hat{\omega}_{(g)} - \hat{\omega})\big) \qquad (*) \\
&\doteq X\big(\Omega^{-1}(\hat{\Omega}_{(g)} - \hat{\Omega})\Omega^{-1}\omega + \Omega^{-1}(\hat{\omega}_{(g)} - \hat{\omega})\big) \qquad (**) \\
&= X\big(\Omega^{-1}\tfrac{1}{G-1}(\hat{\Omega} - \hat{\Omega}_g)\Omega^{-1}\omega + \Omega^{-1}\tfrac{1}{G-1}(\hat{\omega} - \hat{\omega}_g)\big)
\end{aligned}$$

So
$$\begin{aligned}
\mathrm{E_D}(\mathrm{v}^{GJ}(\hat{y}^{GR})) &= \mathrm{E_D}\Big(\tfrac{G-1}{G}\sum_{g=1}^{G}(\hat{y}^{\mathrm{GR}}_{(g)} - \hat{y}^{\mathrm{GR}})^2\Big) \\
&\doteq \mathrm{E_D}\Big(\tfrac{1}{G(G-1)}\sum_{g=1}^{G}(X(\Omega^{-1}\hat{\Omega}_g\Omega^{-1}\omega + \Omega^{-1}\hat{\omega}_g) - X(\Omega^{-1}\hat{\Omega}\Omega^{-1}\omega + \Omega^{-1}\hat{\omega}))^2\Big) \\
&= \mathrm{var_D}(X(\Omega^{-1}\hat{\Omega}\Omega^{-1}\omega + \Omega^{-1}\hat{\omega})) \qquad \text{as in theorem 1} \\
&\doteq \mathrm{var_D}(X(\hat{\Omega}^{-1}\omega + \Omega^{-1}\hat{\omega})) \qquad (**) \\
&\doteq \mathrm{var_D}(X(\hat{\Omega}^{-1}\hat{\omega})) \qquad (*) \\
&= \mathrm{var_D}(\hat{y}^{GR})
\end{aligned}$$

The approximations at (*) arise by adding or subtracting terms of the form $(\hat{\Omega}^{-1} - \Omega^{-1})(\hat{\omega} - \omega)$ or $(\hat{\Omega}^{-1}_{(g)} - \Omega^{-1})(\hat{\omega}_{(g)} - \omega)$. These are terms of order n^{-2}. The approximations at (**) arise by using a Taylor series expansion and ignoring terms of order n^{-2} or greater. This proves the theorem.

4.7 Group jackknife for functions of estimates

Taylor series expansions are useful to show the application of the jackknife to a function of one or more estimates for which the jackknife is valid. For example, suppose that we have estimates $\hat{y}^1, \ldots, \hat{y}^J$ of population values $Y^1, \ldots, Y^J$. Using a Taylor series expansion, a function $\mathrm{f}(\hat{y}_1, \ldots, \hat{y}_J)$ can be written in the form

$$\mathrm{f}(\hat{y}^1, \ldots, \hat{y}^J) = \mathrm{f}(Y^1, \ldots, Y^J) + c_1(\hat{y}^1 - Y^1) + \ldots + c_J(\hat{y}^J - Y^J) + \text{higher order terms}$$

So $E_D(v^{GJ}(f(\hat{y}^1,...,\hat{y}^J))) \doteq E_D(v^{GJ}(c_1\hat{y}^1+...+c_J\hat{y}^J))$ (*)

$\doteq var_D(c_1\hat{y}^1+...+c_J\hat{y}^J)$ if the group jackknife is valid for each estimate

$\doteq var_D(f(\hat{y}^1,...,\hat{y}^J))$ (*)

It would be possible to approach the estimation of variance for such a function by doing the Taylor Series expansion and then using the group jackknife on the resulting linear expression. This appears to be inferior to applying the group jackknife directly - that is, producing a function value from the data dropping out each replicate and then applying (3). The reason is that the approximations at (*) in this derivation can be expected to cancel each other to some extent.

4.8 Properties of the group jackknife approach

Fixed number of replicate weights

To compute the variance of an estimate $\hat{y}$ using the group jackknife approach requires repeating the estimation process G times to obtain the jackknife estimates $\hat{y}_{(g)}$. This can be made straightforward by providing *G replicate weights* alongside the usual weights. The replicate weights are the set of weights $w_{(g)i}$ for which the jackknife estimates of total are obtained by weighted aggregation i.e. $\hat{y}_{(g)} = \Sigma_i w_{(g)i} y_i$. Given these replicate weights, any estimate that can be produced from the weighted data file can also be produced for the G replicates, and so a jackknife variance is available.

The replicate weights are obtained by performing all the steps of weighting but starting with a different set of selection weights. The selection weights for the whole sample w_i^π are replaced by replicate selection weights, which for replicate group g are given by

$$w_{(g)i}^\pi = 0 \qquad \text{for units in group g}$$

$$= \frac{G}{G-1} w_i^\pi \qquad \text{for units not in group g}$$

Importantly, once the replicate weights are provided a user can produce variance estimates from the unit data without needing to know any details of the weighting process, nor details such as stratum and cluster membership of the units. This opens the possibility of making replicate weights available to external researchers of confidentialised unit record files. It also allows a standard and simple process for generating standard errors within a tabulation package such as SUPERCROSS.

Simple application to complex estimates

The group jackknife provides a variance estimate for complex estimates, such as ratios. All that is required is that the estimates be produced separately using each set of replicate weights, and the jackknife formula applied to the results. There are a few exceptions to this - it is not clear that the jackknife will provide good estimates for the variance of quantiles such as the median (see Rao, Wu and Yu 1992). An indirect approach is available for estimating the variance of a quantile based on replicate weights - the method (due to Woodruff (1971)) is used in the software package WESVARPC.

Allowing for multiple steps of weighting

The group jackknife allows for multiple steps of weighting in a fairly natural way. The initial sample is divided into replicate groups, and initial replicate weights are produced. These are zero for units in a given replicate group and the selection weights multiplied by $G/(G-1)$ for other units. Each of the G sets of weights are then taken through all the stages of weighting.

This process is made easier by the GREGWT tool, which caters for a list of input weights for each unit and performs generalised regression weighting for each input weight. GREGWT also allows for the benchmark values to take different values for different replicates, as would arise if they were estimated from an earlier phase of the survey.

