Biometrika (1994), **81**, 1, *pp.* 157–72
Printed in Great Britain

Precision estimation in sample survey inference: A criterion for choice between variance estimators

By ROLF SUNDBERG
Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University, S-10691 Stockholm, Sweden

Summary

We advocate the 'mean squared error of predicted squared error' as a universal criterion for the choice between variance estimators in sample survey inference. The predictive nature of the measure rewards variance estimators adapted to the amount of information in the actual sample. This makes the new measure more satisfactory than the simpler mean squared error of the variance estimator. The criterion turns out to be the same for design-based as for model-based inference, and may also be used for comparison between the design- and model-based theoretical variances. The theory is exemplified for the ratio estimator by a study of six variance estimators suggested in the literature. These calculations are primarily made under a proportional regression model, but consideration is also paid to model robustness. Three of the six estimators are shown to be inferior to the three others, which are approximately equivalent.

Some key words: Expansion estimator; Mean squared error of predicted squared error; Predictive inference; Randomization inference; Ratio estimator; Robust variance estimation.

1. Introduction

In classical sample survey inference concerning a finite population total or some other population statistic, as presented for example in Cochran's (1977) book, the statistical error of an estimator is represented by its bias and its variance evaluated by averaging over all possible samples from the given population according to the randomness specified by the sampling design. Analogously, an estimator of this variance is judged by its sampling bias and variance. As a challenge to this principle of design-based inference, model-based inference regards the population as generated by a stochastic superpopulation model and the actual population characteristic as an outcome to be predicted. Model-based inference allows direct conditioning on the realized sample of units or labels. This is desirable if the sampling units should be given different weight in the inference, for instance in the presence of concomitant variables or in subsampling situations, in particular when the realized sample is extreme. On the other hand model-based inference is model-dependent, and if the actual population is extreme as a realization of the model assumed, a model-based inference may be misleading.

In other words model-based inference provides an answer to the doubts expressed by the question 'Why should my inference depend on samples other than the actual one, samples which could have occurred but did not?', whereas design-based inference responds to the question 'Why should my inference depend on populations other than the actual one?' However, statistical inference requires one kind of randomness or another.

Discussions of model-based versus design-based inference in sample surveys are given by, e.g., Royall (1970), Smith (1976, 1984), Cassel, Särndal & Wretman (1977), Särndal (1978, 1985), Hansen, Madow & Tepping (1983).

Design-based inference does not preclude the consideration of superpopulation models. On the contrary such models give to a population the structure required for guidance in the choice between different sampling procedures, different estimators of a population parameter or, as here, between different estimators of the variance of a chosen estimator. See Särndal, Swensson & Wretman (1992) for a recent text along these lines. We will incorporate design and model aspects in a common framework under the joint assumption of:

(1) a prescribed sampling scheme for the selection of units;

(2) a superpopulation model that might have generated the population values;

(3) a chosen estimator/predictor, unbiased under sampling or model, or both.

We will argue that both theoretical variances, be they model-based or design-based, and variance estimators should be regarded as statistics for prediction of the actual squared error of the estimator/predictor, and we propose a quadratic measure of how close such a statistic is to the actual error, on average under the joint randomness of design and model. The principle 'the closer, the better' will provide a rating of the variances or variance estimators under study.

Empirical comparisons of design- and model-motivated variance estimators have been undertaken in several papers, for example by Royall & Cumberland (1978b, 1981a, b), Cumberland & Royall (1981), Wu & Deng (1983), Deng & Wu (1987) and Valliant (1990), who compare the performance of various variance estimators under simulated sampling from a few real populations. The present paper may be regarded as a theoretical counterpart to this type of papers. There are many previous theoretical comparisons based on the sampling variance of the variance estimator and its expected value over a superpopulation. We will explain why that approach is not good enough.

In §§ 3–5 the ideas will be applied to the expansion and ratio estimators/predictors, and we will draw conclusions as to which are the best variance estimators for the ratio estimator. These conclusions pay consideration also to the important aspect of robustness against deviations from the assumed model.

2. The theoretical concept

Suppose that an unbiased estimator $\hat{\theta}$ has been chosen for the inference about a parameter θ in a statistical model. Conventionally the precision of $\hat{\theta}$ is given by its variance $V = E\{(\hat{\theta}-\theta)^2\}$, which is the best measure of the squared error itself among constants c, in the sense of minimizing over c the expected squared difference

$$E[\{(\hat{\theta}-\theta)^2 - c\}^2]. \tag{2·1}$$

Typically V must be replaced by a statistic v that is a function of available data. The conventional way of considering this replacement is as an estimation problem. Second order properties of an unbiased v are then judged according to the variance of v as an estimator of the parameter V. However, the primary object is to make a statement about the squared error; the role as an estimator of V is secondary. We therefore consider it a more natural principle that v be judged as a predictor of $(\hat{\theta}-\theta)^2$, and in analogy with (2·1) we evaluate v according to the quadratic measure

$$E[\{(\hat{\theta}-\theta)^2 - v\}^2]. \tag{2·2}$$

Note that unbiasedness means the same for estimators of V and predictors of $(\hat{\theta}-\theta)^2$. In many estimation situations there is only a single unbiased v under consideration, and then it is a question of only academic interest whether we select v as predictor or as estimator. In other situations, frequent in sampling theory, there are several competing statistics v, and because they are more or less correlated with $(\hat{\theta}-\theta)^2$ they do not necessarily rank in the same order as predictors of $(\hat{\theta}-\theta)^2$ as they do as estimators of V. The ratio estimator will later be seen to provide a good example.

For a finite population of N units with associated Y-values, consider an estimator $\hat{T}$ of a population parameter $T=T(Y_1,\ldots,Y_N)$, unbiased under the randomness of a specified sampling procedure. The quantity V is then represented by the sampling variance of $\hat{T}$. This variance will depend on the unknown finite population values, and so will any measure of precision of a variance estimator v, be it the expected squared difference (2·2) or the conventional sampling variance of an unbiased estimator v. For guidance in the choice between different statistics v, as well as between different $\hat{T}$, the solution is to make a model-assisted evaluation by averaging over a superpopulation model that reflects the structure of the population. When this is done with (2·2) we obtain the measure

$$E_{(\mathscr{S},\mathscr{M})}[\{(\hat{T}-T)^2-v\}^2], \tag{2·3}$$

where $\mathscr{S}$ and $\mathscr{M}$ stand for the probability distributions defined by the sampling design and the superpopulation model respectively, and $(\mathscr{S},\mathscr{M})$ denotes the joint distribution.

Analogously, if $\hat{T}$ has been chosen as an unbiased predictor of T for a model-based inference, based on a fixed sample of units from a population whose Y-values are imagined as randomly generated, the prediction variance of $\hat{T}-T$ is the best constant measure of quadratic error, with the expected value in (2·1) taken over populations instead of over samples. In the measure (2·2) for the choice between different precision statistics v the expected value is now taken over the superpopulation model. However, this choice would be sample-dependent, that is we must wait until the sample is realized until we can make the choice. In practice we will probably desire more universal guidance in this choice and it is natural to replace the sample-dependent version of criterion (2·2) by its average over all possible samples, that is the measure (2·3) once more.

