Journal of Official Statistics
Vol. 3, No. 1, 1987, pp. 69–81

Strategies for Collapsing Strata for Variance Estimation

Keith Rust and Graham Kalton[1]

Abstract: The collapsed strata variance estimator is frequently used with sample designs in which one primary sampling unit is selected per stratum. The bias and precision of this variance estimator depend upon the way in which strata are collapsed. This paper examines the effects of collapsing strata in pairs, triples, and larger groups on the quality of the variance estimator. The effects of the bias and precision of the variance estimator on the inference for the parameter of interest are considered.

Key Words: Collapsed strata technique; variance estimation; mean square error.

1. Introduction

A feature of many survey sample designs is the selection of a single primary sampling unit (PSU) per stratum. This is achieved either explicitly through fine stratification, implicitly through systematic selection, or through a combination of these methods. The selection of a single PSU per stratum gives efficiency in design since stratification is carried out to the fullest possible extent, but it does not generally permit an unbiased variance estimator to be obtained. A widely used method of variance estimation for this situation is known as the collapsed strata technique. With this technique, strata and their corresponding sample PSUs are collapsed together in groups (collapsed strata) and then the variability among the units within these groups is used to derive a variance estimator (Hansen et al. (1953, Vol. I, §9.15, §9.28), Wolter (1985, §2.5)). This paper discusses the extent of collapsing to use, taking into account both the bias and variance of the resultant variance estimator.

Other approaches to estimating variances with one PSU per stratum designs have been proposed. A method proposed by Chromy (1981) selects one PSU per implicit stratum by a sequential selection procedure. This method permits unbiased variance estimation, but it is not clear that the unbiased estimator is sufficiently precise in many applications.

Other variance estimators for use with one PSU per stratum designs generally require auxiliary information. A model relating the auxiliary information and the survey variables is utilized, and must hold reasonably well if the variance estimator is to be satisfactory. Hartley et al. (1969) have, for instance, proposed a method which uses a linear regression of the stratum means on one or more auxiliary

[1] Keith Rust is a Senior Statistician at Westat, Inc., 1650 Research Blvd., Rockville, MD 20850–3129, USA. Graham Kalton is a Research Scientist at the Survey Research Center, and Professor of Biostatistics, University of Michigan. The authors would like to thank Editor Ingrid Lyberg, the referees and Dr. A.K. Srivastava for valuable comments. We are particularly grateful to the referee who drew our attention to some errors in an earlier version of the paper.

variables. The estimated covariance matrix of the regression residuals is then used to estimate the stratum variance components.

If the strata can be ordered in approximately ascending order of the stratum means, the method of successive differences (see Kish, 1965, §8.6B) is attractive. This method is an extension of collapsing strata in pairs. Frequently these methods have similar biases but the method of successiv differences has somewhat greater precision (DuMouchel et al. (1973)).

Isaki (1983) uses auxiliary information to reduce the bias of the collapsed strata variance estimator. Isaki's theoretical and empirical results suggest that when auxiliary variables highly correlated with the survey variable are available, substantial improvements in the accuracy of variance estimation can be obtained.

In practice, closely related auxiliary information is often not readily available, and this fact, together with the simplicity of execution of the original collapsed strata method, explains its continued popularity. In using the collapsed strata variance estimator, there are choices as to the extent to which strata are collapsed, and the manner in which collapsed strata are formed. This paper considers how best to collapse strata, and the extent of collapsing that gives rise to the most powerful inference concerning the population parameter of interest. To do this, a number of issues need to be addressed. Section 2 gives results on the bias and variance of the collapsed strata variance estimator, and shows how these quantities depend on the method of collapsing. Since the collapsed strata variance estimator is biased, the question of how to compare the performances of biased variance estimators needs to be considered; this is discussed in Section 3 for a simple case. Section 4 then examines the best choice of collapsing method to use in the simple case. The findings are summarized in Section 5.

2. The Collapsed Strata Variance Estimator

Suppose that the population to be sampled is divided into H strata, with stratum h containing N_h PSUs. A multistage sample is drawn by first selecting one PSU from each stratum, and then subsampling from within the selected PSUs. Let π_{hi} denote the probability that PSU i of stratum h is selected ($\sum_{i=1}^{N_h} \pi_{hi} = 1$). A parameter $\mu = \sum_{h=1}^{H} \mu_h$ is estimated using an unbiased linear estimator $\hat{\mu} = \sum_{h=1}^{H} \hat{\mu}_h$, where $\hat{\mu}_h = Y_{hi}/\pi_{hi}$ and Y_{hi} is the value of the characteristic of interest for PSU i of stratum h (if the sample is single stage) or an unbiased estimator of the characteristic of interest derived from the units subsampled from that PSU (if the sample is multistage).

Suppose that the H strata are partitioned into J groups of strata, known as collapsed strata, and that there are $H_j \geq 2$ strata in collapsed stratum j. Stratum h within collapsed stratum j is denoted by $h(j)$. The collapsed strata estimator of the variance of $\hat{\mu}$ is then

$$v_{cs}(\hat{\mu}) = \sum_{j=1}^{J} \frac{H_j}{(H_j - 1)} \sum_{h=1}^{H_j} [\hat{\mu}_{h(j)} - \frac{\hat{\mu}_j}{H_j}]^2,$$

where $\hat{\mu}_{h(j)}$ denotes the unbiased estimator of $\mu_{h(j)}$, the parameter value for stratum h in collapsed stratum j, and $\hat{\mu}_j = \sum_{h=1}^{H_j} \hat{\mu}_{h(j)}$.

