EFFICIENT REPLICATED VARIANCE ESTIMATION

Keith Rust, The University of Michigan and
Australian Bureau of Statistics

1. Introduction

Replicated variance estimation techniques have become standard procedures for use in analyzing sample survey data, particularly since their general usefulness was demonstrated by Kish and Frankel (1974). The techniques provide variance estimates for non-linear estimators, accounting for the complexities of the sample design.

Two forms of replicated variance estimator are used commonly: the jackknife and Balanced Repeated Replication (BRR). Both form estimates of the population parameter of interest, each based on a different subsample (replicate) of the full sample, and use the variability among these estimates to derive a variance estimate for the full sample estimator. In the standard forms of these variance estimators, described in Wolter (1985) for example, the number of replicate estimates of the parameter generated depends upon the sample design, and in particular on the number of primary sampling units (PSUs) selected (the number of sampling units in a single stage design). The number of replicate estimates is generally close to the number of PSUs. For large scale sample surveys the number of such sampling units is typically in the hundreds for multistage surveys and the thousands for single stage surveys. Thus use of the standard forms of the jackknife and BRR variance estimators requires the formation of many replicate estimates for each parameter of interest. Although computers can perform the calculations needed for replicated variance estimation relatively routinely, in the practice of such large scale surveys economies are needed as a result of the scale and complexity of survey estimation. There are several features of this problem.

First, almost always in such surveys, information is collected on a large number of variables, and the results presented contain estimates for many population parameters (tens or hundreds). Thus in situations where replicated variance estimation is typically applied, the derivation of variance estimates for even a subset of the survey variables will involve the calculation of a very large number of replicate estimates.

Second, often survey estimators involve the use of post-stratification and non-response adjustment. These techniques require that a weight, dependent upon the sample data, be attached to each unit, and used in estimation. Such weights should ideally be recomputed for each replicate estimate, using only data from the replicate subsample. Although it is possible to use the single set of whole sample weights for each replicate estimate, Lemeshow (1979) has shown that substantial bias in variance estimation may result from such an approach. The use of separate weighting for each replicate increases the complexity of using replicated methods.

Third, increasingly survey data are being used to examine relationships among the survey variables in the population surveyed. These analyses are generally multivariate in nature, and often involve the use of complex parameter estimators. These estimators may be iterative, such as the estimator of the coefficients in a logistic regression analysis. The use of replicated variance estimation techniques is attractive in such cases, because of the difficulty in obtaining an explicit variance estimator for the parameter estimate. Even when such explicit variance estimators are available, they require specialized computer programs, specific to each estimator. However, as the iterative procedure must be repeated for each subsample replicate, the amount of computation required for replicated variance estimation is very great if there are many replicates.

These features of the circumstances in which replicated variance estimation techniques are used indicate that their usefulness and convenience are enhanced if the number of replicates required for a single variance estimate can be limited, while maintaining adequate precision of variance estimation. The aim of this paper is to show that this is possible in many applications. In Section 2, the requirements for a variance estimator for use in survey inference are considered. The necessary survey design and estimation notation are given in Section 3. Section 4 discusses strategies for using the jackknife with reduced replication. These alternatives are compared analytically in Section 5, and Section 6 illustrates an efficient procedure using a hypothetical population. BRR is considered in Section 7. Aspects of the practical application of these replicated methods are noted in Section 8.

2. The Precision of a Variance Estimator

Sampling variances for survey estimates are required for use in making inference about population parameters. Their major use is in constructing confidence intervals. If an estimator $\hat{\theta}$ with small bias is used to estimate θ, and $v(\hat{\theta})$ provides an approximately unbiased estimator of $V(\hat{\theta})$, the sampling variance of $\hat{\theta}$, then two-sided confidence intervals for θ are constructed in the form $\hat{\theta} \pm t\sqrt{v(\hat{\theta})}$, where t is chosen to be as small as possible while giving the required level of confidence. The appropriate choice of t thus depends primarily on the confidence coefficient $(1-\alpha)$, and on the precision of the variance estimator $v(\hat{\theta})$. The precision of the variance estimator can be expressed as degrees of freedom, r, where $r = 2V(\hat{\theta})^2 \div V(v(\hat{\theta}))$. For large samples the use of the $(1-\alpha/2)$th quantile of the Student t distribution with r degrees of freedom as the value for t in forming confidence intervals will give coverage close to $(1-\alpha)$. For given α, this value decreases with increasing r, and thus with increasing precision of $v(\hat{\theta})$. However, beyond about 25 or 30 degrees of freedom the quantiles of the t distribution vary little with the number of degrees of freedom, being close to those of the normal distribution. Thus it is common practice to use as 95% confidence intervals, intervals of the form $\hat{\theta} \pm 1.96\sqrt{v(\hat{\theta})}$ or $\hat{\theta} \pm 2\sqrt{v(\hat{\theta})}$, provided that r is at least 25 or 30. Hence for the purposes of making inference about a parameter θ, the precision of variance estimation is not of great importance provided that at least 25 to 30 degrees of freedom are attained.

