Survey Methodology, June 1985
Vol. 11, No. 1, pp. 15–31
Statistics Canada

Conditional Inference in Survey Sampling

J.N.K. RAO[1]

ABSTRACT

Conventional methods of inference in survey sampling are critically examined. The need for conditioning the inference on recognizable subsets of the population is emphasized. A number of real examples involving random sample sizes are presented to illustrate inferences conditional on the realized sample configuration and associated difficulties. The examples include the following: estimation of (a) population mean under simple random sampling; (b) population mean in the presence of outliers; (c) domain total and domain mean; (d) population mean with two-way stratification; (e) population mean in the presence of non-responses; (f) population mean under general designs. The conditional bias and the conditional variance of estimators of a population mean (or a domain mean or total), and the associated confidence intervals, are examined.

KEY WORDS: Conditional inference; Conditional bias; Conditional variance; Population mean; Random sample sizes

1. INTRODUCTION

In the conventional set-up for inference in survey sampling the sample design defines the sample space S (set of possible samples s) and the associated probabilities of selection, $p(s)$. The choice of an estimator is based on the criterion of consistency or unbiasedness and on the comparison of mean square errors (MSE), under repeated sampling with probabilities $p(s)$, using the sample space S as the reference set. Thus, an estimator $\hat{\bar{Y}}$ of a population mean $\bar{Y}$ is unbiased if $E(\hat{\bar{Y}}) = \sum_{s \epsilon S} p(s)\hat{\bar{Y}}_s = \bar{Y}$, where $\hat{\bar{Y}}_s$ is the value of $\hat{\bar{Y}}$ for the sample s. The MSE of the estimator $\hat{\bar{Y}}$ is given by $\mathrm{MSE}(\hat{\bar{Y}}) = \sum_{s \epsilon S} p(s)(\hat{\bar{Y}}_s - \bar{Y})^2$, and $\hat{\bar{Y}}$ is consistent if its MSE approaches zero as the sample size increases. A consistent or unbiased estimator of $\mathrm{MSE}(\hat{\bar{Y}})$, denoted as $\mathrm{mse}(\hat{\bar{Y}})$, provides a measure of uncertainty in $\hat{\bar{Y}}$. If $\hat{\bar{Y}}$ is unbiased or consistent, then the observed values $\hat{\bar{Y}}_s$ and $\mathrm{mse}(\hat{\bar{Y}}_s)$ provide a large sample, $(1 - \alpha)$-level, confidence interval given by

$$I_s = \hat{\bar{Y}}_s \pm z_{\alpha/2}\sqrt{\mathrm{mse}(\hat{\bar{Y}}_s)}, \tag{1}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-point of a $N(0, 1)$ variable. The interpretation of (1) is that in repeated sampling with S as the reference set, approximately $100\,(1 - \alpha)\%$ of the intervals, I_s, will contain the true value $\bar{Y}$.

The comparison of unconditional mean square errors, $\mathrm{MSE}(\hat{\bar{Y}})$, is appropriate at the design stage, but the sample space S may not be the relevant reference set for inference after the sample s has been drawn, if the sample contains "recognizable subsets". The concept of recognizable subsets will be illustrated in subsequent sections through examples involving random sample sizes. The choice of relevant reference set, however, is not unique. In fact, the surveyed sample s can be viewed as unique in a real sense, but then no inference under a repeated sampling set-up can be made since the relevant reference set would contain a singleton (Holt and Smith 1979).

[1] J.N.K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6.

Conditional inference has attracted considerable attention and controversy in classical statistics since Fisher (1925). For instance, in testing for independence in a 2×2 table of counts, Fisher argued that the inference should be conditional on the observed row and column marginal totals even if the margins are not fixed by the design. Yates (1984) revived this problem. The choice of relevant reference set is not always clear-cut, but the following guidelines look reasonable: (1) A conditional procedure should be chosen *before* observing the data, especially in the public domain. (2) A conditioning partition of S should be chosen in such a way that the partition contains no (or little) information on the parameters of interest, i.e. the statistic indexing the partition should be an ancillary statistic (Cox and Hinkley 1974, p. 38). (3) If the sample sizes are random (e.g., domain sample sizes) and their population distribution is completely known (or at least partially known), then the inferences should be conditional on the observed sample sizes. In this context, Durbin (1969, p. 643) says "If the sample size is determined by a random mechanism and one happens to get a large sample one knows perfectly well that the quantities of interest are measured more accurately than they would have been if the sample size had happened to be small. It seems self-evident that one should use the information available on sample size in the interpretation of the result. To average over variations in sample size which might have occurred but did not occur, when in fact the sample size is exactly known, seems quite wrong from the standpoint of the analysis of the data actually observed".

The discussion throughout the paper will be confined to conditional inference in the presence of random sample sizes, as in guideline (3) above. Even with this restriction, it will be shown that conditional inferences are not always easy to implement in practice. We begin our discussion with simple examples and then extend it to more complex problems. In the context of sample surveys, Holt and Smith (1979) provide the most compelling arguments in favour of conditional inference, although their discussion was restricted to poststratification of a simple random sample (SRS); see Section 3.1.

Lahiri (1969) pointed out the "difficulties of conveying convincingly the real import of the sample survey estimates to intelligent but lay users of statistical data"; in particular, "the fallacy in implicitly using the (sampling) standard error as a measure of precision of the *observed* (sample) estimate, illustrating this point with a number of examples drawn from the current theory".

2. SIMPLE RANDOM SAMPLING WITH REPLACEMENT

Simple random sampling (SRS) with replacement is seldom used in practice, but it provides a simple introduction to conditional inference.

Suppose a simple random sample, s, of size n is selected fom a population of size N with replacement so that S contains N^n samples s. Let ν denote the number of distinct units in s. Then ν is a random variable with possible values $1, \ldots, n$. Let t_i denote the number of times the i-th population unit is included in s. Then two well-known estimators of the population mean $\bar{Y}$ are given by

$$\bar{y}_n = \frac{1}{n} \sum_{i \in s} t_i y_i, \tag{2.1}$$

the sample mean based on all the n draws, and

$$\bar{y}_\nu = \frac{1}{\nu} \sum_{i \in s} y_i, \tag{2.2}$$

the mean based on the distinct units in s. Both $\bar{y}_n$ and $\bar{y}_\nu$ are unconditionally unbiased under the reference set S, and the unconditional variance of $\bar{y}_\nu$ is always smaller than that of $\bar{y}_n$. Hence, from efficiency considerations $\bar{y}_\nu$ should be preferred over $\bar{y}_n$. The Horvitz-Thompson estimator

$$\bar{y}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} = \frac{\nu}{E(\nu)} \bar{y}_\nu \tag{2.3}$$

is also unconditionally unbiased, where π_i is the probability that unit i is included at least once in the sample:

$$\pi_i = \frac{E(\nu)}{N} = 1 - \left(1 - \frac{1}{N}\right)^n.$$

The comparison of variances of $\bar{y}_\nu$ and $\bar{y}_{HT}$ shows that $\bar{y}_\nu$ is not always better than $\bar{y}_{HT}$.

