Journal of Official Statistics, Vol. 24, No. 3, 2008, pp. 469–488

Weight Adjustments for the Grouped Jackknife Variance Estimator

Richard Valliant[1]*, J. Michael Brick*[2]*, and Jill A. Dever*[3]

The jackknife variance estimator is often implemented by dropping groups of units rather than a single unit at a time. This has the practical advantages of economizing on computation time and file size because a separate weight is appended to the analysis file for each jackknife replicate. If the replicate weight adjustments and the grouped jackknife itself are not appropriately constructed, the variance estimates can have some extremely pathological behavior when estimating totals. When the dropout groups do not all have exactly the same number of first-stage units, the standard version of the grouped jackknife may be a severe overestimate. This problem is most likely to arise in single-stage samples with a large number of first-stage units in many of the strata. The standard grouped jackknife variance estimator and two alternatives are examined for the situation of unequally sized groups through a simulation study of school districts in the 50 United States and the District of Columbia.

Key words: Grouping PSUs; grouping strata; inconsistent variance estimator; replicate weights.

1. Introduction

Replication variance estimation is a standard tool of survey statisticians and researchers. Replication is popular and useful because it provides a simple means of estimating variances without requiring the derivation of explicit variance formulas that are often complicated. Shao (1996), Shao and Tu (1995), and Wolter (2007) review the methods as they apply to finite population estimation and give supporting theory. Rust and Rao (1996), surveyed the uses of replication discussed several practical applications. The general idea is to divide the full sample into subsamples or replicates, compute the basic estimate for the full sample and for each replicate, and then combine these estimates with a simple variance formula.

[1] University of Maryland, Joint Program in Survey Methodology, 1218 Lefrak Hall, College Park, MD 20742, U.S.A. Email: rvalliant@survey.umd.edu
[2] Westat and Joint Program in Survey Methodology, 1650 Research Boulevard, Rockville, MD 20850-3195, U.S.A. Email: mikebrick@westat.com
[3] University of Maryland, Joint Program in Survey Methodology, 1218 Lefrak Hall, College Park, MD 20742, U.S.A. Email: jdever@survey.umd.edu
Acknowledgments: The work of the first author was supported by grant SES-0416662 from the National Science Foundation (NSF). The authors would like to acknowledge the National Center for Education Statistics (NCES), Institute for Education Sciences, for funding part of this research, and in particular the support of the Project Officer, Marilyn Seastrom. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or NCES. We also are grateful for some very helpful comments by the Associate Editor and the referees.

There is a substantial amount of theory available for the replication methods when they are implemented in standard ways. However, in practice these methods are operationalized in ways that often do not fit the standard theoretical requirements. For the jackknife, in particular, the basic approach is to delete one first-stage or primary sample unit (PSU), compute an estimate based on the remaining sample units, cycle through the remaining PSUs, and compute a variance among the resulting set of estimates. In practice, groups of units are formed by combining units within or across strata. Entire groups are then dropped out in order to compute a jackknife variance estimate. Dropping out groups rather than individual units saves on computation and reduces the number of replicate weights on the analysis file.

Theory for the grouped methods exists, but is largely focused on multi-stage samples with a small number of PSUs selected from each stratum. For example, Lu, Brick, and Sitter (2006) report on grouping methods that are most relevant when two PSUs are sampled in each stratum. However, they do not discuss grouping for single-stage samples with many units selected per stratum that are very common in establishment surveys. The theory of grouping is limited in single-stage designs, and guidance for practitioners is almost nonexistent. As a result, some methods used in practice may not have good theoretical properties, as described below.

Sections 2 through 4 briefly review how the jackknife is implemented in practice, how replicate weights are computed, and how PSUs are often grouped to reduce computational burden. Section 5 gives some general theory for when the grouped jackknife is biased and introduces two alternatives to the standard method. In Section 6, we illustrate the performance of the standard grouped jackknife and the alternatives in a simulation study. Technical details of the bias calculations and theory supporting the alternative estimators are in the Appendices. Section 7 summarizes our findings.

2. Delete-one Jackknife and Its Implementation for Survey Data

The basic implementation of the jackknife is known as the delete-one version and is briefly described in this section. Suppose a stratified probability sample of n_h out of N_h PSUs is selected from stratum h ($h = 1, \ldots, H$). Define $\hat{\theta}$ as the estimate based on the full sample, i.e., the estimate in which each sample unit in each sample PSU is weighted by the inverse of the probability of selection or base weight, and $\hat{\theta}_{(hi)}$ as the "replicate" estimate based on omitting PSU i in stratum h. In computing $\hat{\theta}_{(hi)}$, the base weights for the remaining PSUs in stratum h are multiplied by the factor $n_h/(n_h - 1)$ to compensate for dropping PSU (hi). The standard delete-one jackknife variance estimator is calculated from the replicate estimates based on omitting one sample PSU at a time and cycling through all sample PSUs:

$$v_J(\hat{\theta}) = \sum_h \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left[\hat{\theta}_{(hi)} - \hat{\theta}\right]^2 \tag{1}$$

(e.g., Krewski and Rao 1981, Expression 2.4; Wolter 2007, Section 4.5). The estimator $\hat{\theta}$ can be a linear function of the sample data or a smooth (i.e., differentiable) nonlinear function such as a ratio. There are other versions of the jackknife, including one that uses deviations centered on the average replicate estimate, that are addressed in Appendix B. A finite population correction factor (*fpc*) is sometimes added to give $v_J(\hat{\theta}) = \sum_h (1 - f_h)(n_h - 1)n_h^{-1} \sum_{i=1}^{n_h} \left[\hat{\theta}_{(hi)} - \hat{\theta} \right]^2$, where $f_h = n_h/N_h$, the fraction of PSUs in stratum h that are in the sample. If a single-stage sample is selected by stratified simple random sampling and the estimate is linear, then this *fpc* is exact; otherwise its addition is only an approximate way of reflecting a nonnegligible sampling fraction. In many establishment surveys, large sampling fractions are common in strata containing a limited number of sizeable units. In such cases, accounting for nonnegligible *fpc*'s is an important part of variance estimation.

3. Replicate Weights

The usual approach in creating a data file is to include the full sample weight and the replicate weights for each completed case. Software packages like WesVar® (Westat 2000), Stata® (StataCorp 2005) and SUDAAN® (Research Triangle Institute 2001) use the replicate weights to compute $\hat{\theta}_{(hi)}$ and $v_J(\hat{\theta})$. As discussed in Rust and Rao (1996), appending replicate weights to an analysis file has some distinct advantages:

(i) The data analyst's task is simplified because the analyst does not need to know the details of the sample design used to collect the data.
(ii) The data file creators can incorporate into the weights features of estimation such as nonresponse adjustment and poststratification.
(iii) Omitting strata and PSU identifiers can help limit the risk of disclosing the identities of respondents.

If there are $n = \sum_{h=1}^{H} n_h$ PSUs and the delete-one jackknife is used, each record on the data file will have $n + 1$ weights. Descriptive and model-based analyses are then replicated $n + 1$ times–once for the full sample and once for each replicate. Jackknife variance estimates are computed by substituting the results in (1).

In many single-stage samples, there may be thousands of PSUs. Consequently, the standard delete-one jackknife would require that thousands of weights be appended to each record. The number of data fields can easily be exceeded by the number of weights and the size of the data file as measured in megabytes can quickly become unmanageable. Some complex analyses, such as logistic regression, require iterative calculations to obtain estimates of model parameters. For such cases, the iteration must be performed separately for the full sample and for each replicate. Computing time can be unacceptably long when these iterative computations are done hundreds or thousands of times.

One approach to reducing calculation time and data file size is to delete groups of PSUs rather than individual units, as discussed in Amrhein, Hicks, and Kott (1997), Duchesne (2000), Kott (1998, 1999, 2001), Rust (1985, 1986), Rust and Kalton (1987), and Wolter (2007, Chapter 4). Although this is an old technique, there is limited theoretical guidance on how the grouping should be done, and there are wide variations in what is done in real applications. Rust (1984, 1986) discusses the effect of grouping on the variance of a variance estimate, which is related to its approximate degrees of freedom; he does not discuss the potential for bias caused by poor methods of grouping or weight adjustment.

Methods of forming replication groups in this case include the following:

(i) Form groups of PSUs within design strata, or
(ii) Combine design strata into superstrata and form groups of PSUs within a superstratum that cut across design strata.