Thus we make the remarkable observation that, although proponents of design-based and model-based inference follow different principles in their choice of $\hat{T}$ and in their interpretation of the concept of precision of $\hat{T}$, they have reason to agree about the criterion for rating precision estimators. Note that (2·3) provides an adequate criterion for any precision statistic v, be it motivated by design or model considerations, or even utterly unfounded.

We saw that the sampling and prediction theoretical variances both minimize expressions of type (2·1) but with expected values calculated under different probability distributions. We can make them comparable by regarding them like v in (2·3) as precision statistics under joint randomness, and compare

$$E_{(\mathscr{S},\mathscr{M})}[\{(\hat{T}-T)^2-\mathrm{var}_{\mathscr{S}|\mathscr{M}}(\hat{T})\}^2] \tag{2·4}$$

with

$$E_{(\mathscr{S},\mathscr{M})}[\{(\hat{T}-T)^2-\mathrm{var}_{\mathscr{M}|\mathscr{S}}(\hat{T}-T)\}^2]. \tag{2·5}$$

Here $\mathscr{S}|\mathscr{M}$ indicates the sampling distribution given a fixed population generated by the superpopulation model, and $\mathscr{M}|\mathscr{S}$ the distribution defined by the model, given a fixed sample of units. Comparisons under this joint randomness should be of some interest when discussing the merits of the design-based and model-based approaches, see §§ 3 and 4 for examples.

Any quantity of type (2·2)–(2·5) might be called a 'mean squared error of predicted squared error'. We propose that the choice between statistics for variance estimation in specified situations be based on this measure.

From a design-based point of view we could ask why we should not measure the performance of an estimator v by its sampling variance, or mean squared error if v is not unbiased, averaged over $\mathscr{M}$ to get a measure that does not involve the actual unknown population: $E_{\mathscr{M}}\{\text{var}_{\mathscr{S}|\mathscr{M}}(v)\}$. Among others Wu (1982) used this measure, in a theoretical study for the ratio estimator, but Wu & Deng (1983) found empirically that it was unsatisfactory for describing precision and confidence interval coverage, remarking

> In estimating the population mean the purpose of variance estimation is rather for assessing the variability of the ratio estimator than for estimating the variance itself.

The mean squared error of predicted squared error satisfies their demand by measuring how well v predicts $(\hat{T}-T)^2$, thereby paying attention to the amount of information in the sample concerning the error $(\hat{T}-T)^2$, whereas $\text{var}_{\mathscr{S}|\mathscr{M}}(v)$ primarily measures how precisely v estimates $\text{var}_{\mathscr{S}|\mathscr{M}}(\hat{T})$. Only for v that are uncorrelated with $(\hat{T}-T)^2$ will $E_{\mathscr{M}}\{\text{var}_{\mathscr{S}|\mathscr{M}}(v)\}$ be a measure equivalent to (2·3). An analogous argument may be given within model-based inference. It is intuitively reasonable that the argument carries over to confidence intervals, so that intervals based on $(\hat{T}-T)^2/v$ are most correct when v best possible reflects the sampling variability of the estimator $\hat{T}$, rather than when v best estimates a population variance quantity. If so, the conclusions about confidence interval coverage of Wu & Deng (1983) might be explained this way.

Isaki & Fuller (1982) introduced the anticipated variance of $\hat{T}-T$ as being the variance of $\hat{T}-T$ over both design and model jointly. If v satisfies the unbiasedness requirement

$$E_{(\mathscr{S},\mathscr{M})}(v) = E_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2\}, \tag{2·6}$$

the mean squared error of predicted squared error criterion is formally the same for v as a predictor of $(\hat{T}-T)^2$ as the anticipated variance of $\hat{T}-T$ is for $\hat{T}$ as a predictor of T.

Expressions (2·3)–(2·5) may be rewritten in a form more convenient for calculations and comparisons. Assume first quite generally that v is any statistic satisfying relation (2·6). Typically v has been chosen among $\mathscr{S}$-unbiased estimators or $\mathscr{M}$-unbiased predictors of the respective variance of $\hat{T}$ or $\hat{T}-T$, so that (2·6) is satisfied. We may then rewrite (2·3) as

$$E_{(\mathscr{S},\mathscr{M})}[\{(\hat{T}-T)^2 - v\}^2] = \text{var}_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2\} - 2\,\text{cov}_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2, v\} + \text{var}_{(\mathscr{S},\mathscr{M})}(v). \tag{2·7}$$

The first term, $\text{var}_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2\}$, does not depend on v and will therefore be called the common term. It corresponds to the choice $v = \text{const} = \text{var}_{(\mathscr{S},\mathscr{M})}(\hat{T}-T)$, the overall variance, that will minimize (2·3) only among constants, and will be inferior to the conditional variances in (2·4) and (2·5).

The v-dependent remainder of (2·7) will be regarded as a reduction of the common term, to be called the specific reduction, in formulae to be abbreviated spec.red. Hence, for a particular statistic v,

mean squared error of predicted squared error = common term − specific reduction.

In the special cases (2·4) and (2·5) with

$$v = V_{\mathscr{S}} = \text{var}_{\mathscr{S}\mid\mathscr{M}}(\hat{T}), \quad v = V_{\mathscr{M}} = \text{var}_{\mathscr{M}\mid\mathscr{S}}(\hat{T} - T),$$

v is constant under $\mathscr{S}$ and $\mathscr{M}$ respectively, so the covariance in (2·7) equals the following variance and the specific reductions simplify to

$$\text{spec.red.}\,\{V_{\mathscr{S}}\} = \text{var}_{\mathscr{M}}(V_{\mathscr{S}}), \quad \text{spec.red.}\,\{V_{\mathscr{M}}\} = \text{var}_{\mathscr{S}}(V_{\mathscr{M}}). \tag{2·8}$$

If $v = v_m$ is a model-unbiased estimator of the theoretical conditional variance $V_{\mathscr{M}} = \text{var}_{\mathscr{M}\mid\mathscr{S}}(\hat{T} - T)$, its specific reduction may be expressed as

$$\text{spec.red.}\,\{v_m\} = \text{spec.red.}\,\{V_{\mathscr{M}}\} + E_{\mathscr{S}}[2\,\text{cov}_{\mathscr{M}\mid\mathscr{S}}\{(\hat{T} - T)^2, v_m\} - \text{var}_{\mathscr{M}\mid\mathscr{S}}(v_m)]. \tag{2·9}$$

This form is convenient for calculations and is seen from (2·7) by writing

$$\text{var}_{(\mathscr{S},\mathscr{M})}(.) = \text{var}_{\mathscr{S}}\{E_{\mathscr{M}\mid\mathscr{S}}(.)\} + E_{\mathscr{S}}\{\text{var}_{\mathscr{M}\mid\mathscr{S}}(.)\},$$

and correspondingly for $\text{cov}_{(\mathscr{S},\mathscr{M})}(.\,,.)$. Analogously we could express the specific reduction for a sampling-unbiased estimator v_s of $V_{\mathscr{S}} = \text{var}_{\mathscr{S}\mid\mathscr{M}}(\hat{T})$. Note that when we replace a theoretical variance by an estimator, the specific reduction may be increased or reduced, depending primarily on the correlation, $\mathscr{M}\mid\mathscr{S}$ or $\mathscr{S}\mid\mathscr{M}$, between this estimator and the actual squared error $(\hat{T} - T)^2$.