Some disadvantages

The group jackknife variance estimator is likely to be more variable than an estimator which works at cluster level. This is because only G replicates are available from which the variance is estimated. If G is too small the variance estimator could be quite unstable. On the other hand, too large a G may bias the variance, as the G replicates will not resemble each other sufficiently (because of the systematic sampling). Higher values of G also lead to a higher storage and computational cost.

Kott (1998) reports on using the group jackknife in the US National Agricultural Statistics Service, with $G = 15$. He notes that this leads to 95 percent confidence intervals about ten percent longer than usual, because they should based on to a student's t distribution with 15 degrees of freedom, rather than a normal distribution. This is less of an issue for $G = 30$.

4.9 The zoned jackknife

Suppose we have divided the sample into G groups, and have produced G replicate weights using the standard grouped jackknife approach of dropping one group at a time. This gives a variance estimator with only $G-1$ degrees of freedom, which may be too variable for some purposes.

An approach to increasing the stability of the estimator is to divide the units into a number of "zones", in such a way that estimates from the different zones have very little correlation. We can now write an estimate of total in the form $\hat{y} = \Sigma_j \hat{y}_j$, for $\hat{y}_j$ the estimate of the total for zone j. If we assume that estimates from the zones are independent we can estimate the variance of $\hat{y}$ as the sum of the group jackknife variance estimates for the zone estimates $\hat{y}_j$.

In many ABS household surveys estimates from the different states and territories could be considered approximately independent, as usually sampling and estimation across states are independent. The zoned jackknife is then likely to be quite good, with a low bias and lower variability than the group jackknife.

Properties of the zoned jackknife

The zoned jackknife variance estimate can be obtained using the same replicate weights as the group jackknife. The only extra information required is the zone to which each unit belongs. Suitable information for specifying zone may be available on a confidentialised unit record file provided for analysis outside the ABS.

The zoned jackknife should be more stable than the group jackknife since we have increased the number of degrees of freedom in the variance calculation. It may have a bias if there are post-strata or selection strata that lie in more than one "zone".

The zoned jackknife only applies directly to estimates of total. Functions of totals (e.g. ratios) would need to be linearised before the calculation. This will be presented in more detail in section 6.3.

5 A simple simulation

5.1 Simulating a population of clusters

We performed some simple simulations to get a feel for the behaviour of the group jackknife variance estimator (equation (4.4)) when the sample is selected systematically. In the simulations it is compared to two versions of the weighted residuals variance estimator. Standard weighted residuals (equation (3.1)) computes variances at stratum level by comparing all the cluster values. The split-halves version (equation (3.4)) computes variances at stratum level by comparing two variance groups formed systematically within the stratum.

We simulate a population of N clusters by generating cluster values from the distribution:

$$y_i \sim \mathrm{N}(r(2\frac{i-1}{N-1} - 1), 1)$$

so that the underlying means go from - r to r as the cluster identifier i goes from 1 to N. The value r controls how well the ordering of the clusters predicts their value; for r=0 the ordering gives no information. The population is divided into H frame strata, with stratum h containing clusters i for which $(h-1)N/H < i \leq hN/H$.

In each simulation we have K systematic samples of clusters with skip K from this population. (Thus the frame strata are not used in selection.) Each sample gives an estimate of the total $Y = \Sigma_i y_i$. We can evaluate the true variance of this estimate using the K possible systematic samples.

For each sample we can also compute variance estimates of the three types being considered. The mean and standard error of these variance estimates is obtained across the K samples. By comparison to the true mean for the population, the average bias and root mean squared error of the variance estimators can be obtained.

Note that this simulation is very simple, with the estimate of concern being a simple mean of cluster values across the population, rather than a generalised regression estimate. It is interesting because it shows how the various estimators perform when the situation is systematic sampling, the situation assumed by the group jackknife approach. Also note that the larger values of r are not chosen as realistic. Even $r = 2$ suggests clusters that are systematically very different to each other relative to the variability of estimates from the clusters. The $r = 6$ case is presented to show what happens in the extreme, rather than as a realistic possibility.

5.2 Results of simulations

The results shown in Table 1 are average figures from 500 simulations with $N = 123456$ and $K = 177$ (so number of sampled clusters is about 133). Three cases are presented: $r = 0$ (where there is no benefit from systematic sampling), r=2 and r=6 (which have increasing amounts of benefit from systematic sampling). For the group jackknife we tried numbers of groups $G = 15$ and 30. For the weighted residuals method we divided the population into $H = 5$ and 30 strata. Split-halves was computed with 30 strata, but forcing the stratum sample sizes to be even. Finally, the zoned jackknife was applied using G=30 replicate groups, but dividing the population into eight equal zones consisting of consecutive clusters.

Values for bias, standard error and root mean squared error are presented in Table 1 as percentages of the mean of the true standard errors from the 500 simulations (which was about 26.5 for each of the three r values).

Table 1: Bias, standard error and root mean squared error of standard error estimates
averaged over 500 simulations, as percentage of the mean true standard error

average values from 500 simulations	Group jackknife		Weighted residuals		Zoned jackknife	Split-halves
% of true standard error	G=15	G=30	H=5	H=30	G=30*	H=30**
	No benefit of ordering (r=0): True standard error mean=26.34					
bias	-1.7	-0.8	0.1	0.0	0.0	-1.1
standard error	18.7	13.1	2.7	2.7	4.7	12.9
root mean squared error	19.5	14.1	5.4	5.4	6.7	13.9
	Some benefit of ordering (r=2): True standard error mean=26.56					
bias	-1.5	0.2	2.4	-0.2	2.3	-1.0
standard error	18.7	13.1	2.7	2.7	4.7	12.8
root mean squared error	19.4	14.0	5.6	5.2	7.0	13.9
	High benefit of ordering (r=6): True standard error mean=26.31					
bias	2.0	10.4	22.2	1.0	18.3	-1.0
standard error	19.5	14.3	3.2	2.8	5.1	12.9
root mean squared error	20.3	18.2	22.5	5.2	19.0	13.9

* The clusters were divided into eight zones for the zoned jackknife calculation

** Stratum sample size was set to multiple of two for the split-halves calculation.

The weighted residuals variance estimator stands out as superior to the others, provided that the number of "strata" is sufficient to capture most of the benefits of the purposive ordering of the clusters in the systematic sample. The split-halves estimator within strata does not do as well. By using fewer degrees of freedom the estimator is less stable, and this is not compensated for by any reduction in bias.