2.1. *Bias of the collapsed strata variance estimator*

Hansen et al. (1953, Vol II, §9.5) give the expected value of $v_{cs}(\hat{\mu})$, from which it follows that

$$\text{Bias}\,(v_{cs}(\hat{\mu})) = \sum_{j=1}^{J} \frac{H_j}{(H_j - 1)} \sum_{h=1}^{H_j} [\mu_{h(j)} - \frac{\mu_j}{H_j}]^2, \quad (1)$$

where $\mu_j = \sum_{h=1}^{H_j} \mu_{h(j)}$. This bias is non-negative, and is zero only when the strata within each collapsed stratum j have a common value of $\mu_{h(j)}$. The bias is kept small by collapsing together strata with similar μ_h values. It will generally be smallest when little collapsing occurs so that values of μ_h can be kept similar within each collapsed stratum.

It should be noted that the choice of which strata are to be collapsed must be based on prior knowledge and expectations about the μ_h values, not on data from the sample. If sample data are used, for example to collapse strata with similar $\hat{\mu}_h$ values, a severe negative bias in variance estimation can result.

In the special but common case where $H_j = H/J$ ($= \bar{H}$ say) for all j, the bias of $v_{cs}(\hat{\mu})$ can readily be expressed in terms of the average variance of the μ_h within collapsed strata, σ^2_{wc}, where

$$\sigma^2_{wc} = \frac{1}{H} \sum_{j=1}^{J} \sum_{h=1}^{\bar{H}} [\mu_{h(j)} - \frac{\mu_j}{\bar{H}_j}]^2.$$

By comparing σ^2_{wc} with Bias $(v_{cs}(\hat{\mu}))$ in (1) it can be seen that

$$\text{Bias}(v_{cs}(\hat{\mu})) = \bar{H}H\sigma^2_{wc}/(\bar{H}-1).$$

This bias can alternatively be expressed as

$$\text{Bias}(v_{cs}(\hat{\mu})) = H(1-\varrho)\sigma^2_s,$$

where ϱ is the intraclass correlation of the μ_h values within collapsed strata, given by

$$\varrho = 1 - \frac{\bar{H}\sigma^2_{wc}}{(\bar{H}-1)\sigma^2_s},$$

and σ^2_s is the between strata variance given by

$$\sigma^2_s = \frac{1}{H} \sum_{h=1}^{H} [\mu_h - \frac{\mu}{\bar{H}}]^2.$$

The possible values of ϱ range from $-1/(\bar{H}-1)$ to 1.

In the following discussion, it proves more convenient to employ the relative bias rather than the bias of the collapsed strata variance estimator. The relative bias of $v_{cs}(\hat{\mu})$ is defined as

$$\text{Rel Bias}(v_{cs}(\hat{\mu})) = \text{Bias}(v_{cs}(\hat{\mu}))/V(\hat{\mu}).$$

2.2. *The likely magnitude of the relative bias of* $v_{cs}(\hat{\mu})$

As will be seen in Sections 3 and 4, some idea of the likely magnitude of the relative bias of $v_{cs}(\hat{\mu})$ is needed in order to assess alternative collapsed strata variance estimators. The following example gives the order of magnitude of the relative bias that might occur in one particular case. Suppose that a multistage sample with one PSU selected per stratum is drawn to estimate a population proportion μ. Strata are of equal size, and PSUs are selected with probabilities proportional to their exact sizes. A sample of n elements is selected from each selected PSU, giving a total sample size of Hn. Denote the stratum population proportions as P_h and their sample estimates as p_h, and let $\mu_h = P_h/H$ (so that $\mu = \sum_{h=1}^{H} \mu_h$) and $\hat{\mu}_h = p_h/H$. Further, let the variance of the stratum population proportions P_h be

$$\sigma^2_p = \frac{1}{H} \sum_{h=1}^{H} (P_h - \mu)^2 = H^2 \sigma^2_s.$$

For this example, the variance of $\hat{\mu}$ is

$$V(\hat{\mu}) = \sum_{h=1}^{H} \text{Deff}_h P_h(1-P_h)/nH^2$$

where Deff_h is the within stratum design effect (see Kish (1965, § 8.2)). For simplicity Deff_h is taken to be 1. Noting that

$$H\mu(1-\mu) = \sum_{h=1}^{H} P_h(1-P_h) + H\sigma^2_p,$$

the variance of $\hat{\mu}$ can be expressed as

$$V(\hat{\mu}) = [\mu(1-\mu) - \sigma_p^2]/nH.$$

Thus the relative bias of $v_{cs}(\hat{\mu})$ reduces to

$$\text{Rel Bias}\,(v_{cs}(\hat{\mu})) = \frac{n(1-\varrho)\sigma_p^2}{\mu(1-\mu)-\sigma_p^2}. \qquad (2)$$

As an illustration of likely values for ϱ and σ_p^2, consider the following examples. Suppose that the collapsed stratum proportions P_j are equally spaced over a range of 0.4 to 0.6, and that the stratum proportions within each collapsed stratum are equally spaced with a range of 0.1. For $H = 100$ strata using $J = 10$ collapsed strata, this corresponds to the case where one collapsed stratum contains ten strata with proportions (P_h) of .350, .361, .372, .383, ..., .439, .450, a second collapsed stratum has strata with P_h values of .372, .383, .394, ..., .450, .461, .472, and the tenth collapsed stratum has P_h values of .550, .561, .572, ..., .650. In this case the strategy of forming homogeneous collapsed strata has been only moderately successful. With such uniform distributions of proportion values both within and across collapsed strata, the values of σ_{wc}^2 and σ_p^2 are given by

$$\sigma_{wc}^2 = \frac{(0.1)^2(\bar{H}+1)}{12H^2(\bar{H}-1)},$$

and

$$\sigma_p^2 = \frac{(0.1)^2(\bar{H}+1)}{12(\bar{H}-1)} + \frac{(0.2)^2(J+1)}{12(J-1)}.$$

This gives $\sigma_p^2 = .0051$ and $\varrho = 0.78$ when $H = 100$ and $J = 10$ as in the above example.