We will now turn to examples of situations where it may be of interest to compare the specific reductions for theoretical variances or variance estimators. In all examples T will be the finite population total.

3. Sample mean under simple random sampling and exchangeability

This is an almost trivial example but a natural starting point. Let the finite population be $(Y_1, \ldots, Y_N)$, with total $T = N\bar{Y}$. A sample of size n is to be taken by simple random sampling without replacement. We consider two different superpopulation models in which unit labels carry no information about the Y-values.

Model $\mathscr{M}1$. $Y_1, \ldots, Y_N$ is a randomly labelled finite set of fixed values.

Model $\mathscr{M}2$. $Y_1, \ldots, Y_N$ are independent and identically distributed with mean μ and standard deviation σ.

Note that model $\mathscr{M}1$ is obtained from $\mathscr{M}2$ by conditioning on the order statistic of the population.

The natural estimator/predictor of T is the expansion estimator $\hat{T} = N\bar{y}$, where $\bar{y}$ is the sample mean. Here $\hat{T}$ is design- and model-unbiased with variances

$$V_{\mathscr{S}} = \text{var}_{\mathscr{S}\mid\mathscr{M}}(\hat{T}) = \frac{N^2}{n}(1 - f)S^2, \tag{3·1}$$

$$V_{\mathscr{M}1} = \text{var}_{\mathscr{M}1\mid\mathscr{S}}(\hat{T} - T) = \frac{N^2}{n}(1 - f)S^2, \tag{3·2}$$

$$V_{\mathscr{M}2} = \text{var}_{\mathscr{M}2\mid\mathscr{S}}(\hat{T} - T) = \frac{N^2}{n}(1 - f)\sigma^2, \tag{3·1}$$

where $f = n/N$ and $S^2 = (N-1)^{-1}\sum(Y_i - \bar{Y})^2$, when the summation is over the range

$i=1,\ldots,N$. The equality of $V_{\mathscr{S}}$ and $V_{\mathscr{M}1}$ simply reflects the probabilistic equivalence between simple random sampling and random labelling, making the inferences equivalent. In particular $V_{\mathscr{S}}$ and $V_{\mathscr{M}1}$ have identical mean squared error of predicted squared error under $(\mathscr{S}, \mathscr{M}1)$.

Under $(\mathscr{S}, \mathscr{M}2)$ we have the following results; compare with the more general § 4. The common term may be written $2nC_n\{1+O(1/n)\}$, for fixed f, where

$$C_n = N^4\sigma^4(1-f)^2/n^3.$$

For the $\mathscr{M}2$-based precision measure (3·3),

$$\text{spec.red. }\{V_{\mathscr{M}2}\} = 0, \tag{3·4}$$

whereas for $V_{\mathscr{S}}$ under $(\mathscr{S}, \mathscr{M}2)$

$$\text{spec.red. }\{V_{\mathscr{S}}\} = \frac{N^4}{n^2}(1-f)^2 \operatorname{var}_{\mathscr{M}2}(S^2) \geqslant 0. \tag{3·5}$$

More specifically

$$\operatorname{var}_{\mathscr{M}2}(S^2) = \sigma^4\left(\frac{2}{N-1} + \frac{\gamma_2}{N}\right), \tag{3·6}$$

with $\gamma_2 \geqslant -2$ as the coefficient of excess, i.e. the standardized fourth cumulant, of the distribution of Y under model $\mathscr{M}2$. Inserting (3·6) in (3·5) and neglecting the difference between $N-1$ and N we may note for later reference that

$$\text{spec.red. }\{V_{\mathscr{S}}\} = C_n f(2+\gamma_2). \tag{3·7}$$

We conclude from (3·4) and (3·5) that the $\mathscr{M}2$-based precision measure $V_{\mathscr{M}2}$ does not describe the true precision of $\hat{T}$ as well on the average as does $V_{\mathscr{S}}$. The obvious explanation is that the σ^2 of model $\mathscr{M}2$ is not the actual finite population variance. The random labelling model $\mathscr{M}1$ is therefore in principle a more attractive model than $\mathscr{M}2$, in that it only involves the real finite population variance S^2 and no hypothetical σ^2.

In practice, however, neither S^2 nor σ^2 are known, but both are estimated by the same statistic s^2, and in both models the inference becomes equivalent to sample randomization inference under simple random sampling. In particular $V_{\mathscr{S}}$, $V_{\mathscr{M}1}$ and $V_{\mathscr{M}2}$ have the same estimator, $v=(N^2/n)(1-f)s^2$, with specific reduction under $(\mathscr{S}, \mathscr{M}2)$ given up to a factor of order $1+O(1/n)$ by

$$\text{spec.red. }\{v\} = C_n\{-2+\gamma_2(1-2f)\}. \tag{3·8}$$

If the Y-values are generated by a normal distribution, say, for which $\gamma_2=0$,

$$\text{spec.red. }\{v\} \leqslant \text{spec.red. }\{V_{\mathscr{M}2}\} \leqslant \text{spec.red. }\{V_{\mathscr{S}}\}.$$

A simple explanation behind the first of these inequalities is that under normality s^2 and $(\bar{y}-\bar{Y})^2$ are uncorrelated under $\mathscr{M}\,|\,\mathscr{S}$, so when going from $V_{\mathscr{M}2}$ to v there is only a negative contribution by the variance of v.

4. The ratio estimator under simple random sampling and proportional regression

Again the study population is $(Y_1,\ldots,Y_N)$, with total $T=N\bar{Y}$, and the design specifies that the choice of n units, or labels or indices, be made by simple random sampling, without replacement. We assume that there is also available a concomitant variable x, with positive population values $(X_1, \ldots, X_N)$, and a known population average $\bar{X}$. Consider the superpopulation model which specifies that the Y-values are independently generated, given the set of X-values, as

$$E_{\mathscr{M}}(Y_i | X_i) = \beta X_i, \quad \mathrm{var}_{\mathscr{M}}(Y_i | X_i) = \sigma^2 X_i. \tag{4·1}$$

This is the superpopulation model traditionally associated with the ratio estimator. We also assume that the conditional distributions have a common form, with coefficient of excess γ_2. The ratio estimator/predictor is defined by $\hat{T} = N\bar{y}\bar{X}/\bar{x}$, where $\bar{y}$ and $\bar{x}$ denote the sample means of Y- and X-values. This is an approximately unbiased estimator of T under simple random sampling, with good approximation when n is large, and it is an exactly unbiased, and even best linear unbiased, predictor of T under model (4·1); compare Royall (1970).