The group jackknife variance estimator has quite a high variability, due to the reliance on a relatively small number (15 or 30) of groups. The variability is the main contributor to the root mean squared error of the estimator of standard error. There is some danger of introducing an appreciable bias if the systematic sampling is very important (the r=6 case here) and we use too many groups. In such a case the systematic difference between the groups is large enough to bias the standard error estimates upward.

The zoned jackknife has considerably lower variability than the group jackknife, at the cost of an increased bias. The bias remains small compared to the improved variability, particularly for the more realistic cases r = 0 or 2.

6 Example: Labour Force Survey

6.1 Characteristics of the Labour Force Survey

The Labour Force Survey is a straightforward example of a household survey using the ABS household frame. It is a very large survey, with around 30000 households sampled each month. The availability of monthly repeats of the same survey makes it a good setting for comparing the various estimators of variance. It also allows us to demonstrate the use of the group jackknife for estimating the variance of complex estimates that use data from more than one month - e.g. movement of unemployment rate, or trend of employed persons.

The Labour Force Survey estimates are weighted using post-stratified ratio estimation, with 540 post-strata (14 geographic regions by two sexes by 20 age groups).

6.2 Comparison of variance estimators

Variance estimates were obtained for estimates of employed persons and unemployed persons categorised by sex , broad age group and marital status. Each estimate had its variance computed a number of ways. First, the group jackknife with 15, 30 and 60 replicates (which will be labelled J15, J30 and J60 respectively in the graphs). Second, the weighted residuals variance estimator WR using frame strata as strata. Third, the split-halves variance estimator SH, again using frame strata as strata. Finally, the zoned jackknife was used, with 30 replicates and using state of usual residence as the zone variable. (State of usual residence is used in post-stratification, while state of enumeration is used in selection. Which to use as a zone variable appears to be a matter of convenience, since the two are almost identical.)

Graph 1 shows estimates of relative standard error (RSE) for employed persons from January 1993 to December 1999. Given that the sample design is consistent over time, we would expect the true relative standard error to change slowly (except perhaps between September 1997 and April 1998 when a new sample was introduced). The time series is presented to show the variability of the different estimates from month to month. This variability is much higher for the group jackknife estimators and the split-halves estimator than for the weighted residuals estimator. On the other hand, the difference between the estimators would appear to average about zero.

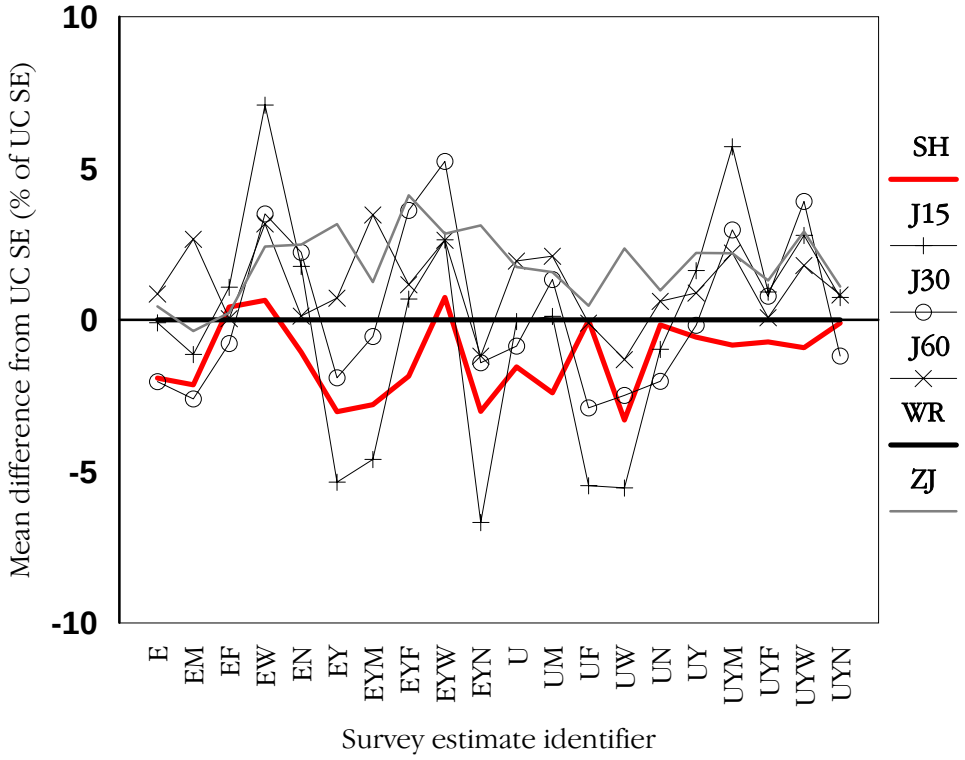
Graph 1: Estimated relative standard error, employed persons, January 1993 to December 1999

The group jackknife using only 15 replicates (J15) gives estimates that are considerably more variable than when using 30 replicates (J30). Using 60 replicates (J60) gives even lower variability. The zoned jackknife (ZJ) has lower variability than J60 even though it uses only the same 30 replicate weights used for the J30 estimator..

The comparison here seems to show that the split-halves estimator (SH) has lower variability than the group jackknife estimators, but greater than the weighted residuals (WR) estimator. The stability of the weighted residuals estimator is very noticeable, with the exception of an odd value in March 1993. The smoothness of the weighted residuals estimates increases our confidence in the reliability of this method.

To get a feel for the likely bias of the standard error estimates, we look at the difference between the average estimate from each method and the average estimate from the WR method, as a percentage of the average standard error using WR. Graph 2 presents these percentage differences for a set of twenty survey estimates.