Following Durbin's (1969) argument, it is clear that for the purpose of inference one should condition on the observed value of ν, i.e., the relevant reference set is the set S_ν of $\binom{N}{\nu}$ samples of effective size ν, and not S. Fortunately, it is easy to implement conditional inference in this case since $P(s_\nu | \nu) = \binom{N}{\nu}^{-1}$, i.e. conditionally, the observed sample, s_ν, of distinct units is a simple random sample of size ν drawn without replacement. It follows that $\bar{y}_\nu$ is conditionally unbiased, i.e. $E_2(\bar{y}_\nu) = \bar{Y}$ where E_2 denotes conditional expectation, whereas $E_2(\bar{y}_{HT}) = [\nu/E(\nu)]\bar{Y} \neq \bar{Y}$ so that $\bar{y}_{HT}$ is conditionally biased. Hence, $\bar{y}_\nu$ should be preferred over $\bar{y}_{HT}$, despite the inconclusive comparison of unconditional variances. Note that $\bar{y}_{HT}$ would be a serious underestimate if the observed ν is much smaller that $E(\nu)$.

A relevant measure of uncertainty is the conditional variance, $V_2(\bar{y}_\nu)$, which is estimated unbiasedly by

$$v(\bar{y}_\nu) = \left(\frac{1}{\nu} - \frac{1}{N}\right) s_{\nu y}^2, \tag{2.4}$$

where $(\nu - 1)s_{\nu y}^2 = \sum_{i \in s} (y_i - \bar{y}_\nu)^2$ and V_2 denotes the conditional variance. The appropriate confidence interval for $\bar{Y}$ is given by

$$I_\nu = \bar{y}_\nu \pm z_{\alpha/2} \sqrt{v(\bar{y}_\nu)}. \tag{2.5}$$

Conditionally, the confidence level of I_ν is $1 - \alpha$ approximately if ν is not small. Another variance estimator

$$v^*(\bar{y}_\nu) = \left[E\left(\frac{1}{\nu}\right) - \frac{1}{N}\right] s_{\nu y}^2 \tag{2.6}$$

is conditionally biased, although unbiased when averaged over the whole sample space, S. It follows from (2.4) and (2.6) that $v(\bar{y}_\nu) < v^*(\bar{y}_\nu)$ if $1/\nu < E(1/\nu)$ and vice versa if $1/\nu > E(1/\nu)$. Thus, the confidence interval based on (2.6) would be too narrow if $E(1/\nu) < 1/\nu$ and hence yield a confidence level less than $1 - \alpha$, and too wide if $E(1/\nu) > 1/\nu$ leading to a confidence level greater than $1 - \alpha$. It may be noted that confidence intervals that are conditionally correct are automatically correct in the unconditional framework.

3. SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT

Suppose a simple random sample of fixed size n is drawn without replacement. In the absence of recognizable subsets, the relevant reference set is the set S of $\binom{N}{n}$ samples s, each of size n, and the sample mean $\bar{y}_n$ is unbiased and its variance is estimated unbiasedly by

$$v(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right)s_{ny}^2 \tag{3.1}$$

where $(n - 1)s_{ny}^2 = \sum_{i \in s} (y_i - \bar{y}_n)^2$. The resulting confidence interval is given by I_s: $\bar{y}_n \pm z_{\alpha/2}\sqrt{v(\bar{y}_n)}$ with confidence level $1 - \alpha$ approximately if n is not small.

Suppose now that recognizable subsets exist in the sense that we observe the sample configuration $\underset{\sim}{n} = (n_1, \ldots, n_k)$ belonging to k post-strata with known weights $W_i = N_i/N$. Ideally, stratified sampling should have been used but the strata frames were not available. The relevant reference set now is the set $S_{\underset{\sim}{n}}$ of $\prod\binom{N_i}{n_i}$ samples having the realized configuration $\underset{\sim}{n}$ since the distribution of $\underset{\sim}{n}$ is completely known.

3.1 All $n_i \geq 1$

If *all* the observed $n_i \geq 1$, then the customary post-stratified estimator

$$\bar{y}_{pst} = \sum W_i \bar{y}_i \tag{3.2}$$

is conditionally unbiased given $\underset{\sim}{n}$ since $P(s|\underset{\sim}{n}) = \prod\binom{N_i}{n_i}^{-1}$, i.e., conditionally the observed sample s is a stratified random sample $(s_1, \ldots, s_k)$ with strata sample sizes n_i. Here $\bar{y}_i$ denotes the sample mean in the i-th stratum. A relevant measure of uncertainty is the conditional variance, $V_2(\bar{y}_{pst})$, which is estimated unbiasedly by

$$v(\bar{y}_{pst}) = \sum W_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)s_{iy}^2, \tag{3.3}$$

provided *all* $n_i \geq 2$, where $(n_i - 1)s_{iy}^2 = \sum_{j \in s_i} (y_{ij} - \bar{y}_i)^2$ (Holt and Smith 1979). The resulting confidence interval, I_{pst}: $\bar{y}_{pst} \pm z_{\alpha/2}\sqrt{v(\bar{y}_{pst})}$, is conditionally correct. Another variance estimator

$$v^*(\bar{y}_{pst}) = \sum W_i^2\left[E\left(\frac{1}{n_i}\right) - \frac{1}{N_i}\right]s_{iy}^2 \tag{3.4}$$

$$\doteq \left(\frac{1}{n} - \frac{1}{N}\right)\sum W_i s_{iy}^2$$

is conditionally biased, although unbiased when averaged over the whole sample space, S (assuming that $P(n_i \leq 1)$ is negligible). The conditional performance of confidence interval based on (3.4) evidently depends on the extent of divergence of the observed values $1/n_i$ from their expectataions $E(1/n_i)$. It may be noted that the interval I_{pst} is also correct in the unconditional framework, provided $P(n_i \leq 1)$ is negligible for all i.