The approximate sampling variance, or mean squared error, of $\hat{T}$ is

$$V_{\mathscr{S}} = \mathrm{var}_{\mathscr{S}|\mathscr{M}}(\hat{T}) \simeq \frac{N^2(1-f)}{n} S^2_{Y|X}, \tag{4·2}$$

with

$$S^2_{Y|X} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - RX_i)^2, \quad R = \bar{Y}/\bar{X},$$

see e.g. Cochran (1977, p. 153). The predictive variance is (Royall, 1970)

$$V_{\mathscr{M}} = \mathrm{var}_{\mathscr{M}|\mathscr{S}}(\hat{T} - T) = \frac{N^2\sigma^2}{n} \frac{\bar{X}(\bar{X} - f\bar{x})}{\bar{x}}. \tag{4·3}$$

Several estimators of $V_{\mathscr{S}}$ and $V_{\mathscr{M}}$ have been proposed in the literature. Here six such statistics will be compared. Various subsets of these six have been discussed by, for example, Rao (1969), Royall & Eberhardt (1975), Royall & Cumberland (1978b, 1981a), Wu (1982), Wu & Deng (1983).

The 'classical' estimator of (4·2), approximately sampling-unbiased, is obtained by simply replacing the population variance in (4·2) by the corresponding sample variance and R by $\hat{R} = \bar{y}/\bar{x}$, with the result

$$v_0 = \frac{N^2(1-f)}{n} \frac{1}{n-1} \sum_s (y_i - \hat{R}x_i)^2. \tag{4·4}$$

Note that $E_{\mathscr{S}|\mathscr{M}}(\hat{R}) \simeq R$ and $E_{\mathscr{M}|\mathscr{S}}(\hat{R}) = E_{\mathscr{M}}(R) = \beta$. Royall & Eberhardt (1975) and Royall & Cumberland (1978b, 1981a) used the notation v_C. Under model (4·1), v_0 is not model-unbiased, but

$$v_H = v_0 \frac{\bar{X}(\bar{X} - f\bar{x})}{\bar{x}^2(1-f)(1-\eta_x^2/n)} \tag{4·5}$$

is, where η_x is the coefficient of variation of the x-sample; this was proposed by Royall & Eberhardt (1975). In his discussion of Royall & Cumberland (1978b), Rao (1978) advocated a design-motivated simpler alternative to v_H,

$$v_2 = v_0(\bar{X}/\bar{x})^2, \tag{4·6}$$

and gave literature references. Wu (1982) and Wu & Deng (1983) studied a wider class of estimators by replacing the exponent 2 in (4·6) by an arbitrary number g. Wu found

$$v_1 = v_0(\bar{X}/\bar{x}) \tag{4·7}$$

to be of particular interest since the exponent value $g = 1$ minimized the expected value under model (4·1) of the approximate sampling variance for this class of estimators. A related reasoning led Godambe & Thompson (1986) to propose the same estimator. Note that all estimators mentioned until now are approximately sampling-unbiased.

A more direct alternative to v_H as a model-unbiased estimator of $V_{\mathscr{M}}$ is obtained by simply replacing σ^2 in (4·3) by the weighted mean square of residuals estimator,

$$\hat{\sigma}_W^2 = \frac{1}{n-1} \sum_s \frac{(y_i - \hat{R}x_i)^2}{x_i}. \tag{4·8}$$

Following Royall & Eberhardt (1975) we denote the resulting statistic by v_W. Royall & Cumberland (1978b, 1981a) used the notation v_L.

The last statistic to be considered is obtained by replacing σ^2 in (4·3) by

$$\hat{\sigma}_D^2 = \frac{1}{n} \sum_s \frac{(y_i - \hat{R}x_i)^2}{\bar{x} - x_i/n}. \tag{4·9}$$

The resulting model-unbiased estimator v_D (Royall & Cumberland, 1981a) was proposed by Royall & Cumberland (1978a) under the notation G_2. Under mild restrictions on the x_i, v_D is easily shown to be approximately sampling-unbiased.

PROPOSITION. *For the ratio estimator/predictor under simple random sampling and the proportional regression model* (4·1), *the specific reductions for the theoretical variances* $V_{\mathscr{S}}$ *in* (4·2) *and* $V_{\mathscr{M}}$ *in* (4·3) *and the variance estimators* v_0, v_1, v_2, v_H, v_D *and* v_W, *given by* (4·4)–(4·9), *are as follows, up to factors of order* $1 + O(1/n)$:

$$\text{spec.red.}\,\{V_{\mathscr{S}}\} = C_n f(2+\gamma_2)(1+\eta_X^2), \quad \text{spec.red.}\,\{V_{\mathscr{M}}\} = \frac{C_n\eta_X^2}{1-f},$$

$$\text{spec.red.}\,\{v_W\} = \text{spec.red.}\,\{V_{\mathscr{M}}\} + C_n\{(1-2f)(2+\gamma_2) - 4(1-f)\},$$

$$\text{spec.red.}\,\{v_H\} = \text{spec.red.}\,\{V_{\mathscr{M}}\} + C_n\{(1-2f)(2+\gamma_2) - 4(1-f)\}(1+\eta_X^2),$$

$$\text{spec.red.}\,\{v_D\} = \text{spec.red.}\,\{v_H\}, \quad \text{spec.red.}\,\{v_0\} = \text{spec.red.}\,\{v_H\} - \frac{C_n\eta_X^2(2-f)^2}{1-f},$$

$$\text{spec.red.}\,\{v_1) = \text{spec.red}\,\{v_H\} - \frac{C_n\eta_X^2}{1-f}, \quad \text{spec.red.}\,\{v_2\} = \text{spec.red.}\,\{v_H\} - \frac{C_n\eta_X^2 f^2}{1-f},$$

where γ_2 *is the coefficient of excess of the distribution of* Y *given* X, η_X *is the coefficient of variation of the* X *population*, $\eta_X = S_X/\bar{X}$, *and*

$$C_n = N^4\sigma^4\bar{X}^2(1-f)^2/n^3.$$

The common term equals $2nC_n$, *except for a factor of order* $1 + O(1/n)$.

Derivations of these expressions are given in the Appendix. Model $\mathscr{M}2$ of § 3 appears for $X \equiv 1$, that is $\bar{X} = 1$ and $\eta_X = 0$, and formulae (3·4), (3·7) and (3·8) of that example are easily checked to be consistent with the present results. We continue by a list of conclusions apparent from a comparison of the various expressions, followed by a numerical example.

Comments and conclusions. (i) Neither $V_{\mathscr{S}}$ nor $V_{\mathscr{M}}$ dominates the other in specific reduction. The predictive variance $V_{\mathscr{M}}$ benefits from a small sampling fraction and, perhaps more surprising, also from a large fraction; $V_{\mathscr{S}}$ benefits from a small relative variation in x. As $\eta_X \to 0$ we approach the situation of § 3, in which the greater value for $V_{\mathscr{S}}$ was explained by the appearance of the superpopulation parameter σ in $V_{\mathscr{M}}$. For evaluation of the factor $(2+\gamma_2)$ in $V_{\mathscr{S}}$, note that $(2+\gamma_2) \geqslant 0$, and that the normal distribution corresponds to $\gamma_2 = 0$, whereas the rectangular distribution has $\gamma_2 = -1{\cdot}2$.