A common assumption in the literature is that equal-sized groups of PSUs are formed, i.e., the groups each contain an equal number of PSUs (see e.g., Duchesne 2000 or Wolter 2007). DiGaetano, Brick, and Flores-Cervantes (1998) describe an application of method (ii) above where the goal is to retain at least 30 degrees of freedom for full population estimates; however, their approach is not directly applicable to single-stage samples with many PSUs per stratum.

4. Standard Grouped Jackknife

The standard implementation of the grouped jackknife variance estimator, analogous to (1), is

$$v_{GJ1}(\hat{\theta}) = \sum_{\tilde{h}=1}^{\tilde{H}} \frac{G_{\tilde{h}} - 1}{G_{\tilde{h}}} \sum_{g=1}^{G_{\tilde{h}}} \left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right]^2 \tag{2}$$

where $\tilde{h}$ denotes a superstratum, $\tilde{H}$ is the number of superstrata, g is a group of PSUs, and $G_{\tilde{h}}$ is the total number of groups in superstratum $\tilde{h}$. Note that summing over all groups in $\tilde{h}$ implicitly sums over all design strata in that superstratum. The replicate estimate $\hat{\theta}_{(\tilde{h}g)}$ is computed from the sample of PSUs and subunits after omitting all units assigned to Group g within superstratum $\tilde{h}$. In cases where the sampling fraction of first-stage units is similar among the strata in a superstratum, an ad hoc *fpc* of the form $1 - f_{\tilde{h}}$ is sometimes inserted in (2), similar to inserting an *fpc* in (1).

Handy shorthand terms for superstratum and group are VarStrat and VarUnit, which is also the jargon used by WesVar®. A VarStrat is either an original design stratum or a combination of design strata. A VarUnit within a particular VarStrat is either a group formed from sample units within a design stratum, or a combination of units from different design strata.

In computing $\hat{\theta}_{(hi)}$ for the delete-one jackknife (1), the base weights associated with all units retained for the replicate within design stratum h are multiplied by the factor $n_h/(n_h - 1)$ to compensate for the dropping of PSU (hi). By analogy, in the grouped jackknife the usual procedure for computing $\hat{\theta}_{(\tilde{h}g)}$ is to multiply each full sample weight within VarStrat $\tilde{h}$ by $G_{\tilde{h}}/(G_{\tilde{h}} - 1)$ since one group is dropped to calculate the replicate estimate. That is, the same weight adjustment is made regardless of which group is dropped. The appropriateness of this adjustment is critically dependent on the groups having the same number of PSUs, as described in the next section.

5. A More General Grouped Jackknife

In this discussion and for the theory derived in the Appendices, the PSUs are assumed to be selected with replacement (WR). If multistage sampling is used, the sampling within PSUs can be done using any method as long as a design-unbiased estimate of each PSU total can be constructed. Although restrictive, the WR assumption is commonly used to simplify theoretical calculations (e.g., see Krewski and Rao 1981) and reasonably describes empirical results in many cases where sampling is actually done without replacement but n_h/N_h is small. As a practical fix-up, we will insert an *fpc* in variance formulae to approximately account for nonnegligible sampling fractions. We consider linear estimators of the population total with the form $\hat{\theta} = \sum_h \hat{y}_h$ where $\hat{y}_h = n_h^{-1} \sum_{i=1}^{n_h} y_{hi}/p_{hi}$, y_{hi} is the estimated total for units in PSU (hi), p_{hi} is the single-draw selection probability of PSU (hi), and n_h is the number of sampled PSUs from Stratum h. Note that we can also write $\hat{y}_h = \sum_{i=1}^{n_h} w_{hi} y_{hi}$ with $w_{hi} = (n_h p_{hi})^{-1}$ so that $\hat{\theta}$ can be written in the usual form as a weighted sum of data. In single-stage samples, y_{hi} is the value for unit (hi). Thus, $\hat{y}_h$ is a "pwr" estimator as discussed in Särndal, Swensson, and Wretman (1992, Section 2.9). Suppose that $S_{\tilde{h}}$ is the set of design strata grouped together into VarStrat $\tilde{h}$, $S_{\tilde{h}h}$ is the set of VarUnits assigned to VarStrat $\tilde{h}$ and design stratum h, and $S_{\tilde{h}hg}$ is the set of PSUs in VarStrat $\tilde{h}$ and design stratum h assigned to VarUnit g. We do not require that the PSUs in h be split among all VarUnits in $\tilde{h}$. Thus, $S_{\tilde{h}hg}$ could be empty, i.e., Group g may contain no PSUs from design Stratum h. The set $S_{\tilde{h}hg}$ cannot contain all of the PSUs from any design stratum h within VarStrat $\tilde{h}$. If it did, then dropping the associated VarUnit would drop the sample from an entire design stratum. A group of PSUs can cross design strata or be contained within a single design stratum.

To fix the idea, Figure 1 illustrates some possibilities for grouping. In this example there are three design strata, which have six, four, and four sample PSUs, respectively. Two VarStrat are formed with VarStrat 1 containing design stratum 1 and VarStrat 2 containing design strata 2 and 3. VarStrat 1 illustrates that different PSUs can be combined within a design stratum in order to create a VarUnit. In VarStrat 1, three VarUnits are created by pairing consecutive PSUs. (The sample PSUs should be randomly ordered within a design stratum to avoid biases that would be created by, for example, consistently pairing the smallest PSUs in a design stratum.) VarStrat 2 illustrates that different design strata can be combined. Two VarUnits are formed in VarStrat 2. The first VarUnit consists of PSUs 1 and 2 from design stratum 2 and PSUs 1 and 2 from design stratum 3. The second VarUnit contains the PSUs numbered 3 and 4 in each of design strata 2 and 3. Thus, when VarUnit 1 in VarStrat 2 is deleted for the grouped jackknife, PSUs from both design strata 2 and 3

VarStrat $\tilde{h}$	Design stratum h	PSU i	VarUnit g
1	1	1	1
		2	1
		3	2
		4	2
		5	3
		6	3
2	2	1	1
		2	1
		3	2
		4	2
	3	1	1
		2	1
		3	2
		4	2

Fig. 1. An illustration of grouping PSUs and design strata

would be dropped. A pitfall to avoid would be creating a VarUnit that would result in deleting an entire design stratum. For instance, if we assigned PSUs 1–4 from design stratum 2 to VarUnit 1 in VarStrat 2, this would result in the full sample from design stratum 2 being dropped for the grouped jackknife. The remaining PSUs could not be used to make a design-unbiased, full population estimate in such a case.

When VarStrat and VarUnits are used, the estimated total can be written as

$$\hat{\theta} = \sum_{\tilde{h}} \sum_{h \in S_{\tilde{h}}} \sum_{g \in S_{\tilde{h}h}} n_h^{-1} \sum_{i \in S_{\tilde{h}hg}} \hat{y}_{hi} \tag{3}$$

with $\hat{y}_{hi} = y_{hi}/p_{hi}$ and y_{hi} being an unbiased estimator of the total for PSU (hi). Suppose that $E(\hat{y}_{hi} - Y_h)^2 = \sigma_h^2$, i.e., the contribution to variance of each 1-PSU estimate, $\hat{y}_{hi}$, of the stratum total, Y_h, is the same. For example, in a stratified, single-stage simple random sample, $\hat{y}_{hi} = N_h y_{hi}$ and $\sigma_h^2 = N_h^2 V_h^2$ where V_h^2 is the stratum h population variance of the y_{hi}'s. With this notation, the variance of the estimated total is

$$\mathrm{var}(\hat{\theta}) = \sum_{\tilde{h}} \sum_{h \in S_{\tilde{h}}} \frac{\sigma_h^2}{n_h} \tag{4}$$

The general form of the grouped jackknife that we consider is

$$v_{GJ}(\hat{\theta}) = \sum_{\tilde{h}} (1 - f_{\tilde{h}}) \sum_{g=1}^{G_{\tilde{h}}} K_{(\tilde{h}g)} \left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} \right]^2 \tag{5}$$

where $K_{(\tilde{h}g)}$ is a constant that depends on VarStrat and the deleted VarUnit. To construct the estimator $\hat{\theta}_{(\tilde{h}g)}$, full-sample weights are used for units in VarStrat other than $\tilde{h}$. The weights for units retained in stratum h within VarStrat $\tilde{h}$, when VarUnit $(\tilde{h}g)$ is omitted, are adjusted by a factor, $a_{h(\tilde{h}g)}$. Weights for units in VarUnit $(\tilde{h}g)$ are set to zero. More specifically, the weights used to construct $\hat{\theta}_{(\tilde{h}g)}$ are

$$w_{hi(\tilde{h}g)} = \begin{cases} w_{hi} & (hi) \notin S_{\tilde{h}} \\ a_{h(\tilde{h}g)} w_{hi} & (hi) \in S_{\tilde{h}}, (hi) \notin S_{\tilde{h}hg} \\ 0 & (hi) \in S_{\tilde{h}hg} \end{cases} \tag{6}$$