If $H = 30$ and $J = 15$, there are 15 collapsed strata, with the first collapsed pair having P_h values of .350 and .450, the second having .363 and .463, the third .377 and .477, and the fifteenth .550 and .650. In this case, the collapsed strata are not very homogeneous with respect to P_h values, and this is reflected by the relatively low value for ϱ of 0.21. The value of σ_p^2 is .0063.

If $H = 12$ and $J = 2$, one collapsed stratum has P_h values of .35, .37, .39, .41, .43, .45 and the other has P_h values of .55, .57, .59, .61, .63, .65. Here the collapsed strata are quite homogeneous with $\varrho = 0.87$, and $\sigma_p^2 = .0112$.

Table 1. Relative biases of $v_{cs}(\mu)$ for a range of values of σ_p^2, ϱ, and n, for $\mu = 0.5$

σ_p^2	ϱ	n			
		5	10	25	50
.001	.00	.020	.040	.100	.201
	.75	.005	.010	.025	.050
	.95	.001	.002	.005	.010
	.99	.000	.000	.001	.002
.003	.00	.061	.121	.304	.607
	.75	.015	.030	.076	.152
	.95	.003	.006	.015	.030
	.99	.001	.001	.003	.006
.005	.00	.102	.204	.510	1.020
	.75	.026	.051	.128	.256
	.95	.005	.010	.026	.051
	.99	.001	.002	.005	.010
.01	.00	.208	.417	1.042	2.083
	.75	.052	.104	.260	.521
	.95	.010	.021	.052	.104
	.99	.002	.004	.010	.021

If the range of P_h values within collapsed strata is reduced from 0.1 to 0.05 in the above three examples, the homogeneity of the collapsed strata increases substantially. The ϱ values become 0.93, 0.72, and 0.97 respectively. The corresponding values for σ_p^2 are .0043, .0044, and .0103.

Table 1 shows the relative bias of $v_{cs}(\hat{\mu})$ for some likely values of σ_p^2, ϱ, and n, in the case where $\mu = 0.5$. It can be seen from the table that the relative bias varies greatly with variation in these three parameters.

For values of μ other than 0.5, the relative biases will be greater than those shown in Table 1 by a factor of $(0.25 - \sigma_p^2)/(\mu(1-\mu) - \sigma_p^2)$ (for the same values of σ_p^2, ϱ and n). For example, if $\mu = 0.2$, the relative biases will be slightly more than 56 % greater than those shown in Table 1. Note, however, that with more extreme proportions, relatively small values of σ_p^2 are likely to be encountered, as the range of P_h values will generally be small. Thus, even though the relative biases will be larger than those shown in the table, they will not be of great magnitude because small σ_p^2 values are associated with smaller relative biases.

Relative biases greater than 0.5 occur in Table 1 only when σ_p^2 and n are both large and ϱ is small. This combination of values is seldom likely to arise in practice. One reason is that when the variation among the stratum proportions is large ($\sigma_p^2 \geq .005$, say), it should generally be possible to collapse strata so that ϱ is large ($\varrho \geq 0.75$). Another reason is that in national surveys of human populations, the average cluster sizes are usually less than $n = 50$. Moreover, the values of n applicable for subclass estimates are smaller than that for the total sample. Thus, as a rule relative biases can be expected to be less than 0.5; only in the case of large cluster sizes ($n \geq 50$) are larger relative biases likely to occur.

2.3. *The variance of* $v_{cs}(\hat{\mu})$

The variance of $v_{cs}(\hat{\mu})$ is given by:

$$\mathrm{Var}(v_{cs}(\hat{\mu})) = \mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D} \qquad (3)$$

where

$$\mathrm{A} = \sum_{j=1}^{J} \left[\sum_{h=1}^{H_j} \mu_{h(j)}^{(4)} - \sum_{h=1}^{H_j} \sigma_{h(j)}^4 \right],$$

$$\mathrm{B} = \sum_{j=1}^{J} \frac{4H_j^2}{(H_j-1)^2} \sum_{h=1}^{H_j} \left[\mu_{h(j)} - \frac{\mu_j}{H_j}\right]^2 \sigma_{h(j)}^2,$$

$$\mathrm{C} = \sum_{j=1}^{J} \frac{4H_j}{(H_j-1)} \sum_{h=1}^{H_j} \left[\mu_{h(j)} - \frac{\mu_j}{H_j}\right] \mu_{h(j)}^{(3)},$$

$$\mathrm{D} = \sum_{j=1}^{J} \frac{4}{(H_j-1)^2} \sum_{h<k}^{H_j} \sum^{H_j} \sigma_{h(j)}^2 \, \sigma_{k(j)}^2,$$

$\mu_{h(j)}^{(4)} = E(\hat{\mu}_{h(j)} - \mu_{h(j)})^4$, $\mu_{h(j)}^{(3)} = E(\hat{\mu}_{h(j)} - \mu_{h(j)})^3$, and $\sigma_{h(j)}^2 = E(\hat{\mu}_{h(j)} - \mu_{h(j)})^2$. An outline of the derivation of this result is given in the appendix. Expression (3) generalizes the result given by DuMouchel et al. (1973) for the case when all $H_j = 2$.