(ii) The exclusively sampling-motivated estimators v_0 and v_1 are uniformly inferior to the estimators v_H, v_D and v_2. For small f the difference in specific reduction between v_H and v_0 is approximately $4C_n \eta_X^2$, which may be of considerable size.

(iii) Variance estimators v_H and v_D are asymptotically equivalent with respect to specific reduction.

(iv) Estimator v_2 is quite close to v_H and v_D in specific reduction, in other words, the further refinement from v_2 that v_H represents is rarely worthwhile since their relative difference is of order f^2. See also § 5.

(v) Even the best estimators typically differ in specific reduction from the theoretical model-based variance $V_{\mathscr{M}}$ by a negative quantity, <0 at least for $\gamma_2 < 2$; compare with the end of § 3.

(vi) The usual negative quantity mentioned under (v) is expanded by a factor $(1+\eta_X^2)$ when we go from v_W to v_H or v_D, so usually v_W is better in mean squared error of predicted squared error than the other estimators. This is explained by the higher statistical efficiency of v_W. More precisely, Royall & Eberhardt (1975) remarked that

$$\frac{\operatorname{var}_{\mathscr{M}|\mathscr{S}}(v_H)}{\operatorname{var}_{\mathscr{M}|\mathscr{S}}(v_W)} \doteq 1 + \eta_x^2,$$

which partially explains the relationship between the specific reductions for v_W and v_H. They stated their result under the assumption $\gamma_2 = 0$, but it holds for arbitrary γ_2. However, we will see in § 5 that spec.red. $\{v_W\}$ is so sensitive to small modelling errors that v_W nevertheless cannot be recommended.

(vii) The specific reductions, and their mutual differences, are of a smaller magnitude than the common term. Hence, in large samples only marginal improvement in mean squared error of predicted squared error is possible by changing from one variance estimator to another.

Numerical illustration. Royall & Cumberland (1978b, 1981a) presented simulation studies of variance estimators in six real populations for which the ratio estimator/predictor was regarded as natural. Let us take these populations as examples for typical parameter values. Their η_X^2 values ranged almost from 0·5 to 3. No information was given about the γ_2 values. Let us assume $\gamma_2 = 0$. The authors used $f \doteq \frac{1}{4}$ for one of the populations with $\eta_X^2 \doteq 0{\cdot}6$, and otherwise $f \doteq \frac{1}{10}$. With such values we obtain the following specific reductions, after division by C_n:

for $f = \frac{1}{4}$, $\eta_X^2 = 0{\cdot}6$, $\gamma_2 = 0$,

$V_{\mathscr{S}}$: $+0{\cdot}8$, $V_{\mathscr{M}}$: $+0{\cdot}8$, v_W: $-1{\cdot}2$, $v_H \doteq v_D \doteq v_2$: $-2{\cdot}4$, v_1: $-3{\cdot}2$; v_0: $-4{\cdot}8$;

for $f = \frac{1}{10}$, $\eta_X^2 = 3$, $\gamma_2 = 0$,

$V_{\mathscr{S}}$: $+0{\cdot}8$, $V_{\mathscr{M}}$: $+3{\cdot}3$, v_W: $+1{\cdot}3$, $v_H \doteq v_D \doteq v_2$: $-4{\cdot}7$, v_1: $-8{\cdot}0$, v_0: $-16{\cdot}7$.

5. Robustness considerations for the ratio estimator/predictor

A finite population is likely to deviate in some respect from whatever simple superpopulation model is used to represent it. To what extent will the estimated precision assigned to an estimator/predictor be sensitive to such deviations? This was the concern of Royall & Cumberland (1978b, 1981a, b) in their empirical studies of variance estimators when they used real populations. We continue our study from § 4 of variance estimators for the ratio estimator by considering here the influence of deviations between the assumed model and the 'true' model.

Suppose there is an undetected intercept $\alpha > 0$ in the true relationship,

$$E_{\mathscr{M}}(Y_i | X_i) = \alpha + \beta X_i. \tag{5·1}$$

With an intercept it is no longer realistic to assume that the true variance of Y given X tends to zero with X, so suppose instead that

$$\operatorname{var}_{\mathscr{M}}(Y_i | X_i) = \tau^2 + \sigma^2 X_i. \tag{5·2}$$

It will turn out that the intercept in the variance function introduced is more crucial to our results that the intercept in the mean function.

Under (5·1) with $\alpha \neq 0$, $\hat{T}$ is no longer an $\mathscr{M}$-unbiased predictor. However, to the usual degree of approximation condition (2·6) will still hold for all our estimators v except v_W, since these estimators are approximately sampling-unbiased. Therefore the specific reductions are still given by the last right-hand terms of (2·7), except for $v = v_W$ which requires that a squared bias term is added to (2·7),

$$\text{bias}\,(v_W)^2 = E_{(\mathscr{S},\mathscr{M})}\{(\hat{T} - T)^2 - v_W\}^2.$$

Elementary approximate calculations show that

$$\text{bias}\,(v_W)^2 \simeq \frac{N^4}{n^2}(1-f)^2 \bar{X}^2 [\tau^2 \{\text{Ave}\,(X^{-1}) - \bar{X}^{-1}\} + \alpha^2 \{\text{Ave}\,(X^{-1}) - \bar{X}^{-1} - \eta_X^2 \bar{X}^{-1}\}]^2, \tag{5·3}$$

where Ave (X^{-1}) stands for the population mean of the X^{-1} values. Note that for fixed τ^2 or α this term is of order nC_n as $n \to \infty$. However, if we let $n \to \infty$ but want to mimic a finite sample situation with deviations from the assumed model that are not quite apparent, we must let these deviations shrink with n. More precisely, we should let $\tau^2 = O(n^{-\frac{1}{2}})$ and $\alpha = O(n^{-\frac{1}{2}})$, so that the intercepts may remain difficult to detect in sample data as $n \to \infty$. Doing so, we see from (5·3) that the intercept in the mean is less crucial than the intercept in the variance function, and that the latter will reduce spec.red. $\{v_W\}$ by a quantity of the same order of magnitude as the specific reductions themselves. Our conclusion is that we cannot rely on the result of § 4 as concerns v_W. A rarely detectable error in the variance function is sufficient to make v_W worse than the others. This result matches findings by Royall & Cumberland (1978b, 1981a). Already Royall & Eberhardt (1975) warned against the use of v_W for its lack of robustness.