Theory for v_{GJ} in (5) and for some special cases is developed in Appendix A; the details are summarized here. When $f_{\tilde{h}}$ is negligible, the expectation of (5) is

$$E\left[v_{GJ}(\hat{\theta})\right] = \sum_{\tilde{h}} \sum_{g=1}^{G_{\tilde{h}}} K_{(\tilde{h}g)} \left\{ V_{(\tilde{h}g)} + B_{(\tilde{h}g)}^2 \right\} \tag{7}$$

where $V_{(\tilde{h}g)} = \sum_{h \in S_{\tilde{h}}} \sigma_h^2 D_{h(\tilde{h}g)}/n_h$ and

$$B_{(\tilde{h}g)} = \sum_{h \in S_{\tilde{h}}} Y_h \left[a_{h(\tilde{h}g)} (1 - p_{hg}) - 1 \right] \tag{8}$$

with $D_{h(\tilde{h}g)} = a_{h(\tilde{h}g)}^2 (1 - p_{hg}) - 2a_{h(\tilde{h}g)} (1 - p_{hg}) + 1$, $p_{hg} = n_{hg}/n_h$, the proportion of PSUs in stratum h assigned to group g, and n_{hg} is the number of PSUs in stratum

h assigned to group g. The term $\sum_{\tilde{h}} \sum_{g=1}^{G_{\tilde{h}}} K_{(\tilde{h}g)} V_{(\tilde{h}g)}$ will equal var($\hat{\theta}$) in (4) if

$$\sum_g K_{(\tilde{h}g)} D_{h(\tilde{h}g)} = 1 \tag{9}$$

for each design stratum h in VarStrat $\tilde{h}$. The term involving $B^2_{(\tilde{h}g)}$ is a type of squared bias since $B_{(\tilde{h}g)} = E\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right]$ and, when (9) holds, must be zero for v_{GJ} to be unbiased.

The combination, $a_{h(\tilde{h}g)} = G_{\tilde{h}}/(G_{\tilde{h}} - 1)$ with $K_{(\tilde{h}g)} = (G_{\tilde{h}} - 1)/G_{\tilde{h}}$, gives the standard grouped jackknife v_{GJ1} in (2). As shown in Appendix A, this estimator is unbiased if the PSUs in design stratum h are equally divided among the groups, in which case every group (VarUnit) in VarStrat $\tilde{h}$ has the same number of PSUs. But, v_{GJ1} can be extremely biased if the groups vary in size. In fact, $N^{-2} \sum_{\tilde{h},g} K_{(\tilde{h}g)} B^2_{(\tilde{h}g)}$ will be $O(1)$ as long as $p_{hg} \neq 1/G_{\tilde{h}}$ so that, on average, the squared bias will dominate v_{GJ} in large samples. As noted in Appendix A, this problem will be most apparent for variables with small values of the population coefficient of variation.

The size of the bias of v_{GJ1} also depends, in an interesting way, on the sizes of groups that are formed. For illustration, consider a single design stratum with n PSUs in which G VarUnits are formed. Suppose that n/G has remainder r and that two sizes of groups are formed:

r groups of size $\lceil n/G \rceil$ where $\lceil x \rceil$ denotes the integer ceiling of x, and
$G - r$ groups of size $\lfloor n/G \rfloor$ where $\lfloor x \rfloor$ denotes the integer floor of x.

These numbers of groups would result if units were put into some order and then numbered serially from 1 to G, cycling back to 1 as needed. Then the squared-bias term in the expectation of v_{GJ} is

$$\begin{aligned}\sum_{g=1}^{G} K_{(g)} B^2_{(g)} \propto \frac{G-1}{G} \Bigg\{ & r\left[\left(\frac{G}{G-1}\right)\left(1 - \frac{\lfloor n/G \rfloor}{n}\right) - 1\right]^2 \\ & + (G-r)\left[\left(\frac{G}{G-1}\right)\left(1 - \frac{\lfloor n/G \rfloor}{n}\right) - 1\right]^2 \Bigg\}\end{aligned} \tag{10}$$

The proportionality in (10) is due to Y_h's being factored out in (8).

Expression (10) is graphed versus G in Figure 2 for $n = 539$, which is the number of PSUs in VarStrat 1 in the simulation study reported in Section 6. The pattern for other values of n is similar. The size of the squared-bias term varies dramatically, depending on G. Choices of G that are near each other, like 90 and 100, can have very different bias characteristics. For example, with $G = 100$, the squared bias term (10) has a local maximum of about 82.7×10^{-5} but with $G = 90$, the squared bias is only 3.4×10^{-5}. When $G = 100$, 61 groups of Size 5 and 39 groups of Size 6 are formed; when $G = 90$, 1 group of Size 5 and 89 groups of Size 6 are formed. When either r or $G - r$ is small (i.e., min $(r, G - r)$ is small), most groups have the same size and the squared-bias term is small. With larger values of min $(r, G - r)$, (10) is larger and the bias of v_{GJ1} will be larger. As the figure makes clear, there are some "lucky Gs" where the bias of v_{GJ1} will be much less noticeable.

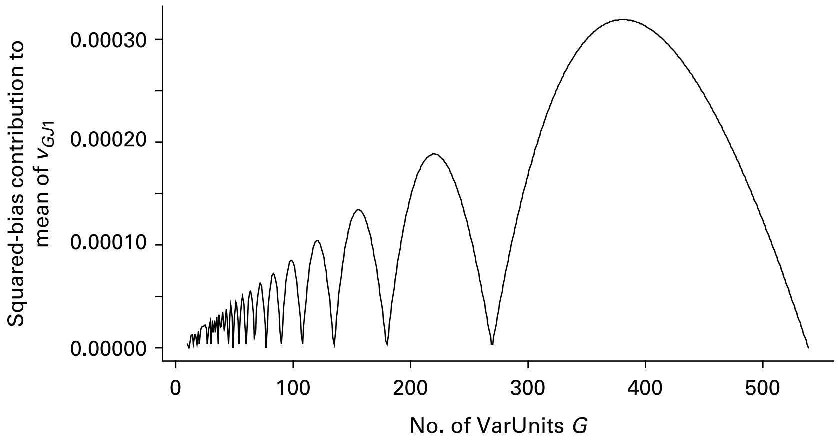

Fig. 2. Dependence of squared-bias term in $E(v_{GJ1})$ on the number of groups $G(n = 539)$

In Appendix B, we show that if v_{GJ1} is defined by centering on the VarStrat mean, $\hat{\theta}_{(\tilde{h})} = \sum_{g=1}^{G_{\tilde{h}}} \hat{\theta}_{(\tilde{h}g)}/G_{\tilde{h}}$, it will also be an overestimate unless all VarUnits have the same number of PSUs. Appendix B also sketches properties for nonlinear estimators. In the important special case of estimating the variance of a ratio, v_{GJ1} will be approximately unbiased, even if the VarUnits are not of equal size. Thus, the problems with the standard grouped jackknife estimator will surface mainly when estimating totals.

A better choice for constructing (5) is $a_{h(\tilde{h}g)} = n_{\tilde{h}}/(n_{\tilde{h}} - n_{\tilde{h}g})$ with $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ where $n_{\tilde{h}g}$ is the number of PSUs in VarStrat $\tilde{h}$ assigned to VarUnit g and $n_{\tilde{h}} = \sum_{g=1}^{G_{\tilde{h}}} n_{\tilde{h}g}$ is the total number of PSUs in the VarStrat. The weight adjustment $a_{h(\tilde{h}g)}$ depends on the number of PSUs retained, i.e., $n_{\tilde{h}} - n_{\tilde{h}g}$, after VarUnit $(\tilde{h}g)$ is dropped. This combination of $a_{h(\tilde{h}g)}$ and $K_{(\tilde{h}g)}$ gives an estimator,

$$v_{GJ2}(\hat{\theta}) = \sum_{\tilde{h}} (1 - f_{\tilde{h}}) \sum_{g=1}^{G_{\tilde{h}}} \frac{(n_{\tilde{h}} - n_{\tilde{h}g})}{n_{\tilde{h}}} \left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} \right]^2$$

that is approximately unbiased even when the VarUnits do not all have the same number of PSUs (see Appendix A, Case 2).