For stratified designs with many sampled PSUs, 25 to 30 degrees of freedom can often be attained with replicated variance estimators using few more than 30 replicates. However, to attain the required precision with a relatively small number of replicates, care is often required in the way replicates are formed, and methods for doing this are discussed below.

3. Notation

Consider a sample design with H strata, with $n_h \geq 2$ PSUs selected independently within each stratum, giving a total of $n = \sum_{h=1}^{H} n_h$ PSUs selected. The sample can be single stage, or subsampling of second and later stage units can take place within selected PSUs, giving a multistage design.

Consider an unbiased linear estimator $\hat{\theta} = \sum_{h=1}^{H} W_h \hat{\theta}_h$ of a parameter $\theta = \sum_{h=1}^{H} W_h \theta_h$, so that the estimator is a sum of stratum estimators with known weights. The unbiased estimator of θ_h, $\hat{\theta}_h$, is given by

$$\hat{\theta}_h = \Big(\sum_{i=1}^{n_h} X_{hi}\Big)/n_h ,$$

where X_{hi} is the unbiased estimate of θ_h derived from the units subsampled within PSU i of stratum h, incorporating appropriate weighting factors. If the sample is single stage, X_{hi} is a measurement on unit i in stratum h, weighted inversely to its probability of selection. Let $\sigma_h^2 = E(X_{hi} - \theta_h)^2$, $\mu_h^{(4)} = E(X_{hi} - \theta_h)^4$ and $\beta_h = \mu_h^{(4)}/\sigma_h^4$, the kurtosis of the PSU estimates for stratum h. Thus $V(\hat{\theta}) = \sum_{h=1}^{H} W_h^2 \sigma_h^2/n_h$.

The precision of a number of replicated variance estimators will be considered below for such unbiased linear estimators. The results generalize approximately for non-linear estimators with small bias by considering the stratum specific terms θ_h, $\hat{\theta}_h$, σ_h^2 and β_h as relating to the appropriate linear substitute for the estimator in question, derived from a Taylor series expansion.

4. The Jackknife Variance Estimator

For an estimator $\hat{\theta}$ of θ, let $\hat{\theta}_{(ih)}$ denote the estimate of θ derived from the subsample consisting of the full sample with all units from PSU i, stratum h omitted. For linear unbiased $\hat{\theta}$, in $\hat{\theta}_{(ih)}$ the term θ_h is estimated by $\sum_{j \neq i} X_{hj}/(n_h - 1)$. The standard (full) jackknife variance estimator of $V(\hat{\theta})$ is

$$v_{FJ}(\hat{\theta}) = \sum_{h=1}^{H} \frac{(n_h - 1)}{n_h} \sum_{i=1}^{n_h} \big(\hat{\theta}_{(ih)} - \hat{\theta}\big)^2 .$$

Thus n replicate estimates $\hat{\theta}_{(ih)}$ are required to derive $v_{FJ}(\hat{\theta})$ from a particular sample. For unbiased linear $\hat{\theta}$, v_{FJ} is unbiased, and its variance is

$$V\big(v_{FJ}(\hat{\theta})\big) = \sum_{h=1}^{H} \frac{W_h^4 \sigma_h^4}{n_h^2} \left[\frac{(\beta_h - 3)}{n_h} + \frac{2}{(n_h - 1)}\right].$$

A more general jackknife variance estimator requiring fewer than n replicates can be obtained by omitting more than one PSU from each replicate, and/or omitting only some PSUs from any of the replicates. A general formulation is given below, and specific cases are studied in more detail.

Let the H strata be partitioned into G combined strata, with combined stratum g consisting of $H_g \geq 1$ strata. Groups of units to be omitted for the formation of replicates are determined in the following way. A single such dropout group consists of s_h $(1 \leq s_h < n_h)$ PSUs from stratum h, for each of the H_g strata in combined stratum g. The s_h values are such that $f_h = s_h/n_h$, the fraction of PSUs from stratum h which are omitted, is constant within each combined stratum, so that $f_h = f_g$. Let $F_g = f_g^{-1}$. The s_h PSUs are selected so as to constitute a simple random sample without replacement of the n_h PSUs in stratum h. A replicate is formed by taking the entire sample and omitting the units in the dropout group. A total of ℓ_g dropout groups (and thus ℓ_g replicates) are formed from each combined stratum by repeating this process ℓ_g times (not necessarily independently). A total of $L = \sum_{g=1}^{G} \ell_g$ replicates are formed by omitting each dropout group in turn.