In the rest of the section we will compare the other estimators with v_H. First, v_D is still asymptotically equivalent with v_H and need not be further discussed. The others, v_0, v_1 and v_2, are all of the form $v = g(\mathscr{s})v_H$, where $g(\mathscr{s})$ only depends on the sample. Remark 2 at the end of the Appendix shows how the calculations of the specific reductions must be modified in this case. The results are as follows:

$$\text{spec.red.}\,\{v_0\} = \text{spec.red.}\,\{v_H\} - \frac{C_n\eta_X^2(2-f)(2-f-2f\theta)}{1-f},$$

$$\text{spec.red.}\,\{v_1\} = \text{spec.red.}\,\{v_H\} - \frac{C_n\eta_X^2(1-2f\theta)}{1-f},$$

$$\text{spec.red.}\,\{v_2\} = \text{spec.red.}\,\{v_H\} - \frac{C_n\eta_X^2 f^2(1-2\theta)}{1-f},$$

where $\theta = \tau^2/(\tau^2+\sigma^2\bar{X}) \leqslant 1$ and, in the expression for C_n, $\sigma^2\bar{X}$ is replaced by the average modified error variance $\tau^2+\sigma^2\bar{X}$.

It is seen that, unless f is large, it is still the case that v_2 is better than v_1, which is better than v_0, as in § 4. The differences remain of the same magnitude unless θ is large. The previous small difference between v_H and v_2, of relative magnitude $O(f^2)$, remains small but changes sign for large θ. As $\theta \to 1$, v_2 becomes optimal.

Conclusion. If it appears that (4·1) is not too bad as a model for the data, use any of v_2, v_H or v_D.

6. Discussion

In this paper we have introduced and advocated the mean squared error of predicted squared error as a criterion for comparing variance estimators, be they design- or model-motivated. The criterion minimizes the expected squared difference between a variance estimator v and the squared error in $\hat{T}$, $(\hat{T}-T)^2$, so in conventional terminology v is not treated here as an estimator of a variance. However, since the role of the variance is to measure the squared error we see no conflict but only potential advantages in the present point of view.

As an example we investigated in § 4 the ratio estimator/predictor under simple random sampling and a proportional regression model. We found not only that the 'classical' variance estimator v_0 is uniformly inferior to the corresponding model-motivated estimator v_H and its close relatives v_D and v_2, but that the same negative conclusion holds for v_1, which has been advocated in the literature as having minimum sampling variance. That our conclusion was a different one is understood from the fact that the mean squared error of predicted squared error pays attention to the correlation between the variance estimator and the actual squared error. In § 5 we showed how robustness considerations could be incorporated in these calculations. In this way we could distinguish one statistic, v_W, as being nonrobust, in agreement with empirical observations.

The calculations of mean squared error of predicted squared error may be regarded as theoretical counterparts to empirical studies like those made by Royall & Cumberland (1978b, 1981a, b) and Wu & Deng (1983), where they compare the variance estimators with the actual variability of the ratio or regression estimators under repeated sampling in simulation studies on a number of real populations. The present measure has the advantage of allowing variation due to both repeated sampling and repeated generation of population values, instead of being confined to one or a few fixed populations. But the principal advantage of the measure itself lies in the possibility it opens for the statistician to select and motivate a precision statistic which is adapted to the expected amount of information in the actual sample, as expected under a reasonable model description of the population data structure. In this way it fits naturally within the framework of model-based inference. From a design-based point of view the consequences are even more significant, but not less satisfactory: the choice of a variance estimator thus adapted to the information in the sample is not in conflict with the randomization inference principle; the role of models is now only to assist in the selection. The present author prefers explicit modelling used in this way to the alternative implicit modelling of Robinson (1987), where the estimator v_2 in (4·6) is motivated as a conditionally adjusted estimator in an asymptotic design-based argument.

In the particular sense described above we obtain a unification of design- and model-based inference. If in a particular situation the statisticians can agree about the estimator/predictor $\hat{T}$, they should be able to agree about the function v of data used to assign a value to the precision of $\hat{T}$.

Future studies could concern other estimators/predictors or other sampling situations. A close relative to the ratio estimator is the regression estimator, $\hat{T}=N\{\bar{y}-\hat{\beta}(\bar{x}-\bar{X})\}$, for which we have strong reasons to anticipate results quite similar to those for the ratio estimator. Other examples could concern estimators/predictors in systematic sampling or in subsampling situations. A variance estimator is not only used per se, as a number indicating precision, but also as a standard error factor in standard confidence intervals under approximate normality. Empirical results by Wu & Deng (1983) and Deng & Wu (1987) indicate that variance estimators with lower mean squared error of predicted squared error tend to yield better coverage properties for the confidence intervals, but it remains to find convincing theoretical support for such findings. Finally, the idea of this paper might also be applied outside the field of sampling, in particular to situations with two different sources of randomness, for example to designed experiments with their random allocation of treatments to units and random experimental errors.

APPENDIX

Proof of the proposition of § 4

We start with the common term, $\text{var}_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2\}$, and make the calculation by conditioning on the sample s. First note that $\hat{T}-T$ may be written $\Sigma a_i\varepsilon_i$, where $\varepsilon_i=Y_i-\beta X_i$ $(i=1,\ldots,N)$ are mutually independent with

$$E_{\mathscr{M}|\mathscr{S}}(\varepsilon_i|X_i)=0,\quad \text{var}_{\mathscr{M}|\mathscr{S}}(\varepsilon_i|X_i)=\sigma^2X_i,\quad \text{var}_{\mathscr{M}|\mathscr{S}}(\varepsilon_i^2|X_i)=(2+\gamma_2)\sigma^4X_i^2,$$

and the coefficients a_i depend only on the sample, more precisely,

$$a_i=\begin{cases}(\bar{X}-f\bar{x})/(f\bar{x}) & \text{for } i\in s,\\ -1 & \text{for } i\notin s.\end{cases} \tag{A·1}$$

We apply the following lemma.

LEMMA 1. *For arbitrary coefficients* $\{a_i\}$, *allowed to depend on the sample,*

$$E_{\mathscr{M}|\mathscr{S}}\{(\Sigma a_i\varepsilon_i)^2\}=\sigma^2\Sigma a_i^2X_i,$$

$$\text{var}_{\mathscr{M}|\mathscr{S}}\{(\Sigma a_i\varepsilon_i)^2\}=2\sigma^4(\Sigma a_i^2X_i)^2+\gamma_2\sigma^4\Sigma(a_i^2X_i)^2.$$

Proof. The lemma is easily proved by writing

$$(\Sigma a_i\varepsilon_i)^2=\Sigma a_i^2\varepsilon_i^2+2\Sigma_{i<j}a_ia_j\varepsilon_i\varepsilon_j$$

and noting that all terms are mutually uncorrelated. □

Application of the lemma with a_i as in (A·1) yields the common term, $\text{var}_{(\mathscr{S},\mathscr{M})}\{(\hat{T}-T)^2\}$,

$$E_{\mathscr{S}}\left[2\sigma^4\left\{n\bar{x}\left(\frac{\bar{X}-f\bar{x}}{f\bar{x}}\right)^2+N\bar{X}-n\bar{x}\right\}^2\right.$$

$$\left.+\gamma_2\sigma^4\left\{n\text{ ave}(x^2)\left(\frac{\bar{X}-f\bar{x}}{f\bar{x}}\right)^4+N\text{ Ave}(X^2)-n\text{ ave}(x^2)\right\}\right]+\text{var}_{\mathscr{S}}\left(\frac{N^2\sigma^2\bar{X}^2}{n\bar{x}}\right),$$

where ave (x^2) and Ave (X^2) represent the sample and population averages of the x^2 and X^2 values, respectively. Here the crude approximations

$$E_{\mathscr{S}}\{g(\bar{x})\}=g(\bar{X})\{1+O(1/n)\},\quad \text{var}_{\mathscr{S}}(1/\bar{x})=O(1/n),$$

compare Lemma 2 below, are sufficient to conclude that the dominating term of the right-hand side is

$$2N^4\sigma^4(1-f)^2\bar{X}^2/n^2=2nC_n,$$

as was to be shown.