A third approach is to use the combination $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$ with $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ as for v_{GJ2}. This estimator has the same form as v_{GJ2} above but with a different weight adjustment, $a_{h(\tilde{h}g)}$, and will be denoted by v_{GJ3}. Notice that the weight adjustment is specific to a design stratum even when these are grouped into a single VarStrat. The choice, v_{GJ3}, is also approximately unbiased when the VarUnits have differing numbers of PSUs (see Appendix A, Case 2). However, the weight adjustments $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$ can be more variable than the $a_{h(\tilde{h}g)} = n_{\tilde{h}}/(n_{\tilde{h}} - n_{\tilde{h}g})$ adjustments in v_{GJ2}, especially when stratum sample sizes, n_h, are not substantially larger than the n_{hg}'s.

The delete-a-group jackknife (DAGJK) studied by Kott (2001) uses the same weight adjustment as v_{GJ3}. For the DAGJK only one VarStrat is used. PSUs are randomly ordered within a design stratum and the full sample is systematically divided into G groups. When random group g is dropped, the weights for the retained units in stratum h are adjusted by

$a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$. The DAGJK, when $n_h \geq G$ for all strata, is defined as

$$v_{DAGJK}(\hat{\theta}) = \frac{G-1}{G} \sum_{g=1}^{G} \left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} \right]^2 \tag{11}$$

Kott (2001) also gives a modification for the case where $n_h < G$ in some strata. If only one VarStrat is used and $n_{hg} = n_h/G$ in all design strata, then $n_g = \sum_h n_{hg} = n/G$ and $K_{(\tilde{h}g)}$ in v_{GJ3} reduces to $(G-1)/G$. Thus, the DAGJK can be viewed as a special case of v_{GJ3} when all sampling fractions are small. When the stratum *fpc*'s are not negligible, v_{DAGJK} may be too conservative. A single average *fpc* could be used in the DAGJK, but if *fpc*'s vary among strata, a more fine-tuned adjustment is desirable, especially for domain estimates that cover subsets of design strata.

Another alternative for without replacement sampling is to use the same adjustment, $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$, as for v_{GJ3} and v_{DAGJK}, but to multiply the base weight for a unit from stratum h by $\sqrt{1-f_h}$. For example, in stratified simple random sampling without replacement the modified full sample weight would be $\sqrt{1-f_h}N_h/n_h$. Define the full sample estimate using the modified base weights as $\hat{\theta}^*$ and the replicate estimate using the modified base weights and the replicate adjustments to be $\hat{\theta}^*_{(\tilde{h}g)}$. The variance of $\hat{\theta}$ is then estimated by

$$v_{GJ4}(\hat{\theta}) = \sum_{g=1}^{G} K_{(\tilde{h}g)} \left[\hat{\theta}^*_{(\tilde{h}g)} - \hat{\theta}^* \right]^2$$

with $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$. More details are given in Appendix A, Case 4. Note that a variation of DAGJK could be created that incorporates *fpc*'s by using $\hat{\theta}^*_{(\tilde{h}g)} - \hat{\theta}^*$ in its construction instead of $\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}$. The estimator, v_{GJ4}, is essentially the same as v_{GJ3} and will not be pursued further in this manuscript.

6. A Simulation Study

To illustrate the differences in performance of the various choices of grouped jackknife, we conducted a simulation using a population of 11,941 school districts in the 50 United States and the District of Columbia – a population also used in Brick et al. (2005). The districts were categorized into 12 design strata based on size (four categories based on the number of students) crossed with percentage of students at or below the poverty line (three categories). A sample of 1,020 school districts was allocated to the strata as shown in Table 1.

One rationale for forming VarStrat is to group strata with similar *fpc*'s and use the grouped jackknife formula (2) with an *fpc* of $1 - f_{\tilde{h}}$. Since the sampling fractions in Table 1 vary substantially among the strata from a low of 0.051 to a high of 0.245, three VarStrat were formed consisting of strata 1–6, 7–9, and 10–12. The similarity of the *fpc*'s is the dominant factor in forming the superstrata, which contrasts with the approach of Lu, Brick, and Sitter (2006) in the case of a multi-stage sample with relatively few sampled PSUs. The total numbers of sample units in the superstrata were 539, 195, and 286, respectively (Table 1).

Table 1. Population distribution and sample allocation for a population of school districts stratified by size of student body (4) and poverty status (3)

Stratum	Population size (N_h)	Sample size (n_h)	Sampling rate (f_h)	Superstratum (VarStrat)	Population size ($N_{\tilde{h}}$)	Sample size ($n_{\tilde{h}}$)
1	615	32	0.052			
2	1,147	59	0.051			
3	1,292	66	0.051			
4	1,720	111	0.065	1	8,972	539
5	2,305	149	0.065			
6	1,893	122	0.064			
7	692	75	0.108			
8	579	63	0.109	2	1,798	195
9	527	57	0.108			
10	342	83	0.243			
11	449	110	0.245	3	1,171	286
12	380	93	0.245			
Total	11,941	1,020			11,941	1,020

Within each VarStrat, $G_{\tilde{h}} = G$ VarUnits were formed using the following procedure: (*i*) order the design strata; (*ii*) randomly sort the sample units within design strata; and (*iii*) serially number the sample units from 1 to G, cycling back to 1 as needed. Six choices of G were used ($G = 20$, 25, 45, 50, 135, and 150) giving total numbers of groups (VarStrat $\times$ G VarUnits) of 60, 75, 135, 150, 405, and 450, respectively. Note that for $G \geq 50$, there are several strata in Table 1 where $n_h < G$. Based on the analysis that led to Figure 2, the values of G were selected to illustrate the different sizes of bias that can occur for nearby values of G. For example, it was anticipated that $G = 25$ would be more biased than $G = 20$; $G = 50$ would be more biased than $G = 45$; and $G = 150$ would be more biased than $G = 135$. The full delete-one jackknife requires 1,020 replicates. In this example, grouping seems to be a reasonable procedure for reducing the amount of computation needed for the jackknife while retaining the ability to include the *fpc*s.

For each of the six values of G, we selected 5,000 stratified simple random samples without replacement of size 1,020 with the allocation in Table 1 and estimated the totals for nine variables using the stratified expansion estimator $\hat{\theta} = \sum_{h=1}^{H} N_h \bar{Y}_{hs}$ where $\bar{Y}_{hs}$ is the sample mean of the variable in Stratum h. The nine variables consisted of three items from the school district file–the number of full-time equivalent administrators (NumAdm), whether a district had a student-teacher ratio larger than 15 (StuTch), and whether a district had 15 or fewer administrators (SmallAdm)–and six artificial variables. These included three chi-square random variables with degrees of freedom 2, 30, and 60, denoted by $X^2(2)$, $X^2(30)$, and $X^2(60)$, and three Bernoulli random variables with parameters $p = 0.5$, 0.95, and 0.995, denoted by Bin(0.5), Bin(0.95), and Bin(0.995). Each coefficient of variation (*CV*) shown in the header of Table 2 is defined as the population standard deviation divided by the population mean. The artificial variables allow us to illustrate the theoretical finding that the size of the *CV* plays a role in the performance of v_{GJ1}. Number of administrators

Table 2. Ratios of average grouped jackknife variance estimates to empirical mean squared errors over 5,000 stratified simple random samples for estimated totals