It can be seen that the terms B, C and D are affected by the method of collapsing strata. The term B is always non-negative, and is minimized at zero, like the bias, when $\mu_{h(j)}$ is constant for all h in j. However, the two terms C and D will not necessarily be minimized by such an arrangement. The term C will be zero if either $\mu_{h(j)}$ or $\mu_{h(j)}^{(3)}$ is constant for all h in j, but this is not necessarily the minimum achievable, since this term may become negative. The term D depends upon the stratum sampling variances alone, and not the μ_h values.

As is evident from the above, the optimal extent of collapsing is not obvious, even if it is possible to collapse the strata so as to give an unbiased variance estimator. Minimal collapsing keeps the bias and the variance term B low, but will generally give rise to a relatively

large value for D. In the absence of a general solution for optimal collapsing, we will consider a case with some simplifying assumptions.

3. Comparing Collapsed Strata Variance Estimators

The comparison of unbiased, or approximately unbiased, variance estimators is commonly made in terms of their precisions (the inverses of their variances). The use of a more precise unbiased variance estimator gives rise to more powerful inference from the sample data. In comparing biased variance estimators, however, the effects of the biases and the precisions of the estimators must both be taken into account.

A measure frequently used to assess the combined effect of the bias and variance of an estimator is the mean square error. The mean square error provides a useful index of the quality of the estimator $\hat{\theta}$ for making inference about θ. However, since the purpose of using $v(\hat{\mu})$ to estimate $V(\hat{\mu})$ is not generally to make inference about $V(\hat{\mu})$, but to make inference about the parameter μ, it is not clear that the mean square error provides an appropriate index of the quality of a variance estimator.

The criterion we propose for choosing between variance estimators is that the preferred variance estimator is the one that leads to the most powerful inference, while still maintaining at least the stated level of confidence. The variance estimator giving the most powerful inference for a given confidence level $(1 - \alpha)$ is taken to be the one with the shortest expected confidence interval with coverage of at least $(1 - \alpha)$. Although we consider this criterion to be a reasonable one, it should be noted that it ignores issues relating to the control that the researcher has over the width of the confidence interval attained for a particular sample.

Confidence intervals are generally constructed as $\hat{\mu} \pm t_{k,(1-\alpha/2)}\sqrt{v(\hat{\mu})}$, where k is an appropriate number of degrees of freedom for the variance estimator. The form of $v(\hat{\mu})$ affects both its bias and variance. The presence of a positive bias increases the expected size of $v(\hat{\mu})$, and hence the expected width of the confidence interval. The variance affects the width of the confidence interval through the choice of the number of degrees of freedom for the t-value.

In this section we examine whether the standard confidence intervals calculated with the collapsed strata variance estimator satisfy the condition of having coverage of at least $(1 - \alpha)$. We then consider whether the mean square error serves as an adequate index of the quality of a variance estimator according to the criterion proposed above. To obtain tractable expressions for analyzing confidence interval coverage and mean square error, we first introduce some simplifying assumptions.

3.1. *A simple model*

In the remainder of the paper, a simple model for the population and sample design will be used to derive a less complex form for the variance of $v_{cs}(\hat{\mu})$ and in particular to derive a simple relationship between the variance and bias of $v_{cs}(\hat{\mu})$. The assumptions made are as follows: suppose that $\sigma^2_{h(j)} = \sigma^2$, $\mu^{(3)}_{h(j)} = \mu^{(3)}$ and $\beta_{h(j)} = \mu^{(4)}_{h(j)}/\sigma^4_{h(j)} = \beta$ are constant for all h, j. Then the variance of $\hat{\mu}$ is $V(\hat{\mu}) = H\sigma^2$. Further, consider only collapsing strategies for which collapsed strata contain equal numbers of strata, so that $H_j = H/J$ $(= \bar{H}$ say) for all j.

Under these assumptions, the relative variance of $v_{cs}(\hat{\mu})$ is obtained, using (3), as

$$\text{Rel Var}(v_{cs}(\hat{\mu})) = \text{Var}(v_{cs}(\hat{\mu}))/V(\hat{\mu})^2$$

$$= \frac{1}{H}[(\beta - 1) + \frac{2}{(\bar{H}-1)}\{1 + 2\bar{H}\ \text{Rel Bias}(v_{cs}(\hat{\mu}))\}].$$

Thus under the simple model, for given H the relative variance of $v_{cs}(\hat{\mu})$ reduces to a function of just three quantities: the relative bias, the extent of collapsing, and β, the common kurtosis of the sample estimators within strata. If $\beta = 3$, the value for a normal distribution, then

$$\text{Rel Var}(v_{cs}(\hat{\mu})) = \frac{2}{(H-J)}[1 + 2 \text{ Rel Bias}(v_{cs}(\hat{\mu}))]. \quad (4)$$

3.2. *Coverage properties of confidence intervals based on collapsed strata variance estimators*

Consider the collapsed strata variance estimator under the above model. In this case the standard method of constructing confidence intervals is to use $\hat{\mu} \pm t_{(H-J),\,(1-\alpha/2)}\sqrt{v_{cs}(\hat{\mu})}$, where the number of degrees of freedom adopted is $(H-J)$. For example, for a 95 % confidence interval ($\alpha = .05$) for a design with $H = 30$ strata collapsed in pairs ($\bar{H} = 2, J = 15$), the coefficient generally used is $t_{15,.975} = 2.1315$. The question considered here is whether such standard confidence intervals have coverage of at least $(1-\alpha)$.