If $n_i = 1$ for some i, no conditionally unbiased variance estimator can be obtained, but it might be satisfactory to use a collapsed strata method or use the model-based solution of Hartley *et al.* (1969) originally proposed for variance estimation in stratified random sampling with one unit per stratum. Empirical studies might throw some light on the applicability of the latter methods.

The customary justification for preferring $\bar{y}_{pst}$ over $\bar{y}$ is that the unconditional variance of $\bar{y}_{pst}$ is approximately equal to the variance under proportional allocation and hence smaller than the unconditional variance of $\bar{y}$. We are also reminded that gains in efficiency under proportional allocation are likely to be modest. It is more important, however, to note that the sample mean $\bar{y}$ is conditionally biased:

$$E_2(\bar{y}) = \sum w_i \bar{Y}_i \neq \sum W_i \bar{Y}_i = \bar{Y}, \; w_i = \frac{n_i}{n}, \tag{3.5}$$

and hence the resulting inferences could be conditionally incorrect.

Example 1. Suppose $k = 2$ (say, male, female strata with known projected census weights W_1 and $W_2 = 1 - W_1$, or small and big hospitals (Royall 1970)). Royall used a super-population model

$$E_m(y_i) = \beta x_i, \; i = 1, \ldots, N, \; \beta > 0, \; x_i > 0 \tag{3.6}$$

to demonstrate that $\bar{y}$ is model-biased conditionally, where E_m denotes the model expectation, i.e.,

$$E_m(\bar{y}) = \beta \bar{x} \neq E_m(\bar{Y}) = \beta \bar{X} \tag{3.7}$$

unless the sample mean $\bar{x}$ coincides with the population mean $\bar{X}$. In his example, x_i = number of beds in the i-th hospital, y_i = number of occupied beds in the i-th hospital, and $x_1, \ldots, x_N$ are known. Royall argues that $\bar{y}$ leads to serious underestimation if the observed sample contains all (or mostly) small hospitals since $B_m(\bar{y}) = E_m(\bar{y}) - E_m(\bar{Y}) = \beta(\bar{x} - \bar{X})$ and $\bar{x} << \bar{X}$. This point can also be illustrated in our conditional framework without assuming a model. The ratio of the conditional bias of $\bar{y}$ to the population of large hospitals, $\bar{Y}_2$, may be expressed as

$$\frac{B_2(\bar{y})}{\bar{Y}_2} = (W_1 - w_1)\delta = (w_2 - W_2)\delta, \tag{3.8}$$

where $B_2(\bar{y}) = E_2(\bar{y}) - \bar{Y}$ denotes the conditional bias of $\bar{y}$, $\delta = (\bar{Y}_2 - \bar{Y}_1)/\bar{Y}_2$ and $0 < \delta < 1$ since the population mean, $\bar{Y}_1$, of small hospitals is smaller than $\bar{Y}_2$. If $w_1 = 1$ (i.e., all small hospitals observed in the sample), then $E_2(\bar{y}) = \bar{Y}_1 << \bar{Y}$ and hence $\bar{y}$ is a serious underestimate. Similarly, if $w_1 >> W_1$ (i.e., mostly small hospitals observed), then it follows from (3.8) that $\bar{y}$ would lead to serious underestimation.

In this example, one should use the post-stratified estimator $\bar{y}_{pst} = W_1\bar{y}_1 + W_2\bar{y}_2$ which is conditionally unbiased unless $n_1 = 0$ or $n_2 = 0$. It might be preferable, in fact, to use a post-stratified ratio estimator

$$\bar{y}_{pst,r} = \frac{\bar{y}_{pst}}{\bar{x}_{pst}} \bar{X}, \tag{3.9}$$

where $\bar{x}_{pst} = W_1\bar{x}_1 + W_2\bar{x}_2$ and $\bar{x}_i$ is the sample mean of x in the i-th stratum. The estimator (3.9) is approximately unbiased conditionally and more efficient than $\bar{y}_{pst}$ if n is large.

Remark 1. In Royall's example, one should, in fact, use a more efficient design than simple random sampling since all the population x-values are known, e.g., stratified random sampling under x-stratification and, perhaps, optimal allocation based on the x-values.

Remark 2. Royall justifies the use of the customary ratio estimator $\bar{y}_r = (\bar{y}/\bar{x})\bar{X}$ under his model (3.6), but it cannot be justified in the conditional (repeated sampling) framework since $\bar{y}_r$ is conditionally biased:

$$B_2(\bar{y}_r) \doteq \bar{X}\left[\frac{w_2\bar{Y}_1 + w_2\bar{Y}_2}{w_1\bar{X}_1 + w_2\bar{X}_2} - R\right], R = \frac{\bar{Y}}{\bar{X}} \tag{3.10}$$

$$\neq 0$$

unless $\bar{y}_1/\bar{x}_1 = \bar{y}_2/\bar{x}_2 = R$. In the extreme case of $w_1 = 1$, $B_2(\bar{y}_r) = \bar{X}(R_1 - R)$ where $R_1 = \bar{Y}_1/\bar{X}_1$. Hence, $B_2(\bar{y}_r) \lesseqgtr 0$ according as $R_1 \lesseqgtr R$.

Remark 3. If the weight W_1 is unknown but $\bar{X}$ is known, we cannot implement either $\bar{y}_{pst}$ or $\bar{y}_{pst,r}$. Royall suggests the use of $\bar{y}_r$ with inference conditional on the observed mean $\bar{x}$. However, the choice $\bar{x}$ is somewhat arbitrary, and the conditional bias of $\bar{y}_r$ could be quite large unless the model (3.6) is true, at least approximately.

If good prior information on W_1 is available, say $W_1^* \leq W_1 \leq W_1^{**}$ where W_1^* and W_1^{**} are known, then one could use the following "pseudo" post-stratified estimator of $\bar{Y}$:

$$\bar{y}_{pst}^* = \tilde{W}_1\bar{y}_1 + \tilde{W}_2\bar{y}_2, \tag{3.11}$$

where $\tilde{W}_1 = w_1$ if $W_1^* \leq w_1 \leq W_1^{**}$, $= W_1^*$ if $w_1 < W_1^*$, $= W_1^{**}$ if $w_1 > W_1^{**}$ and $\tilde{W}_2 = 1 - \tilde{W}_1$. The estimator $\bar{y}_{pst}^*$ and its ratio analogue should perform better conditionally given (n_1, n_2) than $\bar{y}$ and $\bar{y}_r$, although biased. Unconditionally, the MSE of $\bar{y}_{pst}^*$ should be smaller than the MSE of $\bar{y}$, provided $W_1^* \leq W_1 \leq W_1^{**}$. One could also utilize a formal Bayesian approach to estimate W_1 by specifying a prior distribution on W_1.