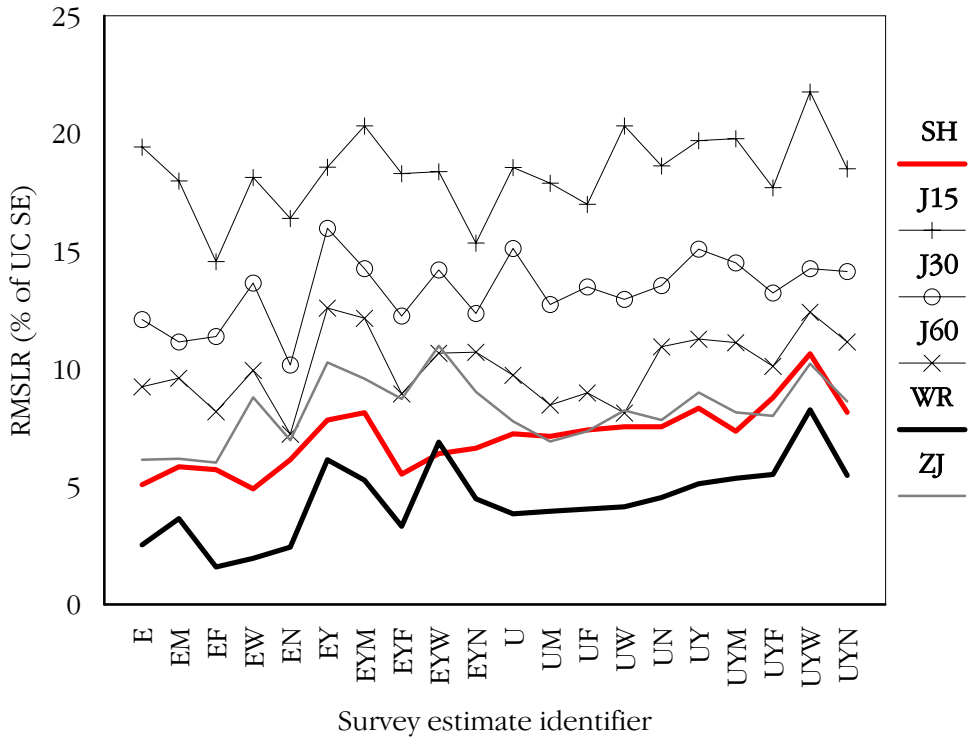
Graph 2: Mean difference from WR estimate for various standard error estimates
(as percent of the mean standard error using the WR estimator)
Averages are taken over the period January 1993 to December 1999
Identifiers for the survey estimates use the code: E/U=Employed/Unemployed; M/F=Male/Female; W/N=Married/Not married; Y=Under 25 years of age



It appears that none of the estimators are systematically different from the others by more than a few percent. The split-halves estimates are somewhat lower than the others, but it is not clear if they are more or less biased than estimates from weighted residuals or the group jackknife. The ZJ estimates are somewhat higher than the others, and this is likely to be a slight positive bias.

To get a measure of the variability of the standard error estimates, suppose that the true relative standard error was a linear function of time. From a given set of standard error estimates we can fit this regression line and obtain a residual from this line at each time point. Averaging the square of this residual over time gives a *root mean squared linear residual* (RMSLR) which measures the variability of the estimates around their linear interpolation. This root mean squared linear residual was computed for each estimator for the set of twenty estimates used in Graph 2. Graph 3 shows the root mean squared linear residuals for the SH, J15, J30, J60 and ZJ estimators as a percentage of the average standard error using WR.

Graph 3: Root mean squared linear residuals for various standard error estimates
(as percent of the average standard error using the WR estimator)
Linear residuals are defined to be the difference from a linear interpolation
Means are taken over the period January 1993 to December 1999
Identifiers for the survey estimates use the code: E/U=Employed/Unemployed; M/F=Male/Female; W/N=Married/Not married; Y=Under 25 years of age



From Graph 3 we see that the variability of the standard error estimates (around a straight line fitted through them) is consistently greater for the group jackknife method than for the weighted residuals method, with the difference reducing as more replicates are used. The split-halves estimates are more variable than the weighted residuals estimates, but less variable than the group jackknife even if 60 replicates are used.

The root mean squared linear residual values in graph 3 are larger than any corresponding biases suggested by graph 2, particularly for the more variable estimators. It thus appears that variability of the standard error estimates is a greater concern than any bias they may have. Even for fitting a variance model (as will be described in 7.1) a bias of the order indicated by graph 2 is unlikely to be of practical importance.

6.3 Standard errors for more complex estimates

Group jackknife calculation straightforward

The application of the group jackknife method to complex estimates was discussed in section 4.8. The complex estimate is simply calculated using each of the replicate weights, and the jackknife variance is then calculated using the standard formula (4.4). This approach is quite straightforward.

Weighted residuals requires linearisation

Obtaining variance estimates using the weighted residuals method is more difficult. A good method is presented by Andersson and Nordberg (1994).

First, note that it is simple to produce the weighted residuals variance estimates for an estimate that is a linear combination of estimates of total. The weighted residuals at stratum by variance group level are in this case obtained by applying the linear combination to the weighted residuals of the estimates of total. The formula (3.1) can then be applied to obtain variance estimates.

For non-linear functions of estimates of total, we can apply this same method to a linear approximation of the function obtained by a Taylor series expansion. This approach is well established - see for instance Särndal, Swensson and Wretman (1992). Binder (1996) discusses a general approach to calculating the linearisation variance estimator for complex weighting including generalised regression estimates. The main difficulty is in programming the method - this has in many agencies restricted its application to a few predefined forms of estimate. Hidiroglou, Bellhouse and Stafford (1997) proposes to use symbolic computation to automate the linearisation computation.

In the ABS a general program has been written to compute weighted residuals variance estimates for complex functions of generalised regression estimates. This program requires the user to specify the original function and the appropriate linearised form, leaving the computer to do the calculations. An automatic approach to the linearising of complex estimates has been applied by Andersson and Nordberg (1994) in the software CLAN.

Applying to a ratio estimate

A simple example that contrasts the group jackknife approach and weighted residuals approaches would be an estimate of ratio (such as unemployment rate). Here the group jackknife computes the ratio estimate separately for each replicate, then applies the jackknife formula (4.4) to obtain a variance estimate. This is straightforward and not computationally burdensome provided the number of groups is not too large.