Total groups	Groups per VarStrat (G)	No. of admins $CV = 1.85$	Student-teacher ratio > 15 $CV = 1.168$	15 or fewer admins $CV = 0.028$	X^2 (2) $CV = 1.0$	X^2 (30) $CV = 0.258$	X^2 (60) $CV = 0.183$	Bin (0.5) $CV = 1.0$	Bin (0.95) $CV = 0.229$	Bin (0.995) $CV = 0.071$
v_{GJ1}[a]										
60	20	1.04	0.99	1.64	1.01	1.07	1.11	1.01	1.11	2.37
75	25	1.01	1.02	3.10	1.03	1.23	1.40	1.03	1.26	5.32
135	45	1.06	1.04	2.23	0.98	1.16	1.28	0.99	1.15	3.76
150	50	1.00	1.02	3.46	0.98	1.23	1.52	0.97	1.32	6.03
405	135	1.06	1.03	3.88	1.01	1.31	1.61	1.04	1.41	7.25
450	150	1.10	1.06	11.29	1.06	2.06	3.03	1.05	2.35	22.59
v_{GJ2}										
60	20	1.03	0.99	1.03	1.01	1.01	0.99	1.00	1.03	1.05
75	25	0.99	1.01	1.05	1.02	1.02	1.00	1.01	1.00	1.10
135	45	1.03	1.03	1.17	0.97	1.04	1.06	0.98	1.01	1.35
150	50	0.98	1.00	1.09	0.96	0.99	1.04	0.95	1.02	1.15
405	135	0.99	1.01	1.24	1.00	1.04	1.06	1.02	1.05	1.47
450	150	1.01	1.00	1.12	0.99	1.01	1.03	0.98	1.05	1.25
v_{GJ3}										
60	20	0.99	1.00	1.00	0.99	1.01	1.02	0.99	0.94	0.99
75	25	1.04	0.98	1.00	0.98	0.99	1.03	0.98	1.03	1.01
135	45	1.02	1.03	1.01	1.00	1.04	1.03	1.00	1.00	1.01
150	50	0.97	1.01	1.00	1.00	1.03	1.01	1.02	1.00	1.01
405	135	1.02	0.98	1.00	1.01	0.98	1.00	1.01	1.01	1.01
450	150	0.97	1.03	1.00	1.00	0.99	1.02	1.02	0.99	1.01
v_{DAGJK}										
60	20	1.27	1.11	1.29	1.10	1.13	1.14	1.10	1.05	1.11
75	25	1.34	1.08	1.28	1.08	1.10	1.14	1.09	1.14	1.12
135	45	1.30	1.12	1.27	1.09	1.13	1.12	1.09	1.10	1.11
150	50	1.23	1.10	1.26	1.09	1.12	1.10	1.11	1.09	1.10
405	135	1.28	1.06	1.26	1.09	1.06	1.08	1.09	1.09	1.09
450	150	1.22	1.12	1.25	1.08	1.08	1.10	1.10	1.08	1.10

[a] The t-tests on the differences in the mean values of v_{GJ1} between all pairs of G values are significantly different from zero with p-values all less than 0.0001 for the variables SmallAdm, $X^2(30)$, $X^2(60)$, Bin(0.95), and Bin(0.995).

and student-teacher ratio have *CVs* larger than 1 while "15 or fewer administrators" has a *CV* of 0.028. We expect the bias of v_{GJ1} to be obvious for the variables with small *CVs*.

The average of the grouped jackknife variance estimates and the empirical mean squared error (MSE) of the estimated totals were computed across the 5,000 samples. (The estimated totals are unbiased in these simulations so that the MSE and the empirical variance are virtually equal.) Table 2 presents the ratios of the average variance estimate to the MSE for the different combinations of variables and numbers of groups. A ratio of 1.0 indicates that the jackknife variance estimate is an unbiased estimate of the MSE. The standard grouped jackknife v_{GJ1} performs reasonably well for the "number of administrators" variable which has a population *CV* of 1.85 and for StuTch with a *CV* of 1.168, as well as for $X^2(2)$, and Bin(0.5) which have population *CVs* of 1. The ratios for these four variables range from 0.97 to 1.10, with no consistent evidence of bias.

In contrast, the grouped jackknife has obvious positive biases for SmallAdm, $X^2(30)$, $X^2(60)$, Bin(0.95), and Bin(0.995). These are the variables with small population *CVs* of 0.028, 0.258, 0.183, 0.229, and 0.071, respectively. The reason for the poor performance of v_{GJ1} in these cases is that the single overall weight adjustment for each replicate is too crude, leading to a nonzero $B_{(\tilde{h}g)}$ term. For example, take the case of 25 VarUnits per VarStrat or 75 VarUnits in all (i.e., the second row in Table 2). In VarStrat 1 with 539 sample districts (Table 1), 14 VarUnits contain 22 PSUs while the remaining 11 VarUnits contain 21 units. The replicate weight adjustment for all 25 groups is $G_{\tilde{1}}/(G_{\tilde{1}} - 1) = 25/24 = 1.04167$. However, more appropriate adjustments based on the number of units actually retained for each replicate are $539/(539 - 22) \cong 1.04255$ for the 14 groups of size 22 and $539/(539 - 21) \cong 1.04054$ for the 1 group of 21. This apparently trivial difference is enough to lead to overestimates of 210% for SmallAdm, 40% for $X^2(60)$, and 432% for Bin(0.995) when $G = 25$.

The consequences for v_{GJ1} of using different values of G, as discussed in Section 5, are apparent in Table 2 even though having three VarStrat with different numbers of PSUs does dilute the effect somewhat. For example, for Bin(0.995) the ratios for $G = 20$ and 25 are 2.37 and 5.32; ratios for $G = 135$ and 150 are 7.25 and 22.59. For the variables with the smaller *CVs*–SmallAdm, $X^2(30)$, $X^2(60)$, Bin(0.95), and Bin(0.995)–the *t*-tests on the differences in the mean values of v_{GJ1} between all pairs of G values are significantly different from zero with *p*-values all less than 0.0001.

The jackknife alternatives that adjust the weights depending on numbers of retained PSUs, v_{GJ2} and v_{GJ3}, essentially eliminate the bias as seen in the second and third banks of ratios in Table 2. The exception is v_{GJ2} for Bin(0.995) where the largest bias is still 47%. The delete-a-group jackknife, v_{DAGJK}, is an overestimate because it omits finite population corrections. Overall, the best choice is v_{GJ3}, which adjusts weights using $n_h/(n_h - n_{hg})$ at the design stratum level and uses $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ at the superstratum level.

Table 3 presents results for estimated ratio means using v_{GJ1}, v_{GJ3}, and v_{DAGJK}. For three of the variables in Table 2, the mean per school district was estimated for the domain of districts whose schools included grades kindergarten through 12. A domain mean was used so that both the numerator and denominator of the ratio were random. All three versions of the jackknife are essentially unbiased for each variable and choice of G, confirming the

Table 3. Ratios of average grouped jackknife variance estimates to empirical mean squared errors over 5,000 stratified simple random samples for estimated means or proportions per district for districts whose schools include grades kindergarten through 12

Total groups	Groups per VarStrat (G)	15 or fewer admins $CV = 0.028$	$X^2(60)$ $CV = 1.83$	Bin(0.995) $CV = 0.071$
v_{GJ1}				
75	25	1.01	0.99	1.00
150	50	1.00	1.01	1.01
450	150	1.02	1.00	1.05
v_{GJ3}				
75	25	1.01	1.01	0.98
150	50	1.00	1.04	1.02
450	150	1.01	1.03	1.02
v_{DAGJK}				
75	25	1.29	1.12	1.09
150	50	1.26	1.13	1.11
450	150	1.27	1.12	1.10

Appendix B theoretical observation that there should be no bias problem for estimating the variance of a ratio.

Coverages of 95% normal theory confidence intervals of the form $\hat{\theta} \pm 1.96\sqrt{v}$ were also tabulated in each set of 5,000 samples using v_{GJ1}, v_{GJ2}, v_{GJ3} and v_{DAGJK} to construct the intervals. To conserve space, we only summarize the results. Over-estimation by v_{GJ1} is generally accompanied by over-coverage. For example, the coverage rates for totals of Bin(0.95) using v_{GJ1} were 97, 98, and 100% for $G = 50, 135,$ and 150. For Bin(0.995) 100% of all confidence intervals using v_{GJ1} cover the population total for all values of G. Coverage rates for totals using v_{GJ2} range from 93 to 96% for eight of the nine variables regardless of the value of G. The exception is Bin(0.995), where the range is 87 to 94%. For the ratio means, corresponding to the results in Table 3, confidence interval coverages ranged from 94 to 96% for all three of the variance estimators and for all variables except Bin(0.995). For Bin(0.995) coverage rates were 85 to 86%. Normal theory intervals are known to perform poorly for extreme proportions so that the under-coverage for this variable is not surprising.

7. Conclusion

The standard procedure in the grouped jackknife of adjusting weights by the factor $G/(G-1)$, where G is the number of groups, leads to biased estimates for variances of totals if the groups are not all the same size. Even a size difference of one first-stage unit in the groups can cause severe overestimates of the variance. This can be a subtle problem because it affects only variables in which the population standard deviation is small compared to the mean (i.e., small coefficient of variation). For variables with population coefficients of variation of, say, 1.0 or larger the problem may not manifest itself. Also, if $\min(r, G-r)$ is small, where r is the remainder in n/G, most groups will have the same size, and the bias will be much less than when $\min(r, G-r)$ is larger. Among statistics commonly computed in surveys, the problem of over-estimation does appear to be isolated to totals. The standard grouped jackknife does not over-estimate the variance of ratio means, for example.