Example 2 (outliers). The problem of estimating a population mean $\bar{Y}$ in the presence of outliers is similar to the hospital example above. Suppose the population is known to contain a small fraction, W_2, of outliers (large observations) but W_2 is unknown, i.e. $W_1 \gg W_2$ and $\bar{Y}_2 \gg \bar{Y}_1$. Then, if the observed sample contains no outliers (i.e., $w_2 = 0$), we would say that $\bar{y}$ is "far from the true value $\bar{Y}$" (Chinnappa 1976) and yet $\bar{y}$ is (unconditionally) unbiased. The meaning of this statement follows from the fact that $E_2(\bar{y}) = \bar{Y}_1 \ll \bar{Y}$, where E_2 is the conditional expectation as before.

On the other hand, we would say that $\bar{y}$ is a serious overestimate if the sample contains outliers. This follows from (3.8) noting that $w_2 \gg W_2$ (since W_2 is very small). For instance, if $N_2 = 1$ then $w_2 = 1/n \gg W_2 = 1/N$. In this situation, we are told to modify the estimate $\bar{y}$ by reducing the weight attached to outliers in the sample. One suggestion is to modify $\bar{y}$ by reducing the weight attached to outliers from $1/n$ to $1/N$ and adjusting the weights for non-outliers such that the n weights sum to 1:

$$\bar{y}^* = \frac{N - n_2}{N}\bar{y}_1 + \frac{n_2}{N}\bar{y}_2. \tag{3.12}$$

The conditional relative bias of $\bar{y}^*$ is given by

$$\frac{B_2(\bar{y}^*)}{\bar{Y}_2} = \left(w_2\frac{n}{N} - W_2\right)\delta, \tag{3.13}$$

whereas $B_2(\bar{y})/\bar{Y}_2 = (w_2 - W_2)\delta$. If $w_2 \frac{n}{N} - W_2 < 0$, then

$$\left| w_2 \frac{n}{N} - W_2 \right| = W_2 - w_2 \frac{n}{N} < w_2 - W_2 \text{ if } 2W_2 < w_2\left(1 + \frac{n}{N}\right).$$

The inequality $2W_2 < w_2(1 + n/N)$ should be satisfied since $w_2 >> W_2$. If $w_2n/N - W_2 > 0$, then

$$\left| w_2 \frac{n}{N} - W_2 \right| = w_2 \frac{n}{N} - W_2 < w_2 - W_2.$$

Hence, the estimator $\bar{y}^*$ should have a smaller absolute value of conditional bias than $\bar{y}$.

The estimator $\bar{y}^*$ is essentially obtained from the post-stratified estimator $\bar{y}_{pst}$ by pretending that $N_2 = n_2$. A more satisfactory solution can be obtained by gathering good prior information on $W_1(= 1 - W_2)$, say from census data, and then using the estimator $\bar{y}^*_{pst}$ or the estimator based on a Bayes estimator of W_1.

Hidiroglou and Srinath (1981) derived the conditional bias and conditional and unconditional MSE of $\bar{y}$, $\bar{y}^*$ and some other modifications of $\bar{y}$, but they did not compare the conditional biases of $\bar{y}$ and $\bar{y}^*$ as above.

3.2 $n_i = 0$ for Some i

If the total sample size, n, is small or if too many post-strata chosen, then n_i could be zero for some i. The post-stratified estimator (3.2) in this case reduces to

$$\bar{y}_{pst} = \Sigma' W_i\bar{y}_i, \tag{3.14}$$

where Σ' denotes summation over strata with nonzero n_i. The estimator (3.14) is conditionally biased:

$$E_2(\bar{y}_{pst}) = \Sigma' W_i\bar{Y}_i \neq \Sigma\, W_i\bar{Y}_i. \tag{3.15}$$

It remains conditionally biased even under the strong assumption $\bar{Y}_i = \bar{Y}$ for all i, which incidentally shows that $\bar{y}_{pst}$ could lead to serious underestimation. It is also unconditionally biased. One commonly used method to overcome these difficulties is to collapse similar strata to ensure that $n_i > 0$ for all i in the reduced set of strata. Fuller (1966) proposed a more efficient solution for the special case of $k = 2$ post-strata, but his framework is unconditional in the sense that the probability, P_1^*, of $n_1 = 0$ given that either $n_1 = 0$ or $n_2 = 0$, is brought into the picture. His estimator is given by

$$\begin{aligned} \bar{y}_F &= \frac{W_1}{P_1^*}\bar{y}_1 \text{ if } n_2 = 0 \\ &= \frac{W_2}{P_2^*}\bar{y}_2 \text{ if } n_1 = 0, \end{aligned} \tag{3.16}$$

where $P_2^* = 1 - P_1^*$. The estimator $\bar{y}_F$ is conditionally unbiased given that either $n_1 = 0$ or $n_2 = 0$, but is conditionally biased given (n_1, n_2), even in the case $\bar{Y}_1 = \bar{Y}_2 = \bar{Y}$.

An unconditionally unbiased estimator is given by

$$\bar{y}_D = \Sigma \frac{a_i}{E(a_i)} W_i\bar{y}_i, \tag{3.17}$$

(Doss *et al.*, 1979), where $a_i = 1$ if at least one unit from stratum i in the sample, $= 0$ otherwise, and $\bar{y}_i$ is defined as $\bar{Y}_i$ if $n_i = 0$ (note that $a_i\bar{y}_i = 0$ if $n_i = 0$ even though $\bar{Y}_i$ is unknown). The estimator $\bar{y}_D$, however, is conditionally biased since

$$E_2(\bar{y}_D) = \sum{}' \frac{W_i\bar{Y}_i}{E(a_i)} \neq \sum W_i\bar{Y}_i = \bar{Y}.$$

It remains conditionally biased even if $\bar{Y}_i = \bar{Y}$ for all i.