Let $\hat{\theta}_{(ig)}$ denote the unbiased estimator of θ based on the replicate formed by omitting the ith dropout group from combined stratum g ($i = 1, 2, ..., \ell_g$). A jackknife variance estimator of $V(\hat{\theta})$ using these L replicates is

$$v_J(\hat{\theta}) = \sum_{g=1}^{G} \frac{(F_g - 1)}{\ell_g} \sum_{i=1}^{\ell_g} \big(\hat{\theta}_{(ig)} - \hat{\theta}\big)^2 .$$

If $\hat{\theta}$ is an unbiased linear estimator of θ, then $v_J(\hat{\theta})$ is an unbiased estimator of $V(\hat{\theta})$.

If in addition each PSU appears in at most one dropout group, and hence is included in either (L − 1) or L replicates, $V(v_J(\hat{\theta}))$ for unbiased linear $\hat{\theta}$ is given by

$$V\big(v_J(\hat{\theta})\big) =$$

$$\sum_{g=1}^{G} \left[\Big(\sum_{h \in g}^{H_g} W_h^4 \sigma_h^4 (\beta_h - 3)/n_h^3\Big)\big(F_g(F_g - 2)^2 + \ell_g(2F_g - 3)\big) + 2\Big(\sum_{h \in g}^{H_g} W_h^2 \sigma_h^2/n_h\Big)^2 \big((F_g - 1)^2 + (\ell_g - 1)\big) \right] \ell_g^{-1}(F_g - 1)^{-2}.$$

This is shown in Rust (1984, Appendix A). The standard jackknife variance estimator, v_{FJ}, is a particular case of v_J with G = H, $H_g = 1$ for all g, $s_h = 1$ for all h, $F_h = \ell_h = n_h$ for all h, and L = n, with each unit appearing in a single dropout group. Among the class of jackknife variance estimators, for which each PSU is included in at most one dropout group, v_{FJ} has the smallest variance, but requires the most replicates. Only jackknife variance estimators with each PSU included in at most one dropout group (giving disjoint dropout groups) will be considered below.

Given a fixed number of replicates, L $(2 \leq L \leq n)$, there are a number of ways in which dropout groups can be formed for use with v_J. Strata can be combined, so that G < H, and some dropout groups will contain PSUs from more than one stratum. PSUs can be grouped within strata and dropped out together, so that $s_h > 1$ for some h. Only some of the units need be included in any dropout group, so that $\ell_g s_h < n_h$ for some or all $h \in g$, g. Three particular cases are considered below.

A. The Grouped Jackknife Variance Estimator (v_{GJ})

For this special case, no strata are combined, so that $H = G$. For each stratum ℓ_h dropout groups are formed, and each contains $s_h = [n_h/\ell_h]$ units (where [ ] denotes the integer part). For example if $n_h = 13$ and $\ell_h = 4$, $s_h = 3$ and 12 of the 13 PSUs in stratum h appear in exactly one dropout group, with the thirteenth never being excluded. Thus $F_h = n_h/[n_h/\ell_h]$ (=13/3 in the example above). The number of replicates formed is $L = \sum_{h=1}^{H} \ell_h$, where $2 \le \ell_h \le n_h$, and the variance estimator is

$$v_{GJ}(\hat{\theta}) = \sum_{h=1}^{H} \frac{(F_h - 1)}{\ell_h} \sum_{i=1}^{\ell_h} (\hat{\theta}_{(ih)} - \hat{\theta})^2 .$$

If (n_h/ℓ_h) is an integer, $F_h = \ell_h$. If this holds for all h

$$v_{GJ}(\hat{\theta}) = \sum_{h=1}^{H} \frac{(\ell_h - 1)}{\ell_h} \sum_{i=1}^{\ell_h} (\hat{\theta}_{(ih)} - \hat{\theta})^2$$

and

$$V(v_{GJ}(\hat{\theta})) = \sum_{h=1}^{H} \frac{W_h^4 \sigma_h^4}{n_h^2} \left[\frac{(\beta_h - 3)}{n_h} + \frac{2}{(\ell_h - 1)} \right] . \quad (1)$$

In general $V(v_{GJ}(\hat{\theta}))$ is not a continuous function of $\underset{\sim}{\ell} = (\ell_1, ..., \ell_H)'$, and thus minimization with respect to $\underset{\sim}{\ell}$ is difficult. However, (1) is tractable for minimization with respect to $\underset{\sim}{\ell}$. As well as giving the value of $V(v_{GJ}(\hat{\theta}))$ when (n_h/ℓ_h) is an integer for all h, (1) approximates $V(v_{GJ}(\hat{\theta}))$ otherwise. Rust (1984, Ch.3) shows that, for the purposes of choosing the best value of $\underset{\sim}{\ell}$, the choice which minimizes (1) will be satisfactory in most applications. Expression (1) is minimized by putting

$$\ell_h = 1 + \frac{(L - H) W_h^2 \sigma_h^2}{n_h \left[\sum_{k=1}^{H} W_k^2 \sigma_k^2 / n_k \right]} \quad (2)$$

with the constraints that $2 \le \ell_h \le n_h$. Rounding these values to an adjacent integer will approximately minimize $V(v_{GJ}(\hat{\theta}))$, and thus maximize the number of degrees of freedom $r(v_{GJ}(\hat{\theta}))$ for fixed L.