Suppose that $\hat{\mu}$ is normally distributed, and that $v_{cs}(\hat{\mu})$ is independent of $\hat{\mu}$. Provided that the relative bias of $v_{cs}(\hat{\mu})$, denoted by RB, is not too large, it is reasonable to assume that $r \cdot v_{cs}(\hat{\mu})/[V(\hat{\mu})(1 + \text{RB})]$ has a chi-squared distribution with r degrees of freedom, where

$$r = 2(1 + \text{RB})^2/\text{Rel Var}(v_{cs}(\hat{\mu})).$$

Hence $(\hat{\mu} - \mu)\sqrt{1 + \text{RB}}/\sqrt{v_{cs}(\hat{\mu})}$ has a central t distribution with r degrees of freedom. With $\beta = 3$, it can be shown using (4) that

$$r = (H-J)(1+\text{RB})^2/(1 + 2\text{RB}).$$

Since $\text{RB} \geq 0$, it follows that the number of degrees of freedom r is at least $(H-J)$. A confidence interval for μ with exact 95 % coverage is given by $\hat{\mu} \pm t_{r,(1-\alpha/2)}\sqrt{v_{cs}(\hat{\mu})}/\sqrt{1+\text{RB}}$. Since $t_{(H-J),\,(1-\alpha/2)} \geq t_{r,\,(1-\alpha/2)}$ and $\sqrt{1+\text{RB}} \geq 1$, it thus follows that the standard confidence interval $\hat{\mu} \pm t_{(H-J),\,(1-\alpha/2)}\sqrt{v_{cs}(\hat{\mu})}$ has coverage of at least $(1-\alpha)$.

Under these assumptions, the standard confidence intervals are doubly conservative in that they use a larger t-value than is required for exact $(1-\alpha)$ coverage, and they also ignore the effect of the positive bias of the variance estimator. When the above distributional assumptions do not hold exactly, but do hold approximately, it seems probable that the standard confidence intervals will still have coverage of at least $(1-\alpha)$.

3.3. *Relationship of the mean square error of $v_{cs}(\hat{\mu})$ to confidence interval widths*

We now turn to consider how well the mean square error of $v_{cs}(\hat{\mu})$ tracks the widths of the standard confidence intervals discussed above. For this purpose, we will examine the relative mean square error of $v_{cs}(\hat{\mu})$, which we define as

$$\begin{aligned}\text{Rel MSE}(v_{cs}(\hat{\mu})) &= \text{MSE}(v_{cs}(\hat{\mu}))/V(\hat{\mu})^2 \\ &= \text{Rel Var}(v_{cs}(\hat{\mu})) \\ &\quad + \{\text{Rel Bias}(v_{cs}(\hat{\mu}))\}^2.\end{aligned}$$

The relative mean square error is preferred to the mean square error because it is expressed more readily in terms of relative bias. The relationship between standard confidence intervals and relative mean square error depends upon the chosen confidence level α. We use $\alpha = .05$, as 95 % confidence intervals are the most frequently utilized when making inference from sample survey data.

When $v_{cs}(\hat{\mu})$ is estimated using different extents of collapsing, the expected widths of the standard 95 % confidence intervals for μ vary in proportion to $t_{(H-J),.975}\sqrt{1+\text{RB}}$. This quantity is therefore used as the basis for the assessment of Rel MSE($v_{cs}(\hat{\mu})$) as a measure of the quality of $v_{cs}(\hat{\mu})$. However, rather than employ this quantity directly, we introduce what we term a *relative 95 % CI width*. This latter quantity is the expected width of the standard two-sided 95 % confidence interval expressed relative to a reference confidence interval width, where the reference is the expected 95 % confidence interval width resulting from the use of a simple random sample of H PSUs, with true sampling variance $V(\hat{\mu})$. The reference is thus the expected confidence interval width for the case where all strata are collapsed together and no bias results. For example, for 30 strata, the reference corresponds to the use of an unbiased variance estimator, together with the t coefficient appropriate for 29 degrees of freedom (2.0452).

Values of relative 95 % CI width and Rel MSE are given in Table 2 for the simple model of Section 3.1 with $\beta = 3$ and for a range of values of relative bias (RB), of numbers of strata (H), and of extents of collapsing ($\bar{H}$). An examination of the table shows that Rel MSE tracks the relative 95 % CI width poorly when there are 30 strata or more. For example, for $H = 30$ strata the table shows that, in terms of relative 95 % CI width, collapsing in tens with .05 relative bias (relative 95 % CI width 1.028) is not as efficient as collapsing in fives with a .02 relative bias (1.019). The corresponding values for Rel MSE (.084, .087) suggest on the contrary that, despite the substantially greater bias, collapsing in tens is superior.

The Rel MSE values track the relative 95 % CI widths better for 12 strata. For example, the alternatives of collapsing into one collapsed stratum of twelve strata with a .05 relative bias, and collapsing into fours with a relative bias of .02 have relative 95 % CI widths of 1.025 and 1.038 respectively. The

Table 2. Comparison of relative 95 % CI width and Rel MSE

	Relative 95 % CI width				Rel MSE			
	RB				RB			
$\bar{H}$	.01	.02	.05	.1	.01	.02	.05	.1
				$\bar{H} = 100$				
2	1.017	1.022	1.037	1.062	.041	.042	.047	.058
5	1.008	1.013	1.028	1.052	.026	.026	.030	.040
10	1.006	1.011	1.026	1.050	.023	.024	.027	.037
				$\bar{H} = 30$				
2	1.047	1.053	1.068	1.093	.136	.139	.149	.170
5	1.014	1.019	1.034	1.058	.085	.087	.094	.110
10	1.008	1.013	1.028	1.052	.076	.077	.084	.099
				$\bar{H} = 12$				
2	1.117	1.123	1.139	1.166	.340	.347	.369	.410
4	1.033	1.038	1.053	1.078	.227	.232	.247	.277
12	1.005	1.010	1.025	1.049	.186	.190	.203	.228

corresponding Rel MSE values of .203 and .232 reflect this ordering. Thus it can be seen that in comparing forms of the collapsed strata variance estimator, with respect to reflecting relative 95 % confidence interval widths, Rel MSE provides a poor criterion, except when only a few strata are involved. With larger numbers of strata, Rel MSE places too little emphasis on bias.