Doss *et al.* criticized $\bar{y}_D$ on the grounds that it is not translation-invariant (i.e., $\bar{y}_D$ does not change to $\bar{y}_D + c$ when each y_i is changed to $y_i + c$, where c is an arbitrary constant), and hence that the variance of $\bar{y}_D$, when y_i is changed to $y_i + c$, can be made arbitrarily large by increasing c sufficiently. On the other hand, the ratio estimator

$$\bar{y}_{rD} = \frac{\sum \frac{a_i}{E(a_i)} W_i\bar{y}_i}{\sum \frac{a_i}{E(a_i)} W_i}, \tag{3.18}$$

proposed by Doss *et al.*, is translation-invariant. It is conditionally biased, but the conditional bias is approximately zero if $\bar{Y}_i = \bar{Y}$ for all i, unlike the conditional bias of $\bar{y}_D$. Another ratio estimator which is similar to $\bar{y}_{rD}$ conditionally is given by

$$\bar{y}_{r(pst)} = \frac{\sum' W_i\bar{y}_i}{\sum' W_i}, \tag{3.19}$$

but it is inconsistent unconditionally, unlike $\bar{y}_{rD}$. Hence, $\bar{y}_{rD}$ may be preferred to $\bar{y}_{r(pst)}$ or $\bar{y}_D$.

If concomitant information on *all* strata is available, then one could fit a model to the observed strata means $\bar{y}_i$ and predict the population means of strata with $n_i = 0$. For example, if the population means $\bar{X}_i$ of a concomitant variable are linearly related to the corresponding $\bar{Y}_i$, then the predicted value of a $\bar{Y}_i$ is given by $\hat{\alpha} + \hat{\beta}\bar{X}_i = \bar{y}_i^*$ (say), where $\hat{\alpha}$ and $\hat{\beta}$ are the least squares estimators obtained by minimising $\sum'(\bar{y}_i - \alpha - \beta\bar{X}_i)^2$. The resulting estimator of $\bar{Y}$ is given by

$$\bar{y}^*_{pst} = \sum{}' W_i\bar{y}_i + \sum{}'' W_i\bar{y}_i^*, \tag{3.20}$$

where $\sum''$ denotes summation over strata with $n_i = 0$. This estimator should have good conditional properties if the fitted model is adequate. It should be clear from this discussion that there is no simple solution if $n_i = 0$ for some of the strata.

4. TWO-WAY STRATIFICATION

Ingenious designs to improve the efficiency of estimators have been proposed in the literature. Bryant *et al.* (1960) proposed a design involving two-way stratification in which the sample sizes n_{ij} are zero for some strata (cells). Their method is supposed to permit estimation of the population mean even when the total sample size n is less than the total number of strata. Using proportional allocation for the marginal sample sizes ($n_{i.}$, $n_{.j}$), they obtained a random allocation n_{ij} such that $E(n_{ij}) = (n_{i.}n_{.j})/n = nW_{i.}W_{.j}$, where $W_{i.}$ and $W_{.j}$ are the row and column marginal totals of cell weights W_{ij}.

Bryant *et al.* proposed the estimator

$$\bar{y}_U = \frac{1}{n} \sum\sum n_{ij}G_{ij}\bar{y}_{ij}, \tag{4.1}$$

where $G_{ij} = n^2 W_{ij}/(n_{i.} n_{.j})$ and $\bar{y}_{ij}$ may be taken as $\bar{Y}_{ij}$ if $n_{ij} = 0$. The estimator $\bar{y}_U$ is unconditionally unbiased. However, the distribution of n_{ij} is completely known (since all W_{ij} are known) and hence the relevant reference set is the set of samples having the observed configuration $\{n_{ij}\}$, i.e., one should treat the design as stratified simple random sampling for inference purposes. The estimator $\bar{y}_U$ is conditionally biased:

$$E_2(\bar{y}_U) = \sum \sum (\frac{n_{ij}G_{ij}}{n})\bar{Y}_{ij} \neq \sum \sum W_{ij}\bar{Y}_{ij} = \bar{Y},$$

noting that $E_2(\bar{y}_{ij}) = \bar{Y}_{ij}$ if $n_{ij} > 0$. It also has the defects of $\bar{y}_D$ in the previous section which can be circumvented by using the ratio estimator

$$\bar{y}_r = \frac{\bar{y}_U}{\bar{a}_U} = \frac{\sum \sum n_{ij}G_{ij}\bar{y}_{ij}}{\sum \sum n_{ij}G_{ij}} \tag{4.2}$$

where $\bar{a}_U = \sum \sum n_{ij}G_{ij}/n$. $\bar{y}_r$ is also conditionally biased, but the conditional bias is approximately zero if $\bar{Y}_{ij} = \bar{Y}$ for all (i, j). The latter condition, however, may be unrealistic in the present context since the strata are different by design.

As in Section 3.1, it seems necessary to use a model connecting the sampled and nonsampled strata. A reasonable model, in the absence of concomitant information, is to assume that

$$y_{ijk} = \mu + \beta_j + \tau_i + \varepsilon_{ijk} \tag{4.3}$$

where y_{ijk} is the k-th observation in the (i, j)-th cell, β_j and τ_i are fixed effects and ε_{ijk} are independent errors with zero mean and common variance σ^2. Unfortunately, the linear combination $\mu + \beta_j + \tau_i$ for nonsampled strata is not estimable from sample data and hence the corresponding $\bar{Y}_{ij}$ cannot be predicted. This difficulty can be avoided by assuming that β_j and τ_i are random variables and then obtaining a predictor $\hat{\mu} + \hat{\beta}_j + \hat{\tau}_i$, but the random effects model may be less realistic than (4.3) in the present context.

Motivated by the above-mentioned difficulty, Bankier (1985) discussed a raking procedure in the context of independent stratified samples according to two different criteria of stratification. His estimator is approximately model-unbiased under the fixed effects model (4.3), while the usual Horvitz-Thompson estimator and its ratio extension are model-biased.