Two simple alternatives were considered to adjust the jackknife based on the numbers of units retained in each replicate. Both of these, denoted v_{GJ2} and v_{GJ3}, were shown theoretically and empirically to essentially eliminate the biases shown to exist in the traditional method. The key feature of v_{GJ2} is a separate set of weight adjustments for each jackknife replicate. The adjustments in a replicate depend on the number of first-stage units retained in each stratum. This approach is similar to making replicate-specific adjustments to reflect the effect on variances of weight adjustments for ineligible units, nonresponding units, and poststratification (e.g., see Yung and Rao 2000). Both alternatives do reduce to the standard version of the grouped jackknife, v_{GJ1}, if all groups in a stratum or combined stratum have the same number of primary sampling units. A delete-a-group jackknife from Kott (2001) is also a substantial improvement over the standard grouped jackknife, but does not incorporate finite population correction factors that vary among strata, unlike v_{GJ2} and v_{GJ3}. However, the DAGJK can be modified to include stratum-level *fpc*'s using the method described for the estimator v_{GJ4}.

Although the theory for the modified estimators, v_{GJ2} and v_{GJ3}, was developed for multistage designs, these variance estimators will be most useful in single-stage designs where a fairly large number of units are selected per stratum. Establishment surveys, in particular, where groups of units are formed for jackknife variance estimation, should consider v_{GJ2} or v_{GJ3}.

Appendix A. Properties of the Grouped Jackknife Estimator

Notation used in this appendix is defined in the main body of the article. As in Section 5, assume that PSUs are selected with replacement and the population total is estimated by $\hat{\theta} = \sum_h \hat{y}_h$ with $\hat{y}_h = n_h^{-1} \sum_{i=1}^{n_h} y_{hi}/p_{hi}$. Assume that, as $n = \sum_h n_h \to \infty$ and $N = \sum_h N_h \to \infty$, the following conditions hold:

(i) $\max_h \frac{N_h}{N} \frac{n}{n_h} = O(1)$

(ii) $\sum_{h \in S_{\tilde{h}}} Y_h = O(N)$

(iii) $\sum_{h \in S_{\tilde{h}}} \frac{\sigma_h^2}{n_h} = O(N^2/n)$

(iv) $K_{(\tilde{h}g)}$, $\tilde{H}$, and $G_{\tilde{h}}$ are all bounded

Condition (i) covers the case of H being bounded and n_h/n and N_h/N converging to constants; or, $H \to \infty$ with N_h/n_h bounded and N/H and n/H converging to constants. Thus, both the case of a fixed number of strata and a large number of PSUs per stratum and the case of a large number of strata and a limited number of PSUs per stratum are covered. Since the number of VarStrat, $\tilde{H}$, is assumed to be bounded in (iv), the number of PSUs in a VarStrat will be large regardless of whether H is bounded or $H \to \infty$.

The expectation of the grouped jackknife in (5) is

$$E\left[v_{GJ}(\hat{\theta})\right] = \sum_{\tilde{h}}(1 - f_{\tilde{h}})\sum_{g=1}^{G_{\tilde{h}}} K_{(\tilde{h}g)}\left\{V_{(\tilde{h}g)} + B^2_{(\tilde{h}g)}\right\} \tag{A.1}$$

where $V_{(\tilde{h}g)} = \text{var}\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right]$ and $B^2_{(\tilde{h}g)} = \left[E(\hat{\theta}_{(\tilde{h}g)} - \hat{\theta})\right]^2$. To evaluate this expectation, first note that the estimated total when group $(\tilde{h}g)$ is deleted is

$$\hat{\theta}_{(\tilde{h}g)} = \sum_{\tilde{h}' \neq \tilde{h}}\sum_{h \in S_{\tilde{h}'}} \hat{y}_h + \sum_{h \in S_{\tilde{h}}}\sum_{\substack{g' \neq g \\ g' \in S_{\tilde{h}h}}} n_h^{-1}\sum_{i \in S_{\tilde{h}hg'}} a_{hi(\tilde{h}g)} y_{hi} \tag{A.2}$$

The term $a_{hi(\tilde{h}g)}$ is the weight adjustment applied to all units in VarStrat $\tilde{h}$, stratum h, and PSU i when VarUnit $(\tilde{h}g)$ is omitted. If there are no PSUs assigned to VarUnit g within design stratum h, i.e., $S_{\tilde{h}hg}$ is empty, then $a_{hi(\tilde{h}g)} = 1$. The full sample PSU weight is $w_{hi} = (n_h p_{hi})^{-1}$, and the adjusted PSU weight, after deleting PSUs in group $(\tilde{h}g)$, is $w_{hi(\tilde{h}g)} = a_{hi(\tilde{h}g)} w_{hi}$ for $i \in S_{\tilde{h}hg'}$. How $a_{hi(\tilde{h}g)}$ and $K_{(\tilde{h}g)}$ are computed determines whether the grouped jackknife is biased or not.

Using (3) and (A.2), the difference between the replicate and the full sample estimates is

$$\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} = \sum_{h \in S_{\tilde{h}}}\left[n_h^{-1}\sum_{\substack{g' \neq g \\ g' \in S_{\tilde{h}h}}}\sum_{i \in S_{\tilde{h}hg'}} a_{hi(\tilde{h}g)}\hat{y}_{hi} - \hat{y}_h\right] \tag{A.3}$$

We consider three cases for the constant $a_{hi(\tilde{h}g)}$:

(1) $G_{\tilde{h}}/(G_{\tilde{h}} - 1)$, the standard choice for weight adjustment when one group is deleted;
(2) $n_{\tilde{h}}/(n_{\tilde{h}} - n_{\tilde{h}g})$ where $n_{\tilde{h}g}$ = number of PSUs in VarStrat $\tilde{h}$ assigned to VarUnit g and $n_{\tilde{h}} = \sum_{g=1}^{G_{\tilde{h}}} n_{\tilde{h}g}$, the total number of PSUs in the VarStrat; and
(3) $n_h/(n_h - n_{hg})$ where n_{hg} = number of PSUs in design stratum h assigned to group g.

Since none of these depend on the particular PSU i, we can simplify the notation for the weight adjustment to be $a_{hi(\tilde{h}g)} \equiv a_{h(\tilde{h}g)}$. Using (A.3) and $\sum_{\substack{g' \neq g \\ g' \in S_{\tilde{h}h}}}\sum_{i \in S_{\tilde{h}hg'}} \hat{y}_{hi} = n_h\hat{y}_h - n_{hg}\hat{y}_{hg}$ with $\hat{y}_{hg} = n_{hg}^{-1}\sum_{i \in S_{\tilde{h}hg}} \hat{y}_{hi}$, we have

$$\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} = \sum_{h \in S_{\tilde{h}}}\left[\left\{a_{h(\tilde{h}g)} - 1\right\}\hat{y}_h - a_{h(\tilde{h}g)} p_{hg}\hat{y}_{hg}\right] \tag{A.4}$$

where $p_{hg} = n_{hg}/n_h$. The expectation of this difference is

$$B_{(\tilde{h}g)} = \sum_{h \in S_{\tilde{h}}} Y_h\left[a_{h(\tilde{h}g)}\left(1 - p_{hg}\right) - 1\right] \tag{A.5}$$

with Y_h being the population total in stratum h. In order for the grouped jackknife to be unbiased, (A.5) must be 0. A sufficient condition for this is $a_{h(\tilde{h}g)}(1 - p_{hg}) = 1$, which implies that $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$, i.e., Case (3) above.

Next, we can use $\text{var}(\hat{y}_h) = \text{cov}(\hat{y}_h, \hat{y}_{hg}) = \sigma_h^2/n_h$ and $\text{var}(\hat{y}_{hg}) = \sigma_h^2/n_{hg}$ to show that

$$V_{(\tilde{h}g)} = \text{var}\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right] = \sum_{h \in S_{\tilde{h}}} \frac{\sigma_h^2}{n_h} D_{h(\tilde{h}g)} \tag{A.6}$$

where $D_{h(\tilde{h}g)} = a_{h(\tilde{h}g)}^2(1 - p_{hg}) - 2a_{h(\tilde{h}g)}(1 - p_{hg}) + 1$. Comparing (4) with (A.1) and neglecting the *fpc,* $1 - f_{\tilde{h}}$, the grouped jackknife will be unbiased if, in addition to (A.5) equaling 0, we have

$$\sum_g K_{(\tilde{h}g)} \sum_{h \in S_{\tilde{h}}} \frac{\sigma_h^2}{n_h} D_{h(\tilde{h}g)} = \sum_{h \in S_{\tilde{h}}} \frac{\sigma_h^2}{n_h} \tag{A.7}$$

Equality is obtained when $\sum_g K_{(\tilde{h}g)} D_{h(\tilde{h}g)} = 1$ for every design stratum h in VarStrat $\tilde{h}$, but this is not possible in general when the group sizes are unequal. Expressions (A.5) and (A.6) are specialized below for the three cases of $a_{h(\tilde{h}g)}$ listed above and some potential values for $K_{(\tilde{h}g)}$ are examined.