The weighted residuals approach can provide a variance estimate after linearising the ratio. For estimates $\hat{y}$ of total Y and $\hat{z}$ of total Z we can write

$$\frac{\hat{y}}{\hat{z}} = \frac{Y+\Delta_y}{Z+\Delta_z} \simeq \frac{Y}{Z}+\frac{1}{Z}(\Delta_y+\Delta_z\frac{Y}{Z})$$

This linearisation is used to produce approximate weighted residuals $\hat{e}_{hg}^{Y/Z}$ at stratum by variance group level based on the weighted residuals $\hat{e}_{hg}^{Y}$ and $\hat{e}_{hg}^{Z}$ of the two variables:

$$\hat{e}_{hg}^{Y/Z} = \frac{1}{Z}(\hat{e}_{hg}^{Y}+\hat{e}_{hg}^{Z}\frac{Y}{Z})$$

These are then used in formula (3.1) to give the variance estimate.

Applying to estimates from multiple months

Either method can be applied to various estimates that are a linear combination of a number of months of data e.g. month to month movement, quarterly average, or even a linear approximation to the X11 trend.

In applying these methods to a repeating monthly survey we need to redefine the first stage sampling unit. The first stage sample is viewed not just as selecting clusters of dwellings but as selecting a sequence of clusters that will be interviewed over a long period. So the dwellings selected at a time point plus all the dwellings that replace them in sample over time are seen as belonging to the same first stage sample unit or cluster. For variance estimation it will be important to keep track of which units are in the same cluster at different times.

This definition of cluster enables the estimation of the variance of estimates based on multiple time points. The contribution of a cluster to a total is simply the sum of the contributions of all the units that are in that cluster at the different time points.

7 Evaluations for complex survey weighting situations

7.1 Modelling variances

The Labour Force Survey provided a good setting for comparing the variability of different variance estimators. In particular, we had the opportunity to compare the same variance estimates for a variety of time points. This option is not available for evaluating one-off surveys. Evaluations of variance estimators in these surveys have focused on how well the variance estimates can be used to fit a model.

Variance models, or generalised variance functions, are used by the ABS to provide an approximate variance for an estimate as a function of the size of the category being estimated — the size often being measured by the estimated number of persons in the category. The usual approach to fitting such a model is to define a large number of categories (such as combinations of state, age group, marital status and employment status) for which estimates and their variances are calculated. Size measures for the categories are also produced - for person estimates these are usually the estimates themselves.

Write R_c for the relative standard error estimate for an estimate for category c with size measure E_c. A variance model is used to fit R_c to E_c, usually taking the form:

$$\log(R_c) = a + b\log(E_c) + c\log(E_c)^2 + \varepsilon_c$$

with the errors ε_c assumed independent normal, often with common variance. The fitted model is published using tables from which a user can predict the relative standard error of any estimate (of the type modelled), given the size measure for the category the estimate applies to.

We can look at the variability of particular relative standard error estimates R_c by observing their spread around the modelled value. Unfortunately, even if we had the true relative standard error values, they would not fit the model exactly. So comparisons of the different variance estimators in this way is limited to whether they lead to different models and whether the modelled values are noticeably more variable.

7.2 Some results for the National Nutrition Survey

The 1995 National Nutrition Survey (NNS) was a complex survey in that it was a subsample from a larger survey, the National Health Survey 1995 (NHS). The weights for the NNS were obtained by performing successively three sets of adjustments to the weights the sampled units had in the NHS. The first set of adjustments dealt with a unit's probability of selection in the NNS. The second set aimed to reduce the effect of differing response rates among various groups in the population. These included differing adjustments for non-response classes based on modelling of the non-response probability. The final adjustment was to perform calibration of the weights to a number of demographic benchmarks.

For application of the group jackknife to this survey the NHS units were divided into 30 replicate groups. For each group in turn, the units excluding that group were put through all the stages of adjustment and given a resulting weight. This gave 30 replicate weights for use in the group jackknife variance estimator.

A comparison was made to a group jackknife based on only 15 groups. The difference in variability of the resulting variance estimates was noticeable, with 15 groups giving a model with increased variability (but little difference in bias)

At the time when this evaluation was done it was usual to use the approximate split-halves variance estimator described in section 3.4 for variance estimation. This estimator depends upon choosing a post-stratification which would approximate the effect of the final calibration

adjustment. The investigation showed that an unrealistically fine choice of this post-stratification could lead to extreme underestimation of the variance. A more realistic choice of post-stratification variable gave much better estimates. Models based on these approximate split-halves estimates of variance were very comparable to models produced from the group jackknife variance estimates. The approximate split-halves variance estimates were slightly more stable than those produced by the group jackknife with 30 replicates. At the time of this analysis there was little software to support the group jackknife estimation, and these results were used to justify continued use of the approximate split-halves estimation approach.

Other evaluations on this data looked at alternative estimations with fewer adjustment steps. This work highlighted the adjustment steps that had made a large impact on the weights. In particular, omitting the non-response adjustments would have led to greatly increased standard errors for the estimates.

7.3 Various other findings from complex surveys

Variance estimation has now been studied for a number of surveys, using the technique of comparing models from different estimators. A number of general observations can be made from these studies.

First, from the point of view of fitting models, there is not much to be gained from improving the variability of variance estimates relative to the previously-used split-halves methodology. This is because the true values do not fit a simple model accurately enough to give much gain from more accurate values. A group jackknife estimator based on 30 replicates gives variances of similar stability to the split-halves approach, and this is accurate enough for fitting variance models. On the other hand, using much lower numbers of replicates (e.g. 15 or fewer) does noticeably affect the variability of the models.

In the Labour Force survey example in section 6, standard errors from split-halves had lower variability than those from the group jackknife with $G = 30$. The situation here is different, in that the modelled standard error estimates include many estimates by state. State estimates are based on much fewer strata, which increases the variability of split-halves standard error estimates, relative to group jackknife estimates (which have the same number of degrees of freedom for state estimates as for Australian estimates).

Bias would be a much more important issue than variability, but there is little evidence of systematic bias in the resulting variance models. Biases have been noted in particular types of estimate. For example, in experiments depending on calibration to very fine benchmarks, the weighted residual or split-halves estimates were biased downwards. The approximate split-halves estimator is biased upwards for categories which are close to benchmark values that were used in the actual weighting but not in the approximate post-stratification chosen.