Bankier's method can be adapted to the two-way stratification problem. The raking ratio estimator of $\bar{Y}$ is given by

$$\bar{y}(p) = \sum \sum \frac{G_{ij}(p)}{n} y_{ij} \tag{4.4}$$

where y_{ij} is the sample total in the (i, j)-th cell ($y_{ij} = 0$ of $n_{ij} = 0$) and $G_{ij}(p)$ are the values obtained in the p-th iteration of the raking procedure such that

$$\sum_j \frac{G_{ij}(p)}{n} n_{ij} \doteq W_{i.} = \sum_j W_{ij} \tag{4.5}$$

and

$$\sum_i \frac{G_{ij}(p)}{n} n_{ij} \doteq W_{.j} = \sum_i W_{ij}.$$

The $G_{ij}(p)$ are obtained as follows: Let $G_{ij}(0) = G_{ij} > 0 \; \forall (i, j)$, and

$$G_{ij}(p) = G_{ij}(p-1) \frac{W_{i.}}{\sum_j \frac{G_{ij}(p-1)}{n} n_{ij}} \quad \textit{if } p \text{ is odd} \tag{4.6}$$

$$= G_{ij}(p-1) \frac{W_{.j}}{\sum_i \frac{G_{ij}(p-1)}{n} n_{ij}} \quad \textit{if } p \text{ is even.}$$

Under the fixed effects model (4.3), we have

$$E_m[\bar{y}(p)] \doteq \mu + \sum_i W_{i.}\tau_i + \sum_j W_{.j}\beta_j = E_m(\sum \sum W_{ij}\bar{Y}_{ij})$$

$$= E_m(\bar{Y}),$$

i.e. $\bar{y}(p)$ is approximately model-unbiased. Since $E(G_{ij}(0)n_{ij}/n) = W_{ij}$ for the choice $G_{ij}(0) = G_{ij}$, these starting values should be good. However, we may encounter convergence problems with the raking process because of the many empty cells ($n_{ij} = 0$) resulting from the Bryant *et al.* design. We hope to investigate these convergence problems as well as the conditional properties of the raking ratio estimator (4.4) in a separate paper.

If the population means $\bar{X}_{ij}$ of a concomitant variable x are known for *all* strata, then one could fit a model to the observed strata means $\bar{y}_{ij}$, as in Section 3.1. For example, the model $\bar{y}_{ij} = \beta\bar{x}_{ij} + b_j + t_i + \bar{\varepsilon}_{ij}$ with random effects b_j and t_i might be reasonable, where $\bar{\varepsilon}_{ij}$ is the sample mean of errors ε_{ijk} in the (i, j)-th cell. A predictor $\hat{\beta}\bar{x}_{ij} + \hat{b}_j + \hat{t}_i$ of $\bar{Y}_{ij}$ for nonsampled strata may be used in conjunction with the observed means $\bar{y}_{ij}$ to arrive at an estimator of $\bar{Y}$. This approach is similar to modelling for small area estimates, except that the parameter of interest here is the overall mean $\bar{Y}$ rather than the individual cell means $\bar{Y}_{ij}$. We hope to investigate the conditional properties of alternative estimators of $\bar{Y}$ in a separate paper.

5. NONRESPONSE

5.1 A Simple Model

Suppose m responses are obtained in a simple random sample of size n. Let W_1 denote the proportion in the response stratum and $\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$ the population mean, where $\bar{Y}_1$ and $\bar{Y}_2$ are the means of response and nonresponse strata respectively, and $W_2 = 1 - W_1$. In this situation, conditioning on the observed value of m can be questioned since the distribution of m depends on the unknown W_1 which is involved in the parameter of interest. Also, the sample mean $\bar{y}_m$ of respondents is unconditionally biased because $E(\bar{y}_m) = \bar{Y}_2 \neq \bar{Y}$. Hence, it is necessary to assume a model for response mechanism even in the unconditional framework, unless a subsample of nonrespondents is also sampled.

A simple model assumes that the probability of response if contacted is the same for all units, say p^*, i.e., data are missing at random. Under this model, the distribution of m depends only on p^*, and hence we should condition on m if p^* is assumed known (or at least partially known or unrelated to $\bar{Y}$). Oh and Scheuren (1983) have shown that conditionally given m the sample s_m of respondents is like a simple random sample of size m from the *whole* population. Hence, $\bar{y}_m$ is conditionally unbiased, and its conditional variance is unbiasedly estimated by

$$v_2(\bar{y}_m) = (m^{-1} - N^{-1})s^2_{my}, \tag{5.1}$$

where $(m - 1)s^2_{my} = \sum_{i \epsilon s_m}(y_i - \bar{y}_m)^2$. The resulting confidence interval $\bar{y}_m \pm z_{\alpha/2}\sqrt{v_2(\bar{y}_m)}$ is conditionally correct, at least approximately, if m is not small.

On the other hand, the Horvitz-Thompson estimator (p^* known):

$$\bar{y}_{HT} = \frac{m}{E(m)}\,\bar{y}_m = \sum_{i \epsilon s_m} \frac{y_i}{np^*} \tag{5.2}$$

is conditionally biased, as in Section 2, although unbiased when averaged over the distribution of m. For general designs, the ratio estimator

$$\hat{\bar{Y}}_{HT,r} = \frac{\sum\limits_{s_m} \dfrac{y_i}{\pi_i p_i^*}}{\sum\limits_{s_m} \dfrac{1}{\pi_i p_i^*}} \tag{5.3}$$

is often used on grounds of efficiency, where π_i is the probability of inclusion and p_i^* is the probability of response if contacted (assumed known) for the i-th unit. In the simple case of $p_i^* = p^*$ and simple random sampling, it is interesting to note that $\hat{\bar{Y}}_{HT,r}$ reduces to $\bar{y}_m$. Hence, the ratio estimator might perform well in a conditional framework, for general designs.

5.2 A More Realistic Model

A more realistic model assumes that data are missing at random within post-strata with known weights W_i. Let n_i and m_i respectively denote the sample size and the respondent sample size in the i-th post-stratum. Then the joint distribution of (n_i, m_i) depends only on the W_i and the response probabilities within post-strata. Hence, we should condition on the observed value of (n_i, m_i) provided the post-stratum response probabilities are either known or unrelated to the parameters of interest, viz., the post-strata means. Conditionally, the observed sample is like a stratified simple random sample with fixed strata sizes m_i (Oh and Scheuren 1983). Hence, the estimator

$$\bar{y}_{pst,m} = \sum W_i \bar{y}_{mi} \tag{5.4}$$

is conditionally unbiased, and its conditional variance is unbiasedly estimated by

$$v_2(\bar{y}_{pst,m}) = \sum W_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)s^2_{miy} \tag{5.5}$$

where $\bar{y}_{mi}$ and s^2_{miy} are the mean and variance of sample respondents in the i-th post-stratum, respectively.