For $V_{\mathscr{S}}$ expression (4·2) is only approximate, but the difference is easily seen to be negligible. Hence, see (2·8), with satisfactory approximation its specific reduction is given by

$$\text{spec.red.}\{V_{\mathscr{S}}\}\simeq\text{var}_{\mathscr{M}}(V_{\mathscr{S}})=\frac{N^4}{n^2}(1-f)^2\,\text{var}_{\mathscr{M}}(S^2_{Y\mid X}).$$

Here we insert the sufficiently precise approximation

$$\text{var}_{\mathscr{M}}(S^2_{Y\mid X})\simeq\text{var}_{\mathscr{M}}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2\right)=\frac{f}{n}(2+\gamma_2)\sigma^4\bar{X}^2(1+\eta_X^2),$$

and the derivation for $V_{\mathscr{S}}$ is finished.

The specific reduction for $V_{\mathscr{M}}$ is obtained by inserting (4·3) in the right-hand equation (2·8),

$$\text{spec.red.}\{V_{\mathscr{M}}\}=\text{var}_{\mathscr{S}}(V_{\mathscr{M}})=(N^4\sigma^4\bar{X}^2/n^2)\,\text{var}_{\mathscr{S}}(\bar{X}/\bar{x}).$$

An expression for the remaining variance is given by the last formula of the next lemma, and insertion of this expression immediately yields the desired formula for spec.red. $\{V_{\mathscr{M}}\}$.

LEMMA 2: *Propagation of errors. For differentiable functions $g(\bar{x})$ bounded within the range of X-values, simple random sampling of x-values, and up to factors of order* $1+O(1/n)$,

$$E_{\mathscr{S}}\{g(\bar{x})\}\simeq g(\bar{X}),$$
$$\text{var}_{\mathscr{S}}\{g(\bar{x})\}\simeq g'(\bar{X})^2\,\text{var}_{\mathscr{S}}(\bar{x})=(1-f)g'(\bar{X})^2S_X^2/n.$$

In particular,

$$E_{\mathscr{S}}(\bar{X}/\bar{x})\simeq 1,\quad \text{var}_{\mathscr{S}}(\bar{X}/\bar{x})\simeq(1-f)\eta_X^2/n.$$

Proof. The lemma is easily proved by linearization of $g(\bar{x})$ around $\bar{X}$. □

Now, let $v_m=V_{\mathscr{M}}\hat{\sigma}^2/\sigma^2$ denote any model-unbiased estimator of $V_{\mathscr{M}}$. By (2·9) we require the difference between

$$2E_{\mathscr{S}}\,\text{cov}_{\mathscr{M}\mid\mathscr{S}}\{(\hat{T}-T)^2,v_m\}=2E_{\mathscr{S}}[V_{\mathscr{M}}\,\text{cov}_{\mathscr{M}\mid\mathscr{S}}\{(\hat{T}-T)^2,\hat{\sigma}^2\}]/\sigma^2, \tag{A·2}$$

and

$$E_{\mathscr{S}}\,\text{var}_{\mathscr{M}\mid\mathscr{S}}(v_m)=E_{\mathscr{S}}\{V^2_{\mathscr{M}}\,\text{var}_{\mathscr{M}\mid\mathscr{S}}(\hat{\sigma}^2)\}/\sigma^4. \tag{A·3}$$

In the variance, (A·3), $\hat{\sigma}^2$ may be replaced, with relative error $O(1/n)$, for each v_m under consideration by the ideal estimators

$$\hat{\sigma}_0^2=\frac{1}{n\bar{x}}\sum\nolimits_s(y_i-\beta x_i)^2=\frac{1}{n\bar{x}}\sum\nolimits_s\varepsilon_i^2 \tag{A·4}$$

in v_H and in v_D, and

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{s}\frac{(y_i-\beta x_i)^2}{x_i} = \frac{1}{n}\sum_{s}\frac{\varepsilon_i^2}{x_i} \tag{A·5}$$

in v_W. Given the sample s, statistics (A·4) and (A·5) have variances $(2+\gamma_2)(1+\eta_x^2)\sigma^4/n$ and $(2+\gamma_2)\sigma^4/n$, respectively. Hence, for v_H and v_D, (A·3) takes the form

$$C_n(2+\gamma_2)E_{\mathscr{S}}\left\{(1+\eta_x^2)\left(1-f\frac{\bar{x}}{\bar{X}}\right)^2 \Big/ (1-f)^2\right\} \simeq C_n(2+\gamma_2)(1+\eta_X^2).$$

The approximation involves an obvious extension of the expected value part of Lemma 2 to functions g of two statistics. For v_W we obtain the same expression except for the factor $(1+\eta_x^2)$.

Next, for the covariance term (A·2), if for $\hat{\sigma}^2$ we use $\hat{\sigma}_0^2$ as expressed in terms of ε_i in (A·4) or (A·5), only terms $a_i^2\varepsilon_i^2$ for $i \in s$ of $(\hat{T}-T)^2$ contribute to the covariance. Elementary calculations as in Lemma 1 yield, with a_s denoting the common value of a_i for $i \in s$,

$$\operatorname{cov}_{\mathscr{M}|\mathscr{S}}\{(\hat{T}-T)^2, \hat{\sigma}_0^2\} = (2+\gamma_2)\sigma^4 a_s^2 \operatorname{ave}(x^2)/\bar{x}$$

for v_H and v_D, and

$$\operatorname{cov}_{\mathscr{M}|\mathscr{S}}\{(\hat{T}-T)^2, \hat{\sigma}_0^2\} = (2+\gamma_2)\sigma^4 a_s^2 \bar{x}$$

for v_W. However, for the covariance between $(\hat{T}-T)^2$ and $\hat{\sigma}^2$ the approximation $\hat{\sigma}_0^2$ of $\hat{\sigma}^2$ is not sufficiently precise. This is because $\hat{\sigma}_0^2$ is obtained from $\hat{\sigma}^2$ by replacing $\hat{R}$ by β, and $\hat{R}$ and $\hat{T}$ are proportional. For v_W a refinement of $\hat{\sigma}_0^2$ by the term $-(\hat{R}-\beta)^2\bar{x} = -\bar{\varepsilon}^2/\bar{x}$ yields a contribution to $\operatorname{cov}_{\mathscr{M}|\mathscr{S}}\{(\hat{T}-T)^2, \hat{\sigma}_W^2\}$ of

$$-n^2 a_s^2 \operatorname{var}_{\mathscr{M}|\mathscr{S}}(\bar{\varepsilon}^2)/\bar{x} = -2a_s^2\sigma^4\bar{x}\{1+O(1/n)\}.$$

When forming $E_{\mathscr{S}}(.)$ of the resulting conditional covariance the crude first order approximation is again sufficient, with the result

$$2E_{\mathscr{S}}\operatorname{cov}_{\mathscr{M}|\mathscr{S}}\{(\hat{T}-T)^2, v_W\} = 2C_n\gamma_2(1-f). \tag{A·6}$$

By analogous but slightly more complicated calculations it can be shown that for v_H and v_D we obtain for (A·2) the right-hand side of (A·6) multiplied by the factor $(1+\eta_x^2)$.