There has been little evidence of gains from replicating the whole sequence of adjustments in the group jackknife. The weighted residual method is applied to only the last stage of adjustments, effectively treating the input weights as fixed when in fact they are the results of previous adjustments. This could bias the variance estimates in cases where the final adjustment is relatively cosmetic, and earlier stages of adjustment had a large effect. We have yet to encounter such a case in our testing. The group jackknife has the advantage of protecting against any such bias.

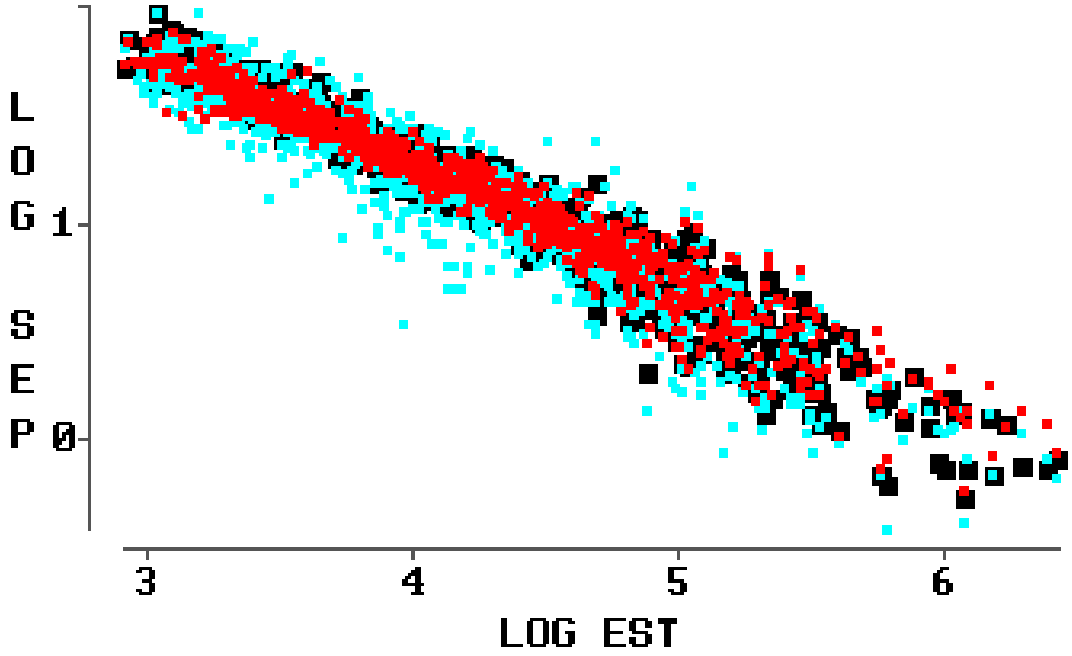
In summary, the group jackknife with around 30 replicates appears sufficiently stable for modelling purposes and is approximately unbiased even for complex weighting processes. The weighted residuals variance estimates have a lower variability, and in most practical cases are approximately unbiased. Before using the weighted residuals method the user should consider

whether the weighting adjustment process used for the survey has any special features which could bias the resulting variance estimates.

An illustrative example

Graph 4 presents a typical graph of points to be modelled, to illustrate the points made above.

Graph 4: Various estimates of relative standard error vs. estimate size
LOG SEP and LOG EST are (base 10) logs of relative standard error % and estimate respectively.
Points are for different estimates from the 1997 Family Characteristics Survey.
Mid-tone points are from approximate split-halves
Black squares (partially hidden) are from group jackknife with 28 replicates
Light points (partially hidden) are from group jackknife with 7 replicates



This graph shows various relative standard error estimates from the 1997 Family Characteristics Survey, plotted against the corresponding estimates on a log-log scale. The mid-tone points in this graph are from an approximate split-halves methodology (as described in section 3.4). We model these points with a quadratic curve on this log-log scale.

Much of the lack of fit in such a curve is due to the fact that the true relative standard errors do not follow such a curve exactly - they themselves form a cloud with not much less spread than the approximate split-halves points. Reducing the variability of the standard error estimates will thus not greatly improve on this model.

Also shown on the graph are relative standard error estimates from the group jackknife with 28 replicates (black squares) and 7 replicates (light points). The extra variability from using only seven replicates is clear. There is also some indication that the approximate split-halves method overestimates the standard error for the largest estimates. This may be because the post-strata used for the approximation do not fully reflect the benefits of the actual estimation approach for large estimates.

8 Discussion

There is continuing pressure to make household surveys more efficient and to extend the range of outputs that they can support. This has led over the past twenty years to an increase in the complexity of the surveys themselves, and of the estimation processes applied to them. This included the introduction of multiple steps of weighting and of calibration of weights to multiple benchmarks.

Over the last few years the ABS has actively investigated calibration methods and methods for obtaining variance estimates for complex survey designs and estimation methods. We have produced an efficient SAS macro, GREGWT, to support calibration using the generalised regression estimator and variants of this estimator. This macro also supports the new variance methods that have been evaluated, particularly the weighted residuals and group jackknife methods.

Properties of variance estimators

This paper looked specifically at the merits of the various variance estimators proposed. A summary of their properties follows.

1. Group jackknife:
   - Simple to apply even for complex estimates
   - Thirty replicates gives sufficient stability for variance modelling
   - Accounts for all steps of the weighting adjustment process
   - Requires a number of replicate weights attached to each survey unit
   - Allows external users to estimate standard errors from confidentialised files of unit data using the replicate weights on the file

2. Weighted residuals
   - Much more stable estimates than group jackknife
   - Approximately unbiased in many practical cases
   - Only accounts for final stage of the weighting process
   - Requires stratum and variance group identifiers for each survey unit
   - Does not allow external users to calculate standard errors from confidentialised unit records, since the required information cannot be released to external users because of confidentiality issues

3. Split-halves
   - Similar properties to weighted residuals but without the low variability
   - There is no practical evidence that split-halves is less biased than weighted residuals
   - Approximate split-halves may be useful where stratum and variance group are known but the exact details of weighting are unknown

4. Zoned jackknife
   - Uses the same replicate weights as the group jackknife, plus a zone indicator
   - Can give variance estimates with considerably lower variability than the group jackknife, though with the danger of an increased bias
   - Not as simple to apply as the group jackknife, particularly for complex estimates
   - Could be used by an external user where the accuracy of a single variance estimate is critical

Application for modelling

Historically, the ABS has presented standard errors for its household survey estimates using models. These models relate the standard error to the size of the estimate itself or the category it is produced for. These models are necessarily inexact at predicting the standard error of an individual estimate.