If $\beta_h = 3$ (the value for a normal distribution) for all h, then substituting the values of ℓ_h given by (2) gives degrees of freedom $r(v_{GJ,OPT}(\hat{\theta})) = L - H$, provided that these optimum ℓ_h values are such that $2 < \ell_h < n_h$ for every stratum, where $v_{GJ,OPT}$ denotes the form of V_{GJ} obtained by using (2) to assign dropout groups to strata. As Kish (1965, Section 8.6D) argues, for designs with many second stage units selected per PSU, β_h will be close to three in most cases. Thus in order to estimate variances with 30 degrees of freedom, for example, approximately (30 + H) replicates will be required for such designs, and somewhat more if $\ell_h = 2$ or $\ell_h = n_h$ for some h. Note that at least 2H replicates must be formed, since $\ell_h \ge 2$ for all h.

B. The Combined Strata Grouped Jackknife Variance Estimator (v_{CJ})

This is an extension of the grouped jackknife, in which strata are combined for the purpose of forming groups. In this case $G < H$, and replicates are formed by dropping units from more than one stratum at a time. The formation of combined strata, and the assignment of dropout groups ℓ_g $(2 \le \ell_g \le \min_{h \epsilon g} n_h)$ must be done in such a way as to give ℓ_g a value no greater than the largest common factor of the n_h for $h \epsilon g$. Thus if two strata with 7 and 14 PSUs selected respectively are combined, $\ell_g \le 7$. However, two strata with 7 and 13 PSUs selected respectively cannot be combined, and if two strata with 8 and 14 PSUs respectively are combined, the only possible number of dropout groups is $\ell_g = 2$. Each dropout group consists of $[n_h/\ell_g]$ units from stratum h. Thus if strata with 7 and 14 PSUs are combined, and $\ell_g = 3$ dropout groups are formed, each will contain 2 PSUs from the first stratum and 4 PSUs from the second. In this case, $F_g = 3.5$. The total number of replicates is $L = \sum_{g=1}^{G} \ell_g$, and the variance estimator is

$$v_{CJ}(\hat{\theta}) = \sum_{g=1}^{G} \frac{(F_g - 1)}{\ell_g} \sum_{i=1}^{\ell_g} (\hat{\theta}_{(ig)} - \hat{\theta})^2 .$$

If (n_h/ℓ_g) is an integer for all $h \epsilon g$, $g = 1, ..., G$, then

$$v_{CJ}(\hat{\theta}) = \sum_{g=1}^{G} \frac{(\ell_g - 1)}{\ell_g} \sum_{i=1}^{\ell_g} (\hat{\theta}_{(ig)} - \hat{\theta})^2$$

and for unbiased linear $\hat{\theta}$

$$V(v_{CJ}(\hat{\theta})) = \sum_{h=1}^{H} W_h^4 \sigma_h^4 (\beta_h - 3) / n_h^3$$

$$+ 2 \sum_{g=1}^{G} \Big(\sum_{h \epsilon g}^{H} W_h^2 \sigma_h^2 / n_h \Big)^2 / (\ell_g - 1) . \quad (3)$$

As for the grouped jackknife, (3) will be used to approximate $V(v_{CJ}(\hat{\theta}))$ in general, with the error of approximation having little consequence for practical application in forming dropout groups efficiently. $V(v_{CJ}(\hat{\theta}))$ is minimized approximately by using

$$\ell_g = 1 + (L-G) \Big(\sum_{h \epsilon g}^{H} W_h^2 \sigma_h^2 / n_h \Big) \Big(\sum_{k=1}^{H} W_k^2 \sigma_k^2 / n_k \Big)^{-1} \quad (4)$$

for $g = 1,2,...,G$, rounding ℓ_g to an appropriate integer within the range $2 \le \ell_g \le \min_{h \epsilon g} n_h$. With $\beta_h = 3$ for all h, and provided that $2 < \ell_g < \min_{h \epsilon g} n_h$ for all g, $r(v_{CJ}(\hat{\theta})) = L - G$.

The best choice of ℓ_g is thus to assign one dropout group to each combined stratum, and then assign the remainder in proportion to the sampling variance contributed by each combined stratum. If this is possible, approximately (r + G) replicates are required to give r degrees of freedom. Thus fewest replicates are required for a given level of precision when as many strata as possible are combined. However, it must be remembered that there are restrictions as to which strata can be combined, and excessive combining will give $\ell_g = \min_{h \epsilon g} n_h$ being optimal for one or more combined strata g. If this occurs, (r + G) replicates will give fewer than (r + G) degrees of freedom, and in fact it may not be possible to form (r + G) replicates if G is small. In Section 6 it will be seen that a lesser degree of combining can be more precise, because the allocation of dropout groups to combined strata can be undertaken more efficiently.