4. Choice of the Extent of Collapsing for Use with $v_{cs}(\hat{\mu})$

We now consider the choice of the extent of collapsing to use with the collapsed strata variance estimator, employing the 95 % confidence interval width as the index of quality. As in Section 3, two-sided 95 % confidence intervals are used, these being the most frequently employed in practice. It must be noted that the conclusions reached regarding the desirable degree of collapsing may differ from those reached if a different level of confidence is used. Under the assumptions of the simple model of Section 3.1, Table 3 shows the effect of a range of extents of collapsing on 95 % confidence interval widths in the presence of varying degrees of relative bias. Parts (a), (b), and (c) of the table are for $H = 100$,

Table 3. Relative 95 % CI widths for different extents of collapsing

(a) $H = 100$

	$\bar{H}$:	2	3	4	5	10
Rel	0.0	1.012	1.006	1.004	1.003	1.001
Bias	.005	1.015	1.009	1.007	1.006	1.004
	.01	1.017	1.011	1.009	1.008	1.006
	.02	1.022	1.016	1.014	1.013	1.011
	.03	1.027	1.021	1.019	1.018	1.016
	.05	1.037	1.031	1.029	1.028	1.026
	.1	1.062	1.055	1.053	1.052	1.050
	.2	1.109	1.102	1.100	1.099	1.097
	.4	1.198	1.190	1.188	1.187	1.185

(b) $H = 30$

	$\bar{H}$:	2	3	5	6	10
Rel	0.0	1.042	1.020	1.009	1.007	1.003
Bias	.005	1.045	1.023	1.012	1.010	1.006
	.01	1.047	1.025	1.014	1.012	1.008
	.02	1.053	1.030	1.019	1.017	1.013
	.03	1.058	1.035	1.024	1.022	1.018
	.05	1.068	1.045	1.034	1.032	1.028
	.1	1.093	1.070	1.058	1.056	1.052
	.2	1.142	1.117	1.106	1.103	1.099
	.4	1.233	1.207	1.194	1.192	1.187

(c) $H = 12$

	$\bar{H}$:	2	3	4	6	12
Rel	0.0	1.112	1.048	1.028	1.012	1.000
Bias	.005	1.115	1.050	1.030	1.015	1.002
	.01	1.117	1.053	1.033	1.017	1.005
	.02	1.123	1.058	1.038	1.022	1.010
	.03	1.128	1.063	1.043	1.027	1.015
	.05	1.139	1.074	1.053	1.037	1.025
	.1	1.166	1.099	1.078	1.062	1.049
	.2	1.218	1.148	1.126	1.109	1.095
	.4	1.315	1.240	1.216	1.198	1.183

30, and 12 strata respectively, and in each case $\beta = 3$. The table entries are relative 95 % CI widths, derived in the same way as those in Table 2.

Table 3(a) shows that when there are many strata ($H = 100$), little is to be gained from extensive collapsing. Reading across the rows of the table shows that there is little reduction in confidence interval width when the relative bias remains constant as collapsing is increased. If additional collapsing leads to even a modest increase in bias, the quality of variance estimation decreases. For instance, collapsing in tens with a .03 relative bias (relative 95 % CI width = 1.016) is slightly inferior to collapsing in pairs with a .005 relative bias (1.015). Since increased collapsing is likely to result in increased bias, the risks inherent in the use of extensive collapsing outweigh the potential benefits. In this case, collapsing in pairs would seem the safest strategy. Even in the presence of a relative bias as large as .05 collapsing in pairs will result in a 95 % confidence interval that is only 3.7 % wider than the reference confidence interval.

When there are only 12 strata (Table 3(c)), the situation is different. Appreciable gains in the precision of variance estimation result from a greater degree of collapsing, resulting in a noticeable reduction in confidence interval width, even if substantial bias is introduced. For example, collapsing six strata together and incurring a relative bias of 0.1 (relative 95 % CI width = 1.062) is markedly superior to collapsing in pairs with no bias (1.112). When there is a large variation in stratum means or proportions (large σ_s^2 or σ_p^2) and collapsing into relatively homogeneous groups (large ϱ) is possible, and collapsing into groups of six or four strata appears a good strategy. Forming one collapsed stratum containing all 12 strata ($\varrho = -.09$) will give a high relative bias in this case.