For the purpose of fitting such a model the key requirement is unbiased estimates of standard error. The split-halves estimator used historically has provided these well enough for this purpose. The weighted residuals estimator is an improvement, with much lower variability. However, either of these methods are potentially biased for certain types of weighting processes. The group jackknife estimator is sufficiently stable for modelling purposes and is straightforward to apply in a complex weighting situation so as to capture all contributions to the variability.

Application for providing standard errors for individual estimates

The choice of a variance estimation approach depends on such factors as ease of computation, accuracy and reliability of the variance estimators, information required for calculation (and availability of such information on confidentialised files) and the use to which the variance estimates will be put.

For the purpose of producing a broad summary of standard errors in publications, the ABS constructs models. However, to provide models that cover the full variety of estimates being published is quite time-consuming. Although it is possible to develop separate models for different grouping of variables, in practice, very few users make use of these detailed models. Standard errors from the models can be subject to moderate model errors - a standard error directly estimated for each variable would not be subject to this limitation..

An alternative to the modelling approach would be to provide a facility to estimate the standard error for individual estimates. The technology to do this will be made available through the SUPERCROSS tabulation package. This package efficiently produces a wide range of estimates, and can be modified to provide standard errors for the group jackknife approach. The simplicity of the group jackknife calculations makes this method ideal for this context.

The main negative for the group jackknife is that the standard error estimates can be quite variable. If this becomes the usual source of standard error estimates from the SUPERCROSS package, it would be appropriate to provide a warning to users. This could state that the standard errors are only estimates, and perhaps suggest smoothing them across a number of similar estimates in cases where a more reliable standard error estimate is needed.

Conclusion

The above considerations suggest that the group jackknife approach is suitable for use as the standard approach to estimating variances for ABS household surveys for the purpose of publication. Other methods such as the weighted residuals method or the zoned jackknife may be appropriate in situations where the quality of standard error estimates is critical (and the assumptions underlying the method used are reasonable). This could apply, for example, in evaluating different methodologies, as in such studies the precision of the standard errors may be critical to detecting small effects.

Appendix: Algorithms implemented in GREGWT macro

Introduction

A SAS macro GREGWT has been written in the ABS to perform calibrated weighting, included generalised regression estimation and the variants described in section 2.8. The standard generalised regression method does not place any restrictions on the size of the weights - this can lead to problems such as negative weights. Variants described in section 2.8 arise as responses to these problems.

All the approaches modify some initial weights to produce weights that aggregate to known benchmark constraints. The methods are described in Singh and Mohl (1996). The algorithms used in GREGWT correspond to Method 5, the "Truncated Linear" method, and Method 6, 'the Truncated Exponential" method, of Singh and Mohl (1996), They introduce range restrictions on the weights in addition to the benchmark constraints. Using these methods the weights will always meet the range restrictions. If convergence is not achieved then the benchmark constraints will not be met, rather than the range restrictions being ignored. Convergence problems generally suggest some problem with the weighting situation being addressed.

GREGWT applies the algorithms without at any stage storing the intermediate weights produced at each iteration. This allows efficient computation of all iterations within a single SAS data step, without storing the weights in internal memory. The algorithms are given below in this form, rather than as presented by Singh and Mohl (1996).

Generalised regression method

Generalised regression as described in section 2.5 corresponds to Method 1 of Singh and Mohl (1996). As in section 2, suppose that x_i is a row vector of auxiliary variables, and X is a corresponding row vector of benchmark values and w_i^A are input weights. The generalised regression weights w_i^{GR} are obtained by the following calculations:

$$\hat{x}^{\mathrm{A}} = \sum_i w_i^{\mathrm{A}} x_i$$

$$T^{\mathrm{A}} = \sum_i a_i^{(0)} x_i x_i' / c_i$$

Calculate A a solution to the equation $(X - \hat{x}^{\mathrm{A}}) = AT^{\mathrm{A}}$

$$w_i^{\mathrm{GR}} = w_i^{\mathrm{A}}(1 + A x_i' / c_i)$$

This requires finding A a solution to an equation $(X - \hat{x}^{\mathrm{A}}) = AT^{\mathrm{A}}$. GREGWT finds this solution by decomposing T^{A} into the form $T^{\mathrm{A}} = U'U$ for U an upper triangular matrix. It is then straightforward to successively solve $(X - \hat{x}^{\mathrm{A}}) = A\,U'U$ for AU', and then for A itself.

This will lead to weights that fulfil the benchmark constraints $\sum_i w_i^{\mathrm{GR}} x_i = X$, provided that the matrix T^{A} is singular. For non-singular T^{A} this calculation ignores any benchmark constraints corresponding to rows of T^{A} that are linearly dependent on previous rows. These ignored benchmarks may still be met, if the values provided for them are consistent with the constraints that were not ignored. If they are not met any differences will be flagged by the macro.

The truncated linear regression method

The idea of the truncated linear regression method is to perform the standard generalised regression calculations above, then to truncate the weights so that they lie within specified bounds $[L_i, U_i]$ for each unit i. These bounds could be constant across units or proportional to the original weights, or specific to individual units. If truncation occurs, the benchmark constraints will not be met by the weights. This leads to an iterative approach which should match the benchmark constraints after a few iterations.

The truncated linear regression algorithm proceeds as follows,

Step 1: Initialise: (note that the superscript in parentheses denotes the iteration)

$a_i^{(0)} = w_i^{\mathrm{A}}$ for all units i

$X^{(0)} = \sum_i a_i^{(0)} x_i$

$T^{(0)} = \sum_i a_i^{(0)} x_i x_i' / c_i$

Calculate $A^{(0)}$ a solution to the equation $(X - X^{(0)}) = A^{(0)} T^{(0)}$

Go to step 2 for iteration $m = 1$.