If the W_i are unknown, it is a common practice to replace W_i in (5.4) by its estimate $w_i = n_i/n$. In this case, conditional inference can be questioned since the distribution of (n_i, m_i) depends on the unknown weights W_i and since W_i are involved in the parameter $\bar{Y} = \sum W_i\bar{Y}_i$. If partial information on W_i, in the form of bounds on W_i, is available, we can proceed with conditional inference as in Example 1, Remark 3, although the resulting estimator is still conditionally biased (but likely to be better than (5.4) with W_i replaced by w_i).

6. DOMAIN ESTIMATION (SRS)

6.1 Domain mean

Under simple random sampling (SRS), the usual estimator of a subpopulation (domain) mean, $\bar{Y}_i$, is given by the sample mean

$$\bar{y}_i = \sum_{j\epsilon s_i} \frac{y_j}{n_i}, \; n_i > 0 \tag{6.1}$$

where s_i is the sample falling in the domain and n_i is the corresponding size.

If the domain size, N_i, is known, then one should condition on the observed value, n_i. The estimator $\bar{y}_i$ is conditionally unbiased if $n_i > 0$ since conditionally s_i is a SRS sample of fixed size n_i from the domain. An unbiased estimate of the conditional variance is

$$v(\bar{y}_i) = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)s^2_{iy}, \; n_i > 0 \tag{6.2}$$

and the resulting confidence interval $\bar{y}_i \pm z_{\alpha/2}\sqrt{v(\bar{y}_i)}$ is conditionally correct.

The estimator $\bar{y}_i$, however, is unstable for small domains (small areas) with small n_i. Also $\bar{y}_i$ is not defined if $n_i = 0$. One solution to the latter problem, suggested in the literature, is to use a modified estimator.

$$\bar{y}_i' = \frac{a_i}{E(a_i)}\bar{y}_i, \; n_i \geq 0 \tag{6.3}$$

where $a_i = 1$ if $n_i \geq 1$; $= 0$ if $n_i = 0$ and $\bar{y}_i$ is taken as $\bar{Y}_i$ if $n_i = 0$. The estimator $\bar{y}_i'$, however, is conditionally biased:

$$E_2(\bar{y}_i') = \frac{a_i}{E(a_i)}\bar{Y}_i.$$

It is an underestimate if $n_i = 0$, and an overestimate if $n_i \geq 0$, although unconditionally unbiased. The extent of overestimation depends on the magnitude of $E(a_i) = P(n_i \geq 1)$. If, for example, $P(n_i \geq 1) = 0.75$, then $E_2(\bar{y}_i') = (4/3)\bar{Y}_i$ if $n_i \geq 1$.

Sarndal (1984) proposed the following estimator in the context of small area estimation:

$$\bar{y}_{iS} = \bar{y} + \frac{w_i}{W_i}(\bar{y}_i - \bar{y}), \; n_i \geq 0, \tag{6.4}$$

where $\bar{y} = \sum w_j \bar{y}_j$ is the overall sample mean and $w_i = n_i/n$. The estimator is approximately unconditionally unbiased, but conditionally biased unless $w_i = W_i$:

$$B_2(\bar{y}_{iS}) = (\frac{w_i}{W_i} - 1)(\bar{Y}_i - \bar{Y}'), \tag{6.5}$$

where $\bar{Y}' = \sum w_i \bar{Y}_i$. If $n_i = 0$, the estimator $\bar{y}_{iS}$ reduces to the "synthetic" estimator $\bar{y}$. The extent of under- (or over-) estimation of $\bar{y}_{iS}$ depends on both $w_i/W_i - 1$ and $\bar{Y}_i - \bar{Y}'$ and hence more complex to analyse than the bias of $\bar{y}_i'$. However, $\bar{y}_{iS}$ would have a larger absolute conditional bias* than $\bar{y}$ if $w_i > 2W_i$ (and hence a larger conditional MSE). Also, the conditionally unbiased estimator $\bar{y}_i$ has a smaller conditional variance than $\bar{y}_{iS}$ if $w_i > W_i$ (neglecting the variance of $\bar{y}$ relative to that of $\bar{y}_i$) and hence smaller conditional MSE.

Hidiroglou and Sarndal (1985) proposed a modification of $\bar{y}_{iS}$:

$$\bar{y}_{iS}^{**} = \begin{cases} \bar{y}_i \text{ if } w_i \geq W_i \\ \bar{y}_{iS}^{*} = \bar{y} + \left(\frac{w_i}{W_i}\right)^2 (\bar{y}_i - \bar{y}) \text{ if } w_i < W_i. \end{cases} \tag{6.6}$$

The estimator $\bar{y}_{iS}^{**}$ is conditionally unbiased if $w_i \geq W_i$, while its conditional absolute bias is smaller than that of $\bar{y}$ if $w_i < W_i$. A motivation for $\bar{y}_{iS}^{**}$ is that the conditional variance of $\bar{y}_{iS}^{*}$ (or $\bar{y}_{iS}$) is larger than that of $\bar{y}_i$ (neglecting the variance of $\bar{y}$ relative to that of $\bar{y}_i$) if $w_i > W_i$, while the conditional variance of $\bar{y}_{iS}^{*}$ is smaller than that of $\bar{y}_{iS}$ if $w_i < W_i$. However, the absolute conditional bias of $\bar{y}_{iS}^{*}$ is larger than that of $\bar{y}_{iS}$ if $w_i < W_i$. Hence, the choice between $\bar{y}_{iS}^{*}$ and $\bar{y}_{iS}$ in the case $w_i < W_i$ is not clear-cut and no simple recipe seems to exist.

Drew *et al.* (1982) proposed another sample size dependent estimator which depends on a parameter K_0. In the SRS case and the choice $K_0 = 1$, their estimator reduces to

$$\bar{y}_{iD} = \begin{cases} \bar{y}_i \text{ if } w_i \geq W_i \\ \bar{y}_{iS} \text{ if } w_i < W_i. \end{cases} \tag{6.7}$$

As noted above, the choice between $\bar{y}_{iS}$ and $\bar{y}_{iS}^{*}$ in the case $w_i < W_i$ is not clear-cut. Consequently, the choice between $\bar{y}_{iD}$ and $\bar{y}_{iS}^{**}$ is also not clear-cut.

If N_i is unknown, the conditional argument may still be relevant provided N_i is unrelated to the parameter of interest $\bar{Y}_i$. It is also relevant when partial information on N_i is available, such as bounds on N_i.