Inserting the expressions for (A·2) and (A·3) derived above we obtain the desired formulae for the difference between spec.red. $\{V_{\mathscr{M}}\}$ and spec.red. $\{v_m\}$ when $v_m = v_H$, v_D and v_W, respectively.

The estimators v_0, v_1 and v_2 will be treated in parallel. We start from the specific reduction as expressed in (2·7); the unbiasedness requirement is satisfied to a sufficient order of magnitude. They may all be written in product form $g(s)v_H$, with only sampling variation in the factor $g(s)$. We apply the laws for propagation of errors on this product, that is an extension of Lemma 2 to functions of several variables, but noting the factors are correlated. To the desired order of magnitude we obtain after some calculations the remarkably simple result

$$\text{spec.red.}\{g(s)v_H\} \simeq \text{spec.red.}\{v_H\} - C_n n \operatorname{var}_{\mathscr{S}}\{g(s)\}. \tag{A·7}$$

Further propagation of errors through the nonlinear $g(s)$ by application of Lemma 2 yields: for v_0

$$n \operatorname{var}_{\mathscr{S}}\{g(s)\} \simeq \eta_X^2\frac{(2-f)^2}{1-f},$$

for v_1

$$n \operatorname{var}_{\mathscr{S}}\{g(s)\} \simeq \eta_X^2\frac{1}{1-f},$$

and for v_2

$$n\,\text{var}_{\mathscr{S}}\{g(s)\} \simeq \eta_X^2 \frac{f^2}{1-f}.$$

Inserting in (A·7) we find differences from spec.red. $\{v_H\}$ as asserted in the Proposition. □

Remark 1. As a by-product of (A·7) we see that v_H is optimal among all estimators obtained from v_H by a modification factor of type $g(s)$.

Remark 2. If the assumed error variance function is incorrect, var $(\varepsilon_i) = V_i \neq \sigma^2 X_i$, so v_H and the other estimators are no longer model-unbiased, (A·7) no longer holds, but we must add to its right-hand side a term

$$2\frac{N^2}{n}(1-f)\bar{V}\,\text{cov}_{\mathscr{S}}[g(s), E_{\mathscr{M}|\mathscr{S}}\{(\hat{T}-T)^2 - v_H\}].$$

Also $\sigma^2 \bar{X}$ in C_n should be replaced by $\bar{V}$.

Remark 3. In the derivations above we chose to condition on the sample s. Alternatively we could have conditioned on the population Y values. We would then have linearized the ratios and used the higher moments calculation technique developed by Tukey (1950) and Wishart (1952).

References

Cassel, C. M., Särndal, C. E. & Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*. New York: Wiley.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

Cumberland, W. G. & Royall, R. M. (1981). Prediction models and unequal probability sampling. *J. R. Statist. Soc.* B **43**, 353–67.

Deng, L.-Y. & Wu, C. F. J. (1987). Estimation of variance of the regression estimator. *J. Am. Statist. Assoc.* **82**, 568–76.

Godambe, V. P. & Thompson, M. E. (1986). Parameters of superpopulation and survey population: Their relationships and estimation. *Int. Statist. Rev.* **54**, 127–38.

Hansen, M. H., Madow, W. G. & Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sampling surveys (with discussion). *J. Am. Statist. Assoc.* **78**, 776–807.

Isaki, C. T. & Fuller, W. A. (1982). Survey design under the regression superpopulation model. *J. Am. Statist. Assoc.* **77**, 89–96.

Rao, J. N. K. (1969). Ratio and regression estimators. In *New Developments in Survey Sampling*, Ed. N. L. John and H. Smith, pp. 213–34. New York: Wiley.

Rao, J. N. K. (1978). Comments on papers by Basu and Royall and Cumberland. In *Survey Sampling and Measurement*, Ed. N. K. Namboodiri, pp. 323–9. New York: Academic Press.

Robinson, J. (1987). Conditioning ratio estimates under simple random sampling. *J. Am. Statist. Assoc.* **82**, 826–31.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57**, 377–87.

Royall, R. M. & Cumberland, W. G. (1978a). Variance estimation in finite population sampling. *J. Am. Statist. Assoc.* **73**, 351–8.

Royall, R. M. & Cumberland, W. G. (1978b). An empirical study of prediction theory in finite population sampling: Simple random sampling and the ratio estimator. In *Survey Sampling and Measurement*, Ed. N. K. Namboodiri, pp. 293–309. New York: Academic Press.

Royall, R. M. & Cumberland, W. G. (1981a). An empirical study of the ratio estimator and estimators of its variance (with discussion). *J. Am. Statist. Assoc.* **76**, 66–88.

Royall, R. M. & Cumberland, W. G. (1981b). The finite-population linear regression estimator and estimators of its variance—An empirical study. *J. Am. Statist. Assoc.* **76**, 924–30.

Royall, R. M. & Eberhardt, K. R. (1975). Variance estimates for the ratio estimator. *Sankhyā* C **37**, 43–52.

Särndal, C. E. (1978). Design-based and model-based inference in survey sampling (with discussion). *Scand. J. Statist.* **5**, 27–52.

Särndal, C. E. (1985). How survey methodologists communicate. *J. Offic. Statist.* **1**, 49–63.

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

SMITH, T. M. F. (1976). The foundations of survey sampling: A review (with discussion). *J. R. Statist. Soc.* A **139**, 183–204.

SMITH, T. M. F. (1984). Present position and potential developments: Some personal views. Sample surveys (with discussion). *J. R. Statist. Soc.* A **147**, 208–21.

TUKEY, J. W. (1950). Some sampling simplified. *J. Am. Statist. Assoc.* **45**, 501–19.

VALLIANT, R. (1990). Comparisons of variance estimators in stratified random and systematic sampling. *J. Offic. Statist.* **6**, 115–31.

WISHART, J. (1952). Moment coefficients of the *k*-statistics in samples for a finite population. *Biometrika* **39**, 1–13.

WU, C. F. (1982). Estimation of variance of the ratio estimator. *Biometrika* **69**, 183–9.

WU, C. F. J. & DENG, L. Y. (1983). Estimation of variance of the ratio estimator. An empirical study. In *Scientific Inference, Data Analysis, and Robustness*, Ed. G. E. P. Box et al., pp. 245–77. New York: Academic Press.

[*Received July* 1992. *Revised July* 1993]