Step 2. For each unit i:

$\tilde{w}^{(m)} = w_i^{\mathrm{A}}(1 + A^{(m-1)} x_i' / c_i)$

if $\tilde{w}^{(m)} < L_i$ then set $w_i^{(m)} = L_i$ and $a_i^{(m)} = 0$

else if $\tilde{w}^{(m)} > U_i$ then set $w_i^{(m)} = U_i$ and $a_i^{(m)} = 0$

else set $w_i^{(m)} = \tilde{w}_i^{(m)}$ and $a_i^{(m)} = w_i^{\mathrm{A}}$

Step 3. $X^{(m)} = \sum_i w_i^{(m)} x_i$

$T^{(m)} = \sum_i a_i^{(m)} x_i x_i' / c_i$

Calculate $\tilde{A}^{(m)}$ a solution to the equation $(X - X^{(m)}) = \tilde{A}^{(m)} T^{(m)}$

$A^{(m)} = A^{(m-1)} + \tilde{A}^{(m)}$

Step 4. Convergence is achieved if for all elements (indexed by p) of row vectors X, $X^{(m)}$, $A^{(m-1)}$ and $A^{(m)}$ one of the two conditions below is met:

either $\left| X_p - X_p^{(m)} \right| < \varepsilon^X$ (benchmark constraints met)

or $\left| A_p^{(m-1)} - A_p^{(m)} \right| < \varepsilon^A$ (no improvement)

for specified small values ε^X and ε^A.

If convergence is achieved or the maximum number of iterations is reached then stop. Otherwise, increment m and repeat from step 2.

At convergence, the final weights are given by $w_i^{(m)}$.

According to Singh and Mohl, the iterations above constitute the Newton-Raphson steps for minimising the generalised least squares distance function (2.11) for w_i in $[L_i, U_i]$, subject to the benchmark constraints.

The truncated exponential method

The truncated exponential method minimises the distance function

$$F^{EXP} = \Sigma_i c_i[w_i \log(w_i/w_i^A) - w_i + w_i^A] \quad (2.13)$$

for w_i in $[L_i, U_i]$, subject to the benchmark constraints. The Newton-Raphson steps for minimising this function are the same as for the linear distance function, except at step 2, which is revised as follows.

Step 2*. For each unit i:

$\tilde{w}^{(m)} = w_i^A \exp(A^{(m-1)} x_i'/c_i)$

if $\tilde{w}^{(m)} < L_i$ then set $w_i^{(m)} = L_i$ and $a_i^{(m)} = 0$

else if $\tilde{w}^{(m)} > U_i$ then set $w_i^{(m)} = U_i$ and $a_i^{(m)} = 0$

else set $w_i^{(m)} = \tilde{w}_i^{(m)}$ and $a_i^{(m)} = \tilde{w}_i^{(m)}$

This is method 6 described in Singh and Mohl (1996). Setting $a_i^{(m)} = \tilde{w}_i^{(m)}$ appears in the appendix of Sing and Mohl (1996), though the description in the text uses $a_i^{(m)} = w_i^A$, as for the linear case. In my experience either formula appears to work, but the former leads to more rapid convergence; this is the formula used in the GREGWT macro.

Bibliography

Andersson, C. and Nordberg, L. 1994, 'A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys - Theory and Software Implementation', *Journal of Official Statistics*, Vol. 10, No. 4, pp. 399-405

Bethlehem, J.G. and Keller, W.J. 1987, 'Linear Weighting of Sample Survey Data', *Journal of Official Statistics*, Vol 3, pp. 141-153.

Binder D.A. 1996, 'Linearization Methods for Single Phase and Two-phase samples: A Cookbook Approach', *Survey Methodology*, Vol.22, pp. 17-22.

Cochran, W.G. 1963, *Sampling Techniques, Second Edition*, John Wiley and Sons, New York

Deville, J.C. and Särndal, C.E. 1992, 'Calibration Estimators in Survey Sampling', *Journal of the American Statistical Association*, Vol. 87, pp. 376-382.

Deville, J.C., Särndal, C.-E. and Suatory, O. 1993, 'Generalized Raking Procedures in Survey Sampling', *Journal of the American Statistical Association*, Vol. 88, pp. 1013-1020.

Horvitz, D.G. and Thompson, D.J. 1952 'A Generalisation of Sampling Without Replacement from a Finite Universe', *Journal of the American statistical* Association, Vol. 47, pp. 663-685.

Hidiroglou, M.A., Bellhouse, D.A., and Stafford, J.E. 1997, 'Generalised Estimation Systems with Future Enhancements Using Symbolic Logic', *Bulletin of the 51st session of the International Statistical Institute*, Book 1, pp 483 -486

Hidiroglou, M.A,, Fuller, W.A. and Hickman, R.D. 1980 *SUPERCARP*, Department of Statistics, Iowa State University, Ames, Iowa.

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.

Lemaître, G. and Dufour, J. 1986, 'Comparative Study of Weighting Methods for Family Data' Proceedings of the Survey Methods Research Section of the American Statistical Association.

Rao, J.N.K. and Singh, A.C. 1997, 'A Ridge-Shrinkage Method for Range-Restricted Weight Calibration in Survey Sampling', *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp.57-64.

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

Särndal, C.-E., Swensson, B. and Wretman, J.H. 1989, 'The Weighted residual Technique for Estimating the Variance of the General Regression Estimator for the Finite Population Total', *Biometrika*, Vol. 76, pp. 527-537.

Sarndal, C.E., Swensson, B, and Wretman J.H. 1992, *Model Assisted Survey Sampling*, Springer-Verlag.

Singh, A.C. and Mohl, C.A. 1996, 'Understanding Calibration Estimators in Survey Sampling', *Survey Methodology*, Vol. 22, pp. 107-115.

Stukel, D.M., Hidiroglou M.A. and Särndal, C.-E. 1996, 'Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization', *Survey Methodology*, Vol. 22, pp. 117-125

Yung, W. and Rao J.N.K. 1996, 'Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling', *Survey Methodology*, Vol. 22, pp. 23 - 31.

Wolter, K.M. 1985, *Introduction to Variance Estimation,* Springer-Verlag, New York.

Woodruff, R.S. (1952) Confidence intervals for medians and other position measures, *Journal of the American Statistical Association*, Vol. 57, pp. 622-627.