Case 1. Standard Replicate Weight Adjustment Based on Deleting One Group: v_{GJ1}

The standard procedure is to adjust all weights in VarStrat $\tilde{h}$ to account for deletion of one group. In particular, $a_{h(\tilde{h}g)} = G_{\tilde{h}}/(G_{\tilde{h}} - 1)$ for each unit in the PSUs that remain after deleting VarUnit $(\tilde{h}g)$. In that case, $a_{h(\tilde{h}g)}(1 - p_{hg})$ in (A.5) reduces to $G_{\tilde{h}}(G_{\tilde{h}} - 1)^{-1}(1 - p_{hg})$. If PSUs in design stratum h are equally divided among the groups so that $p_{hg} = 1/G_{\tilde{h}}$, a situation that is not true in general, then (A.5) reduces to zero. Note that $p_{hg} = 1/G_{\tilde{h}}$ implies that $n_{hg} = n_h/G_{\tilde{h}}$ so that $n_{\tilde{h}g} = \sum_{h \in S_{\tilde{h}}} n_{hg} = \sum_{h \in S_{\tilde{h}}} n_h/G_{\tilde{h}} = n_{\tilde{h}}/G_{\tilde{h}}$. In other words, every group (VarUnit) in VarStrat $\tilde{h}$ has the same number of PSUs. We also have $D_{h(\tilde{h}g)} = \left[1 + G_{\tilde{h}} p_{hg}(G_{\tilde{h}} - 2)\right]/(G_{\tilde{h}} - 1)^2$, which reduces to $1/(G_{\tilde{h}} - 1)$ when $p_{hg} = 1/G_{\tilde{h}}$. In the case of equal size groups, we can set $K_{(\tilde{h}g)} = (G_{\tilde{h}} - 1)/G_{\tilde{h}}$ and $v_{GJ1}(\hat{\theta})$ will be unbiased since $\sum_g K_{(\tilde{h}g)} D_{h(\tilde{h}g)} = 1$.

However, when the n_{hg} vary, v_{GJ1} can have a substantial bias. Under assumptions (i) – (iv), given earlier, $D_{h(\tilde{h}g)} = O(1)$ and $N^{-2}V_{(\tilde{h}g)} = O(n^{-1})$, implying that the first term in $E(v_{GJ1})/N^2$ will be $O(n^{-1})$. Conditions (ii) and (iv) imply that the second term in $E(v_{GJ1})/N^2, N^{-2}\sum_{\tilde{h},g} K_{(\tilde{h}g)} B_{(\tilde{h}g)}^2$, will be $O(1)$ as long as $p_{hg} \neq 1/G_{\tilde{h}}$ so that (A.5) is not 0. As a result, the terms $B_{(\tilde{h}g)}^2$ will dominate the expected value of v_{GJ1}, causing it to be an overestimate. Judging from (A.5) and (A.6), the overestimation will be particularly acute if σ_h^2 is small compared to Y_h^2 in all or most design strata. For example, in stratified simple random sampling (*stsrs*), $\sigma_h^2/Y_h^2 = N_h^2 V_h^2/Y_h^2 = V_h^2/\bar{Y}_h^2$ with V_h^2 being the population variance in stratum h. Thus, the positive bias of v_{GJ1} in *stsrs* should be most apparent when the stratum population coefficients of variation of Y are small. One situation where this can occur is when Y is an indicator for a prevalent characteristic.

Case 2. Weight Adjustment Based on Number of PSUs in Deleted Group: v_{GJ2}

When $a_{h(\tilde{h}g)} = n_{\tilde{h}}/(n_{\tilde{h}} - n_{\tilde{h}g})$, we have $a_{h(\tilde{h}g)}(1 - p_{hg}) = (1 - p_{hg})/(1 - p_{\tilde{h}g})$ with $p_{\tilde{h}g} = n_{\tilde{h}g}/n_{\tilde{h}}$. If n_{hg} is the same for all groups within each design stratum so that

$p_{\tilde{h}g} = p_{hg} = 1/G_{\tilde{h}}$, the expectation in (A.5) is zero. If $p_{\tilde{h}g}$ and p_{hg} are close to equal, (A.5) will still be numerically small. We also have

$$D_{h(\tilde{h}g)} = \frac{1 - p_{hg}}{(1 - p_{\tilde{h}g})^2} - 2\frac{1 - p_{hg}}{1 - p_{\tilde{h}g}} + 1$$

which reduces to $1/(G_{\tilde{h}} - 1)$ when $p_{\tilde{h}g} = p_{hg} = 1/G_{\tilde{h}}$. In that case, we can set $K_{(\tilde{h}g)} = (G_{\tilde{h}} - 1)/G_{\tilde{h}}$ and $v_{GJ}(\hat{\theta})$ reduces to the standard estimator in Case 1. Somewhat more generally, if $p_{\tilde{h}g} \cong p_{hg}$, then setting $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ gives an approximately unbiased estimator:

$$v_{GJ2}(\hat{\theta}) = \sum_{\tilde{h}} (1 - f_{\tilde{h}}) \sum_{g=1}^{G_{\tilde{h}}} \frac{(n_{\tilde{h}} - n_{\tilde{h}g})}{n_{\tilde{h}}} \left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right]^2$$

The unbiasedness follows when $(1 - p_{hg})/(1 - p_{\tilde{h}g}) \cong 1$ so that (A.5) is approximately zero and because $D_{h(\tilde{h}g)} = n_{\tilde{h}g}/(n_{\tilde{h}} - n_{\tilde{h}g})$ from which it follows that

$$\sum_g K_{(\tilde{h}g)} D_{h(\tilde{h}g)} = \sum_g \frac{(n_{\tilde{h}} - n_{\tilde{h}g})}{n_{\tilde{h}}} \frac{n_{\tilde{h}g}}{n_{\tilde{h}} - n_{\tilde{h}g}} = \sum_g p_{\tilde{h}g} = 1$$

Case 3. Weight Adjustment Specific to Design Stratum and Group: v_{GJ3}

In the case of v_{GJ3}, $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$ and $B_{(\tilde{h}g)}$ in (A.5) is always zero. If $p_{\tilde{h}g} \cong p_{hg}$, then setting $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ gives an approximately unbiased estimator for the reasons cited above for v_{GJ2}.

Case 4. Weight Adjustment Specific to Design Stratum and Group and Incorporating fpc*:* v_{GJ4}

The weight adjustment for v_{GJ4} is $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$ but base weights are modified by multiplying the weight for a unit in design stratum h by $\sqrt{1 - f_h}$. Define $\hat{\theta}^* = \sum_h \sqrt{1 - f_h}\, n_h^{-1} \sum_{i=1}^{n_h} y_{hi}/p_{hi}$ and $\hat{\theta}^*_{(\tilde{h}g)}$ to be the replicate estimate in (A.2) using $y^*_{hi} = \sqrt{1 - f_h} y_{hi}$ in place of y_{hi}. Then

$$\hat{\theta}^*_{(\tilde{h}g)} - \hat{\theta}^* = \sum_h \sqrt{1 - f_h} \frac{p_{hg}}{1 - p_{hg}} (\hat{y}_h - \hat{y}_{hg}) \quad \text{(A.8)}$$

Consequently, $E\left[\hat{\theta}^*_{(\tilde{h}g)} - \hat{\theta}^*\right]^2 = \sum_{h \in S_{\tilde{h}}} (1 - f_h) \sigma_h^2 n_{hg}/[n_h(n_h - n_{hg})]$. The condition for unbiasedness is $\sum_g K_{(\tilde{h}g)} n_{hg}/(n_h - n_{hg}) = 1$ for each stratum h in VarStrat $\tilde{h}$. As for v_{GJ2} and v_{GJ3}, setting $K_{(\tilde{h}g)} = (n_{\tilde{h}} - n_{\tilde{h}g})/n_{\tilde{h}}$ gives an approximately unbiased estimator as long as $p_{\tilde{h}g} \cong p_{hg}$. Substituting $a_{h(\tilde{h}g)} = n_h/(n_h - n_{hg})$ into (A.4), shows that $\hat{\theta}_{(\tilde{h}g)} - \hat{\theta} = \sum_h p_{hg}(\hat{y}_h - \hat{y}_{hg})/(1 - p_{hg})$ for v_{GJ3}. Comparing this to (A.8) shows that v_{GJ4} will be approximately the same as v_{GJ3} when the *fpc*'s are similar for all design strata within VarStrat $\tilde{h}$.