C. The Sample Jackknife Variance Estimator (v_{SJ})

For this method, no strata are combined, so G = H. As for the full jackknife, each dropout group contains only one unit $(s_h = 1$ for all h), and thus $F_h = n_h$. The number of replicates used is $L = \sum_{h=1}^{H} \ell_h$, with $1 \le \ell_h \le n_h$. Thus L PSUs are omitted from exactly one replicate, and (n − L) are included in every replicate. The variance estimator is

$$v_{SJ}(\hat{\theta}) = \sum_{h=1}^{H} \frac{(n_h - 1)}{\ell_h} \sum_{i=1}^{\ell_h} (\hat{\theta}_{(ih)} - \hat{\theta})^2 .$$

For unbiased linear $\hat{\theta}$

$$V(v_{SJ}(\hat{\theta})) = \sum_{h=1}^{H} \frac{W_h^4 \sigma_h^4}{n_h^2 (n_h - 1)^2 \ell_h^2} \left\{ (\beta_h - 3) \left[(n_h - 2)^2 + \frac{\ell_h}{n_h} (2n_h - 3) \right] + 2 \left[n_h (n_h - 2) + \ell_h \right] \right\} .$$

This is minimized using ℓ_h values given by

$\ell_h = 1$ if $n_h = 2$

$$\ell_h = \frac{(L - H^*) a_h}{\sum_{k=1}^{H} a_k} \quad \text{if } n_h > 2$$

where H^* is the number of strata with $n_h = 2$, and

$$a_h = \frac{W_h^2 \sigma_h^2}{n_h (n_h - 1)} \sqrt{(n_h - 2)} \left((\beta_h - 3)(n_h - 2) + 2n_h \right)^{1/2} ,$$

with the additional condition that $1 \le \ell_h \le n_h$.

Note that for the sample jackknife variance estimator the optimal allocation of dropout units is a function of the stratum kurtoses. This makes the practical application of this method more difficult than is the case for the grouped jackknife methods described above, since in practice β_h values will often be imprecisely known.

If $\beta_h = 3$ for all h, and if (n / H) is large, r replicates will give almost r degrees of freedom in most cases, for r ≥ H.

5. Comparisons Among Jackknife Procedures

In the preceding section, three alternative jackknife procedures are described which permit the user to determine to a great extent the number of replicates to be used. In comparing these the precisions for linear estimators for a given number of replicates L will be considered. The approximate bias of the variance estimators for non-linear estimators will also be considered.

The grouped jackknife is a special case of the combined strata grouped jackknife. Thus in comparing these two, the issue is to what extent the combining of strata is desirable. As seen in Section 4, if optimization of the choice of ℓ_g values is possible, $V(v_{CJ}(\hat{\theta})) = A + B / (L - G)$ where A and B are functions of the design and population, but not the variance estimator, L is the number of replicates used and G is the number of combined strata used. As B > 0, the most precise variance estimation will result from the use of the fewest combined strata (smallest G) consistent with the assignment of ℓ_g values in accordance with (4), rather than having $\ell_g = 2$ or $\ell_g = \min_{h \in g} n_h$. Essentially, one degree of freedom is "lost" for every combined stratum used, so the use of a smaller number of combined strata is desirable.

Furthermore, the efficiency of combining strata will be more extensive for domains (subclasses whose members are contained within a subset of the survey strata). For a fixed number of replicates L, ℓ_g values will be larger for smaller G. This means that each domain will be assigned to a greater number of dropout groups when much combining occurs, leading to greater precision of variance estimation for domain estimates.

Rust (1984, Appendix A) shows that, ignoring terms of order n^{-3} and higher, for $\hat{\theta}$ an appropriately differentiable function of linear estimators,

$$\text{Bias}(v_J(\hat{\theta})) \approx \sum_{g=1}^{G} (F_g - 1)^{-1} \sum_{h \in g} d_h$$

where d_h is a function of the estimator and sample design in stratum h. To this order of approximation the absolute bias will be less for large values of F_g if d_h is of constant sign for all h, which is likely to occur in practice. Large values of F_g arise if few units from any one stratum are included in a single dropout group, ($[n_h | \ell_g]$ small) and thus in general for fixed number of replicates L less absolute bias in variance estimation is likely if strata are combined to a greater degree (large ℓ_g and small G).

The relative precision of the sample jackknife in comparison to the grouped jackknife depends upon the stratum kurtoses β_h. As discussed in Rust (1984, Ch. 4), for β_h values close to three the precisions of the two approaches are similar. However, the grouped jackknife provides some protection against the presence of larger β_h values, in that its precision declines much less rapidly with increasing β_h than does the precision of the sample jackknife. For example consider the case where $W_h^2 \sigma_h^2$, n_h and β_h are constant across strata. Table 1 shows the number of degrees of freedom for the grouped jackknife and sample jackknife, both for the case where the common $\beta_h = 3$, and for $\beta_h = 10$.