For a design with 30 strata (Table 3(b)), results between these two extremes are found. In this case the best choice of a collapsing strategy is heavily dependent upon the extent to which the relative bias increases with the extent of collapsing. There are worthwhile gains to be had from a high level of collapsing if minimal additional bias is introduced, but these are easily lost if an appreciable increase in bias occurs. Table 3(b) shows that collapsing in triples is superior to collapsing in pairs provided that the resulting increase in relative bias is less than about .04. Collapsing in fives is superior to collapsing in triples if the increase in relative bias is less than .02. Additional collapsing beyond fives leads to little gain at best, and this will disappear with only a little increase in bias. A strategy of collapsing in triples thus appears to be a robust approach. Collapsing in triples with a .03 relative bias, for instance, gives confidence intervals that are only 3.5 % wider than those attainable under complete collapsing with no bias. Collapsing in fives will reduce confidence interval widths by only $(1.035 - 1.024)/1.035 = 1.1$ % if no further bias is introduced, whereas if the relative bias increases to 0.10 as a result of such collapsing, the interval widths will increase by $(1.058 - 1.035)/1.035 = 2.2$ %.

If a reasonable assessment of the relative bias resulting from different collapsing strategies can be made, the optimum collapsing strategy can be chosen by calculating the relative CI widths. For example, with 30 strata and using 95 % confidence intervals, suppose that collapsing in pairs gives a relative bias of .01. Collapsing in triples will be superior if the relative bias is below $(1.047^2/1.020^2) - 1 = .055$. If the relative bias from collapsing in triples is certainly lower than this, say around .02, then collapsing in triples is clearly the superior strategy. Collapsing in fives is superior to collapsing in triples with a relative bias of .02 if the resultant relative bias is below $(1.030^2/1.009^2) - 1 = .042$. Thus, if collapsing in fives is thought likely to result in a relative bias of .05 or greater, it should not be used.

In making an assessment of the magnitude of the relative bias to be expected from a given collapsing strategy, one must remember that the relative bias varies with the sample size per PSU (see, for example, equation (2)). This implies that a greater extent of collapsing will be appropriate for subclasses than for the total sample, as is discussed below.

The relative bias of a collapsed strata variance estimator depends upon the bias of the estimator and on the size of the sampling variance being estimated. The first of these quantities is a function of the variation among stratum population means within each collapsed stratum. For a given set of collapsed strata this variation is likely to be approximately the same for a subclass, and particularly for a crossclass spread evenly across PSUs, as for the total sample. In consequence, the bias of the subclass variance estimator will be of similar size to that of the corresponding total sample variance estimator. For a greater extent of collapsing, the variation among stratum means is likely to increase, thus increasing the bias of the variance estimator. On the other hand, the sampling variance of the subclass estimator will be larger than that of the total sample estimator, but will be unaffected by the manner in which strata are collapsed. It therefore follows that the absolute difference between the relative biases resulting from two different extents of collapsing will be smaller for the subclass variance estimator than for the total sample variance estimator. From Table 3 it can be seen that when the absolute difference in relative bias between two alternative extents of collapsing is small, the variance estimator using the greater extent of collapsing is preferable. Hence in general a greater degree of collapsing should be used to estimate a subclass variance than a total sample variance.

The criterion used for the assessment of collapsed strata variance estimators in the above discussion has been the expected width of the confidence interval with coverage of at least $(1 - \alpha)$. Applying this criterion, a greater extent of collapsing is beneficial when the use of a smaller t coefficient (arising from a greater number of degrees of freedom in the variance estimator) outweighs the effect of the extra bias in the variance estimator. Another distinct consideration concerns the random variation in the width of the confidence interval from sample to sample. Should two different extents of collapsing give rise to the same expected confidence interval width, the one with the greater extent of collapsing yields the more stable width. The more stable the width from sample to sample, the more control the survey researcher has over the width of the confidence interval that will be obtained from a particular sample.

5. Summary of Findings

The following factors need to be considered in deciding on the extent of collapsing to be used:

i. The number of strata in the design. Less collapsing is indicated for designs with many strata than for designs with few strata.
ii. Differences between stratum means within collapsed strata. If these are great for a given degree of collapsing, a lower degree of collapsing is recommended (provided that more homogeneous collapsed strata can be formed as a result).
iii. The final stage sample size per selected PSU. If this is small, a higher level of collapsing should be used.
iv. Subclass (and particularly crossclass) estimates. A greater level of collapsing is desirable for variances of subclass estimates.

As a rule the routine practice is to collapse strata in pairs. This is probably generally appropriate for national estimates from large-scale surveys with 60 or more PSUs. However,

a greater degree of collapsing may be appropriate when a smaller sample of PSUs is selected, and especially so when the number of PSUs is as few as, say, 20.

6. References

Chromy, J.R. (1981): Variance Estimators for a Sequential Sample Selection Procedure. In Current Topics in Survey Sampling, edited by D. Krewski, R. Platek, and J.N.K. Rao. Academic Press, New York, pp. 329–347.

DuMouchel, W.H., Govindarajulu, Z., and Rothman, E. (1973): A Note on Estimating the Variance of the Simple Mean in Stratified Sampling. Canadian Journal of Statistics, 1(2), pp. 267–274.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953): Sample Survey Methods and Theory. Vol. I. Methods and Applications. Vol. II. Theory. John Wiley and Sons, New York.

Hartley, H.O., Rao, J.N.K., and Keifer, G. (1969): Variance Estimation with One Unit per Stratum. Journal of the American Statistical Association, 64, pp. 841–851.

Isaki, C.T. (1983): Variance Estimation Using Auxiliary Information. Journal of the American Statistical Association, 78, pp. 117–123.

Kish, L. (1965): Survey Sampling. John Wiley and Sons, New York.

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. Springer-Verlag, New York.

Received April 1986
Revised March 1987

Appendix

Variance of the Collapsed Strata Variance Estimator

Equation (3) in Section 2 gives the variance of the collapsed strata variance estimator, $v_{cs}(\hat{\mu})$. An outline of the derivation of this expression is given below.