Appendix B. Mean Centering and Nonlinear Estimators

An alternative jackknife is found by centering around the stratum mean of $\hat{\theta}_{(\tilde{h}g)}$:

$$v_{GJ-M}(\hat{\theta}) = \sum_{\tilde{h}}(1 - f_{\tilde{h}})\sum_{g=1}^{G_{\tilde{h}}} K_{(\tilde{h}g)}\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}_{(\tilde{h})}\right]^2$$

where $\hat{\theta}_{(\tilde{h})} = \sum_{g=1}^{G_{\tilde{h}}}\hat{\theta}_{(\tilde{h}g)}/G_{\tilde{h}}$ with $\hat{\theta}_{(\tilde{h}g)}$ defined in (A.2). In the particular case of $a_{h(\tilde{h}g)} = G_{\tilde{h}}/(G_{\tilde{h}} - 1)$ and $K_{(\tilde{h}g)} = 1/a_{h(\tilde{h}g)}$, $v_{GJ-M}(\hat{\theta}) = v_{GJ1}$, which is easily verified. After some computation, it can be shown that

$$B_{(\tilde{h}g)} \equiv E\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}_{(\tilde{h})}\right] = \sum_{h \in S_{\tilde{h}}} Y_h\left[a_{h(\tilde{h}g)}(1 - p_{hg}) - G_{\tilde{h}}^{-1}\sum\nolimits_g^{G_{\tilde{h}}} a_{h(\tilde{h}g)}(1 - p_{hg})\right] \quad \text{(B.1)}$$

If the groups are all the same size in VarStrat $\tilde{h}$, i.e., $p_{hg} = 1/G_{\tilde{h}}$, then (B.1) reduces to 0. Otherwise, if the term in brackets on the right-hand side of (B.1) is $O(1)$, then $B^2_{(\tilde{h}g)}$ is also $O(1)$, and the bias squared term of $E\left[v_{GJ-M}(\hat{\theta})\right]$ will be dominant, leading to the same overestimation as for v_{GJ1}. We anticipate that the same result will hold for the other versions of the jackknife studied by Krewski and Rao (1981).

Results can also be derived for nonlinear estimators of the form $\hat{\theta} = g(\hat{y}_1, \ldots, \hat{y}_p)$ where g is a differentiable function and $\hat{y}_k = \sum_h \hat{y}_{kh} \quad (k = 1, \ldots, p)$ is a linear estimator of the form examined in Appendix A. The grouped jackknife, v_{GJ1}, is sometimes, but not always, subject to the same biases as in the linear case. We will only sketch results for a ratio, $\hat{\theta} = \hat{y}_2/\hat{y}_1$, where the bias should be negligible. This estimator is commonly used and includes the case of a mean in which the denominator is a sample sum of weights. The usual linear approximation gives $\hat{\theta} - \theta \cong \sum_h \hat{z}_h$ where $\hat{z}_h = n_h^{-1}\sum_{i=1}^{n_h} z_{hi}/p_{hi}$, $z_{hi} = (y_{2hi} - \theta y_{1hi})/Y_1$, y_{1hi} and y_{2hi} are estimated totals of the two variables in PSU hi, $\theta = Y_2/Y_1$, and Y_1 and Y_2 are the population totals for the denominator and numerator variables in the ratio. Using the same reasoning as in Appendix A,

$$B_{(\tilde{h}g)} = E\left[\hat{\theta}_{(\tilde{h}g)} - \hat{\theta}\right] = \sum_{h \in S_{\tilde{h}}} Z_h\left[a_{h(\tilde{h}g)}(1 - p_{hg}) - 1\right] \quad \text{(B.2)}$$

where $Z_h = (Y_{2h} - \theta Y_{1h})/Y_1$ with Y_{1h} and Y_{2h} being the stratum population totals. Expression (B.2) will be zero if $a_{h(\tilde{h}g)}(1 - p_{hg}) = 1$ or $Z_h = 0$ for all strata. The former condition is the same as for linear estimators while the latter holds only if the ratio, Y_{2h}/Y_{1h}, equals θ in every stratum. However, note that $Z_h = Y_{1h}(\theta_h - \theta)/Y_1$ with $\theta_h = Y_{2h}/Y_{1h}$. If θ_h and Y_{1h}/Y_1 are bounded, then $B_{(\tilde{h}g)} = O(1)$ and the second term in $E(v_{GJ1})/N^2$, $N^{-2}\sum_{\tilde{h},g} K_{(\tilde{h}g)}B^2_{(\tilde{h}g)}$, is $O(N^{-1})$. Similar analysis to that in Appendix A shows that $N^{-2}V_{(\tilde{h}g)} = O(n^{-1})$. As long as $\sum_g K_{(\tilde{h}g)}D_{h(\tilde{h}g)} \doteq 1$, as in Case 2 of Appendix A with $K_{(\tilde{h}g)} = (G_{\tilde{h}} - 1)/G_{\tilde{h}}$ and $p_{hg} = 1/G_{\tilde{h}}$, then (A.7) will be approximately satisfied. Thus, $E(v_{GJ1})/N^2 = O(n^{-1}) + O(N^{-1})$ with the $O(n^{-1})$ term being approximately equal to the variance of the ratio. Consequently, the standard grouped jackknife is expected to be nearly unbiased for the variance of a ratio when the sampling fraction, n/N, is small.

8. References

Amrhein, J., Hicks, S., and Kott, P.S. (1997). An Application of a Two-phase Ratio Estimator and the Delete-a-group Jackknife. Proceedings of the American Statistical Association, Section on Survey Research Methods, 175–179.

Brick, J.M., Jones, M.E., Kalton, G., and Valliant, R. (2005). Variance Estimation with Hot Deck Imputation: A Simulation Study of Three Methods. Survey Methodology, 31, 151–159.

DiGaetano, R., Brick, J.M., and Flores-Cervantes, I. (1998). Preserving Degrees of Freedom in a Multi-mode, Multi-site Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 475–480.

Duchesne, P. (2000). A Note on Jackknife Variance Estimation for the General Regression Estimator. Journal of Official Statistics, 16, 133–138.

Kott, P.S. (1998). Using the Delete-a-group Jackknife Variance Estimator in Practice. Proceedings of the American Statistical Association, Section on Survey Research Methods, 763–768.

Kott, P.S. (1999). Some Problems and Solutions with a Delete-a-group Jackknife. Federal Committee on Statistical Methodology Research Conference, 4, 129–135.

Kott, P.S. (2001). The Delete-a-group Jackknife. Journal of Official Statistics, 17, 521–526.

Krewski, D. and Rao, J.N.K. (1981). Inference from Stratified Samples: Properties of the Linearization, Jackknife, and Balanced Repeated Replication Methods. Annals of Statistics, 9, 1010–1019.

Lu, W., Brick, J.M., and Sitter, R. (2006). Algorithms for Constructing Combined Strata Grouped Jackknife and Balanced Repeated Replications with Domains. Journal of the American Statistical Association, 101, 1680–1692.

Research Triangle Institute (2001). SUDAAN User's Manual, Release 8.0. Research Triangle Park, NC: Research Triangle Institute.

Rust, K. (1984). Techniques for Estimating Variances for Sampling Surveys. Unpublished Ph.D. dissertation, Ann Arbor, MI: University of Michigan.

Rust, K. (1985). Variance Estimation for Complex Estimators in Sample Surveys. Journal of Official Statistics, 1, 381–397.

Rust, K. (1986). Efficient Replicated Variance Estimation. Proceedings of the American Statistical Association, Section on Survey Research Methods, 81–87.

Rust, K. and Kalton, G. (1987). Strategies for Collapsing Strata for Variance Estimation. Journal of Official Statistics, 3, 69–81.

Rust, K. and Rao, J.N.K. (1996). Variance Estimation for Complex Estimators in Sample Surveys. Statistics in Medical Research, 5, 381–397.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

Shao, J. (1996). Resampling Methods in Sample Surveys (with Discussion). Statistics, 27, 203–254.

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

StataCorp (2005). Stata Statistical Software: Release 9. College Station, TX: StataCorp LP.

Westat (2000). WesVar 4.0 User's Guide. Available at www.westat.com/wesvar. Rockville MD: Westat.

Wolter, K.M. (2007). Introduction to Variance Estimation. New York: Springer-Verlag.

Yung, W. and Rao, J.N.K. (2000). Jackknife Variance Estimation under Imputation for Estimators Using Poststratification Information. Journal of the American Statistical Association, 95, 903–915.

Received May 2006
Revised March 2008