TABLE 1
Degrees of Freedom for Jackknife Variance Estimators
n=200 H=10 L=40 n_h=20 $W_h^2 \sigma_h^2$ Constant

Common β_h	Grouped Jackknife (v_{GJ})	Sample Jackknife (v_{SJ})
3	30	40
10	20	9.5

In the case where $\beta_h = 3$, both variance estimators provide adequate precision for use in making inference about θ, and give rise to similar width confidence intervals. When $\beta_h = 10$, the sample jackknife gives relatively poor precision and will give 95% confidence intervals 8% wider than those given by the grouped jackknife. Using 95% confidence intervals of the form $\hat{\theta} \pm 2\sqrt{v_{SJ}(\hat{\theta})}$, based on the sample jackknife, will give an overstated level of confidence if $\beta_h = 10$. For more extreme values of β, the performance of the sample jackknife deteriorates rapidly. Thus the range of β_h values for which the sample jackknife gives adequate precision for given L is much smaller than for the grouped jackknife. In stratified single stage samples for measuring quantitative variables, kurtoses (β_h values) substantially in excess of 3 are quite likely to arise. Thus in practice the grouped jackknife appears to be a more robust method than the sample jackknife, less sensitive to outliers, which are reflected by the presence of large stratum kurtosis values β_h.

In summary, the desirable properties for a jackknife variance estimator with fixed number of replicates are:

1) A high proportion of the PSUs should be included in some dropout group
2) Within a single dropout group there should be few units from any single stratum
3) Dropout groups should be assigned more heavily to strata (or combined strata) which contribute large proportions of the total sampling variance.

Among the alternatives considered in this section the combined strata grouped jackknife, with optimal allocation of dropout groups, most successfully combines these attributes across the range of types of sample design for which replication methods are likely to be used.

6. Example of the Application of the Combined Strata Grouped Jackknife (v_{CJ})

Consider the stratified population and sample shown in Table 2. For all strata $\beta_h = 3$. If the full jackknife variance estimator v_{FJ} were used for variance estimation, 200 replicate estimates would be required, and about 103 degrees of freedom would be attained.

TABLE 2
Hypothetical Stratified Population and Sample

Stratum (h)	$W_h^2\sigma_h^2$	n_h	$W_h^2\sigma_h^2/n_h$
1,2	1	20	.05
3,4	2	20	.1
5,6	5	20	.25
7,8	10	20	.5
9,10	20	20	1
Total	–	200	3.80

By using the combined stratum grouped jackknife, adequate precision of variance estimation can be attained using much fewer than 200 replicates. One possible approach is to combine all strata, and form 20 dropout groups consisting of one PSU from each stratum. This will give 20 replicates and 19 degrees of freedom. The use of this variance estimator will give 95% confidence intervals which are 5.2% wider than those obtained using the full jackknife with 200 replicates.

Greater precision can be obtained if less combining of strata is used (because only 20 replicates can be formed in the case of a single combined stratum). For example, suppose that we wish to attain about 25 degrees of freedom, and that strata 1 to 4 comprise a domain of special interest (this domain has been heavily sampled). We thus would like strata 1 to 4 to be represented in as many combined strata as possible. Table 3 shows one possible way of forming combined strata, and the appropriate rounded allocation of dropout groups to combined strata. In order to attain 25 degrees of freedom with 4 combined strata, 29 dropout groups have been assigned.

For dropout groups in Combined Stratum 1, ten units are omitted each time, five from each of strata 1 and 7. In Combined Stratum 4, six units are omitted each time, two from each of strata 4, 6 and 10. Such an arrangement gives 24.8 degrees of freedom for the whole sample estimate, and

TABLE 3
Formation of Combined Strata and Dropout Groups for Hypothetical Population

Combined Stratum	Contrib. Strata	$\sum_{h\epsilon g} W_h^2\sigma_h^2/n_h$	Optimum ℓ_g	Rounded ℓ_g	s_h
1	1,7	0.55	4.62	4	5
2	2,8	0.55	4.62	5	4
3	3,5,9	1.35	9.88	10	2
4	4,6,10	1.35	9.88	10	2
Total	–	3.80	29	29	–

24.5 degrees of freedom for the domain consisting of strata 1 to 4. Note that almost 25 degrees of freedom are attained in each case, despite substantial rounding in the assignment of ℓ_g values.