Hansen et al. (1953, Vol. II, § 9.5) show that

$$E\left[\sum_{j=1}^{J}\frac{H_j}{(H_j-1)}\sum_{h=1}^{H_j}(\hat{\mu}_{h(j)}-\frac{\hat{\mu}_j}{H_j})^2\right]=\sum_{j=1}^{J}\left[\sum_{h=1}^{H_j}\sigma^2_{h(j)}+\frac{H_j}{(H_j-1)}\sum_{h=1}^{H_j}(\mu_{h(j)}-\frac{\mu_j}{H_j})^2\right],$$

where the notation is as given in Section 2. It then follows that

$$\mathrm{Var}(v_{cs}(\hat{\mu}))=\sum_{j=1}^{J}\frac{H_j^2}{(H_j-1)^2}E\left[\sum_{h=1}^{H_j}\left\{(\hat{\mu}_{h(j)}-\frac{\hat{\mu}_j}{H_j})^2-(\mu_{h(j)}-\frac{\mu_j}{H_j})^2-\frac{(H_j-1)}{H_j}\sigma^2_{h(j)}\right\}\right]^2. \qquad (1A)$$

Substituting

$$(\hat{\mu}_{h(j)}-\mu_{h(j)})^2+\frac{1}{H_j^2}(\hat{\mu}_j-\mu_j)^2+2(\hat{\mu}_{h(j)}-\mu_{h(j)})(\mu_{h(j)}-\frac{\mu_j}{H_j})-\frac{2}{H_j}(\mu_{h(j)}-\frac{\mu_j}{H_j})(\hat{\mu}_j-\mu_j)$$

$$-\frac{2}{H_j}(\hat{\mu}_{h(j)}-\mu_{h(j)})(\hat{\mu}_j-\mu_j)$$

for $(\hat{\mu}_{h(j)} - \frac{\hat{\mu}_j}{H_j})^2 - (\mu_{h(j)} - \frac{\mu_j}{H_j})^2$ in (1A) and expanding the square gives

$$\mathrm{Var}(v_{cs}(\hat{\mu})) = \sum_{j=1}^{J} \frac{H_j^2}{(H_j-1)^2} E\Bigg[\sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^4 + \sum_{h \neq k}^{H_j} \sum^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^2 (\hat{\mu}_{k(j)} - \mu_{k(j)})^2$$

$$+ \frac{(H_j-1)^2}{H_j^2} \sum_{h=1}^{H_j} \sigma_{h(j)}^4 + \frac{(H_j-1)^2}{H_j^2} \sum_{h \neq k}^{H_j} \sum^{H_j} \sigma_{h(j)}^2 \sigma_{k(j)}^2 + \frac{1}{H_j^2} (\hat{\mu}_j - \mu_j)^4$$

$$+ 4 \sum_{h=1}^{H_j} (\mu_{h(j)} - \frac{\mu_j}{H_j})^2 (\hat{\mu}_{h(j)} - \mu_{h(j)})^2$$

$$+ 4 \sum_{h \neq k}^{H_j} \sum^{H_j} (\mu_{h(j)} - \frac{\mu_j}{H_j}) (\hat{\mu}_{h(j)} - \mu_{h(j)}) (\mu_{k(j)} - \frac{\mu_j}{H_j}) (\hat{\mu}_{k(j)} - \mu_{k(j)})$$

$$- \frac{2(H_j-1)}{H_j} (\sum_{h=1}^{H_j} \sigma_{h(j)}^2) \sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^2 - \frac{2}{H_j} (\hat{\mu}_j - \mu_j)^2 \sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^2$$

$$+ 4 \sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^3 (\mu_{h(j)} - \frac{\mu_j}{H_j}) + 4 \sum_{h \neq k}^{H_j} \sum^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)})^2 (\hat{\mu}_{k(j)} - \mu_{k(j)}) (\mu_{k(j)}$$

$$- \frac{\mu_j}{H_j}) + \frac{2(H_j-1)}{H_j^2} (\hat{\mu}_j - \mu_j)^2 \sum_{h=1}^{H_j} \sigma_{h(j)}^2 - \frac{4(H_j-1)}{H_j} (\sum_{h=1}^{H_j} \sigma_{h(j)}^2) \sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)}) (\mu_{h(j)}$$

$$- \frac{\mu_j}{H_j}) - \frac{4}{H_j} (\hat{\mu}_j - \mu_j)^2 \sum_{h=1}^{H_j} (\hat{\mu}_{h(j)} - \mu_{h(j)}) (\mu_{h(j)} - \frac{\mu_j}{H_j}) \Bigg].$$

Using the relations

$$E(\hat{\mu}_j - \mu_j)^4 = \sum_{h=1}^{H_j} \mu_{h(j)}^{(4)} + 3 \sum_{h \neq k}^{H_j} \sum^{H_j} \sigma_{h(j)}^2 \sigma_{k(j)}^2 ,$$

$$E(\hat{\mu}_{h(j)} - \mu_{h(j)})^2 (\hat{\mu}_j - \mu_j)^2 = \mu_{h(j)}^{(4)} + \sigma_{h(j)}^2 \sum_{k \neq h}^{H_j} \sigma_{k(j)}^2 ,$$

$$E(\hat{\mu}_j - \mu_j)^2 = \sum_{h=1}^{H_j} \sigma_{h(j)}^2 ,$$

$$E(\hat{\mu}_{h(j)} - \mu_{h(j)})(\hat{\mu}_j - \mu_j)^2 = \mu_{h(j)}^{(3)} ,$$

where terms are as defined in Section 2, and collecting terms, gives the stated result. Additional details of the derivation are given in Rust (1984).