7. Balanced Repeated Replication (BRR)

The method of BRR was introduced by McCarthy (1966, 1969) and has been described recently in detail by Wolter (1985). The full method with H strata requires the formation of T half-sample replicate estimates, where T is a multiple of 4 satisfying $H \leq T \leq H + 3$. Depending upon the precise form of variance estimator used, T complementary half-sample replicate estimates may be required also. If $\hat{\theta}'_t$ is the estimate of θ derived from the t th half-sample, one popular form of BRR variance estimator for estimating $V(\hat{\theta})$ is

$$v_{BRR-S}(\hat{\theta}) = \frac{1}{2T}\sum_{t=1}^{T}\left[(\hat{\theta}'_t - \hat{\theta})^2 + (\hat{\theta}''_t - \hat{\theta})^2\right].$$

The discussion below relates to BRR variance estimators of this form. This variance estimator requires 2T replicate estimates. For unbiased linear $\hat{\theta}$, v_{BRR-S} is unbiased, and using the same notation as in the previous sections but with $n_h = 2$ for all h,

$$V(v_{BRR-S}(\hat{\theta})) = \sum_{h=1}^{H}\frac{W_h^4\sigma_h^4(\beta_h + 1)}{8}.$$

Two methods have been proposed for adapting the BRR method for use with fewer than the 2T replicate estimates required for full balance. The method of Partially Balanced Repeated Replication (PBRR), introduced by McCarthy (1966) and developed by Lee (1972, 1973) requires the formation of groups of strata and uses a set of half-sample replicates which are fully balanced within each group, but not across groups. Fewer replicates are required to achieve this partial balancing.

In Lee's development, all groups contain the same number of strata. If H strata are divided into P groups of H/P strata each, instead of requiring about H half-samples and their complements, as for full balance, only about H/P half-samples and their complements are required. Lee discusses strategies for forming such groups so as to maintain good precision of variance estimation for a given number of groups P.

An alternative method of reducing the number of half-samples required for BRR is to form combined strata and balance across combined strata but not within. If G combined strata are formed, this approach results in the formation of about G half-sample replicates and their complements. The number of strata within each combined stratum g, H_g, need not be constant across g.

It can be easily seen (see Rust, 1984, Ch.2) that the methods of PBRR and combined strata BRR are equivalent. However, the combined strata method provides a more convenient approach to handling the case where the number of strata per combined stratum varies (or equivalently for PBRR, the number of strata per balance group varies).

When G combined strata are formed, the gth of which contains H_g strata, the combined strata BRR variance estimator uses T' half-sample estimators and their complements, where $G \leq T' \leq G + 3$. Denoting the variance estimator as v_{CB}, it can be shown that

$$V(v_{CB}(\hat{\theta})) = \frac{1}{2}\left[\sum_{h=1}^{H} W_h^4\sigma_h^4(\beta_h-3)/4 + \sum_{g=1}^{G}\Big(\sum_{h\epsilon g} W_h^2\sigma_h^2\Big)^2\right].$$

If the values of $W_h^2\sigma_h^2$ are known approximately, then $V(v_{CB}(\hat{\theta}))$ can be minimized for fixed G by equalizing the values of $\sum_{h\epsilon g} W_h^2\sigma_h^2$. Thus ideally the G combined strata should each contribute equally to the sampling variance of $\hat{\theta}$. The strategies proposed by Lee (1972, 1973) for obtaining precise variance estimation using PBRR are effectively aimed at achieving this equalization. If $\sum_{h\epsilon g} W_h^2\sigma_h^2$ is constant, then

$$V(v_{CB}(\hat{\theta})) = \frac{1}{2}\left[\sum_{h=1}^{H}\frac{W_h^4\sigma_h^4(\beta_h - 3)}{4} + \frac{1}{G}\Big(\sum_{h=1}^{H} W_h^2\sigma_h^2\Big)^2\right]$$

and if $\beta_h = 3$ for all H, $v_{CB}(\hat{\theta})$ has G degrees of freedom for variance estimation.

Thus, with knowledge of the approximate values of β_h, and the relative values of $W_h^2\sigma_h^2$, the number of replicates required to achieve a prescribed number of degrees of freedom can be calculated. When $\beta_h = 3$ for all h, a total of approximately r half-sample replicates and their complements, obtained from r combined strata, will suffice to give about r degrees of freedom. The combined strata jackknife also requires 2r replicates in this case. The most efficient formation of combined strata is the same for these two replicated approaches. For linear estimators the jackknife and BRR variance estimators are equivalent for 2 PSU per stratum designs if the same set of combined strata is used for each.

8. Practical Application of the Methods

For both the BRR variance estimator, and those jackknife variance estimators (including the grouped jackknife and the combined stratum grouped jackknife) for which $[n_h/\ell_g] = s_h$ for all h,g, so that no "sampling" occurs in the allocation of PSUs to dropout groups, an alternative form of variance estimator is available. In the analyses above, the full complement of replicates formed have been used for variance estimation. For each variance estimator, an alternative form of variance estimator is available, equivalent for linear estimators to the full form, described in Sections 4 and 7. For this alternative form, G of the replicates are ignored, corresponding to one dropout group from each (combined) stratum. The variance estimation formula is adjusted appropriately, so as to give unbiased variance estimation for unbiased linear estimators. In the case of BRR, this variance estimator, based only on the half-sample estimates and ignoring the complementary half-sample replicates, is denoted by Kish and Frankel(1974) and others as v_{BRR-H}. By using such "adjusted" replicated variance estimators, G fewer replicate estimates are required, but there is little loss in precision of variance estimation. Thus these forms of replicated variance estimator are attractive in practice when limitation of the number of replicate estimates derived is desirable. The results given above hold for these adjusted variance estimators, equally as for those for which estimates from all replicates formed are used. For the jackknife variance estimators each combined stratum provides $(\ell_g - 1)$ dropout groups, and approximately r such replicate estimates are required to give r degrees of freedom. Similarly for BRR, approximately r degrees of freedom can be obtained by using r combined strata, efficiently formed, and deriving a single half-sample estimate for each replicate, requiring about r replicate estimates in total.

The development in this paper has assumed independent selection of PSUs within strata, with at least two PSUs selected per stratum. In most designs the former assumption is violated, and in many multistage designs only one PSU is selected per stratum, so that strata must be collapsed for replicated variance estimation. Thus the use of replicated procedures leads frequently to biased variance estimation in practice. For linear estimators the magnitude of this bias is not affected by the form of replicated procedure used, and for all types of estimators is generally both small and positive. It is thus not of major concern for the methods considered in this paper.

For both jackknife and BRR procedures, exact optimization requires knowledge of the relative values of $W_h^2\sigma_h^2 / n_h$, the contribution of sampling variance from each stratum. This will not generally be available, but exact optimization is not generally required. Provided that there is some indication available as to the relative values of $W_h^2\sigma_h^2$ and β_h, adequate precision of variance estimation can be obtained using a moderate number of replicates. The optimum will differ for different survey variables, but approximate optimization for one major survey variable, or a compromise for a few main variables, is likely to be good for most. For non-linear estimators, if the component linear estimators have optima which are similar, their non-linear composite is likely to have an optimum close to these in many cases.

Frequently, variation in $W_h^2\sigma_h^2$ will be dominated by variation in W_h^2, and thus the size of the strata can be used as a proxy for $W_h\sigma_h$. This approach should be used with care for designs where some strata have more stages of selection than others, as σ_h^2 values are likely to be greater for the strata with more stages. Note that for designs with constant σ_h and $n_h = 2$, the recommended approach to combining strata for use with BRR does not involve the formation of combined strata of equal size $\left(\sum_{h \epsilon g} W_h\right)$ as might be expected intuitively. Many small strata can be combined safely, but large strata should be kept distinct.

9. Summary

The aim of this paper is to demonstrate an approach to replicated variance estimation in which the formation of a modest number of replicates will give adequate precision of variance estimation in a wide range of applications. There are a number of features of this approach important for its application in practice. 1) It can be applied to a wide range of sample designs, regardless of the allocation of sample to strata, or other features of the complex design. 2) As long as the researcher has a general idea of the relative contribution of the different survey strata to the total sampling variance, and an awareness of the likelihood of occurrence and severity of long-tailed sampling distributions within strata, it is possible to control adequately the precision of variance estimation, and simultaneously to limit the amount of replication. 3) The procedure is robust, in that it is only necessary to form replicates in a manner somewhat near optimum in order to achieve close to optimum precision. 4) Although different sets of replicates can be formed for different estimators if necessary, in many cases a single set of replicates will be adequate for a wide range of estimators from a single survey. If particular domains (subsets of the survey strata) of interest are identified in advance, a single set of replicates can be formed in a manner which ensures adequate precision of variance estimation for such domains, as well as the whole population.

9. References

Kish, L. (1965). *Survey Sampling.* New York:Wiley.

Kish, L. and Frankel, M.R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society*, Ser B 36, 1-37.

Lee, K-H. (1972). Partially balanced designs for half sample replication method of variance estimation. *Journal of the American Statistical Association* 67, 324-334.

Lee, K-H. (1973). Using partially balanced designs for the half sample replication method of variance estimation. *Journal of the American Statistical Association* 68, 612-614.

Lemeshow, S. (1979). The use of unique statistical weights for estimating variances with the balanced half-sample technique. *Journal of Statistical Planning and Inference* 3, 315-323.

McCarthy, P.J. (1966). *Replication. An Approach to the Analysis of Data from Complex Surveys.* Vital and Health Statistics, Series 2, No. 14. U.S. Dept. of Health, Education and Welfare. Washington:U.S. Government Printing office.

McCarthy, P.J. (1969). Pseudo-replication: half-samples. *International Statistical Review* 37, 239-264.

Rust, K.F. (1984). *Techniques for Estimating Variances for Sample Surveys.* Ph.D. thesis, University of Michigan.

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