If a concomitant variable x with known domain mean $\bar{X}_i$ is available, the ratio estimator

$$\bar{y}_{ir} = \frac{\bar{y}_i}{\bar{x}_i} \bar{X}_i \tag{6.8}$$

---

*Sarndal's estimator, however, should perform better in the case of a one-way model. The estimator is obtained by pooling estimators of the form (6.4) over two or more groups.

and a regression-type estimator (Battese and Fuller 1981)

$$\bar{y}_{ir}^{*} = \bar{y}_i + \frac{\bar{y}}{\bar{x}} (\bar{X}_i - \bar{x}_i) \tag{6.9}$$

are both conditionally unbiased (approximately), but $\bar{y}_{ir}^{*}$ is likely to be more efficient if a regression model (through the origin) with a common slope holds true, at least approximately, for the small areas. If the slopes are varying, then an empirical Bayes estimator, which is more complex, might be more relevant (Dempster *et al.* 1981).

6.2 Domain Total

If N_i is known, then an estimate of domain total $Y_i = N_i \bar{Y}_i$ is simply obtained by multiplying a chosen estimator of $\bar{Y}_i$ by N_i. On the other hand, the usual unbiased estimator

$$\hat{Y}_i = \hat{N}_i \bar{y}_i = \frac{N}{n} \sum_{j \in s_i} Y_j , \ n_i \geq 1 \tag{6.9}$$

is used if N_i is unknown, where $\hat{N}_i = N w_i$ is the unbiased estimator of N_i and $P(n_i = 0)$ is assumed to be negligible.

Suppose now that we have prior information, say $N_i^* \leq N_i \leq N_i^{**}$. Then the conditional argument may be relevant. The conditional bias of $\hat{Y}_i$ is

$$B_2(\hat{Y}_i) = (\hat{N}_i - N_i)\bar{Y}_i. \tag{6.10}$$

It follows from (6.10) (assuming $\bar{Y}_i > 0$) that $B_2(\hat{Y}_i) > 0$, i.e., overestimation, if $\hat{N}_i > N_i$ and that $B_2(\hat{Y}_i)$ increases as the domain sample size n_i increases. Similarly, $B_2(\hat{Y}_i) < 0$, i.e., underestimation, if $\hat{N}_i < N_i$ and $|B_2(\hat{Y}_i)|$ increases as n_i decreases; the conditional bias is zero if $\hat{N}_i = N_i$.

Utilizing the prior information, we can modify $\hat{Y}_i$ as

$$\hat{Y}_i^* = \begin{cases} N_i^* \bar{y}_i & \text{if } \hat{N}_i < N_i^* \\ \hat{N}_i \bar{y}_i & \text{if } N_i^* \leq \hat{N}_i \leq N_i^{**} \\ N_i^{**} \bar{y}_i & \text{if } \hat{N}_i > N_i^{**}. \end{cases} \tag{6.11}$$

The absolute conditional bias of $\hat{Y}_i^*$ is smaller than that of $\hat{Y}_i$ if either $\hat{N}_i < N_i^*$ or $\hat{N}_i > N_i^{**}$, while $\hat{Y}_i^* = \hat{Y}_i$ in the interval $N_i^* \leq \hat{N}_i \leq N_i^{**}$. Hence, $\hat{Y}_i^*$ is conditionally better than the unbiased estimator $\hat{Y}_i$. Also the unconditional MSE of $\hat{Y}_i^*$ is smaller than that of $\hat{Y}_i$, although $\hat{Y}_i^*$ is unconditionally biased. Unfortunately, there is no simple way to improve upon $\hat{Y}_i^*$ in the range $N_i^* \leq \hat{N}_i \leq N_i^{**}$. In any case, $\hat{Y}_i^*$ should be preferred over $\hat{Y}_i$. Good supplementary information on the domain size is necessary in estimating a domain total efficiently.

7. GENERAL DESIGNS

Post-stratification adjustment is commonly employed in complex large-scale surveys, mainly to increase the efficiency of estimators, e.g., the age-sex adjustment in the Canadian Labour Force Survey (LFS). A general theory of unconditional inference is also available.

The estimator of total Y is given by

$$\hat{Y}_{pst} = \sum M_i \frac{\hat{Y}_i}{\hat{M}_i} \tag{7.1}$$

where $\hat{Y}_i$ and $\hat{M}_i$ are the usual unbiased domain estimators of the i-th post-stratum total Y_i and size M_i respectively. In the LFS, projected census counts are used for the M_i. The estimator $\hat{Y}_{pst}$ reduces to $\sum N_i \bar{y}_i$ in the SRS case (see (3.2)) and we have already seen that $\sum N_i \bar{y}_i$ is conditionally unbiased in the SRS case (assuming all $n_i \geq 1$). However, for complex designs it seems difficult to investigate the conditional properties of (7.1); even the choice of reference set is not so clear-cut. To illustrate this difficulty, consider stratified SRS with $L = 2$ strata and $k = 2$ post-strata. If we condition on the observed post-strata sample sizes (n_{h1}, n_{h2}) in each stratum h, the theory is straightforward provided the post-strata sizes N_{hi} in each stratum are known. However, in practice we will run into problems with zero sample sizes n_{hi} and also the sizes N_{hi} in each stratum may not be available or the projections inaccurate, although $N_{.i} = \sum_h N_{hi} = M_i$ are available. Hence, we may prefer to condition on the observed total sample sizes $(n_{.1}, n_{.2})$, where $n_{.i} = \sum_h n_{hi}$.

The estimator $\hat{Y}_{pst}$ in this special case of stratified SRS ($L = 2$, $k = 2$) reduces to

$$\hat{Y}_{pst} = N_{.1} \frac{N_{1.} \frac{y_{11}}{n_{1.}} + N_{2.} \frac{y_{21}}{n_{2.}}}{N_{1.} \frac{n_{11}}{n_{1.}} + N_{2.} \frac{n_{21}}{n_{2.}}} + N_{.2} \frac{N_{1.} \frac{y_{12}}{n_{1.}} + N_{2.} \frac{y_{22}}{n_{2.}}}{N_{1.} \frac{n_{12}}{n_{1.}} + N_{2.} \frac{n_{22}}{n_{2.}}} \tag{7.2}$$

where $N_{h.} = N_{h1} + N_{h2}$ and $n_{h.} = n_{h1} + n_{h2}$ are the strata population and sample sizes respectively, and y_{hi} are the sample totals in the (h, i)-th cell. The conditional expectation of (7.2) given $(n_{.1}, n_{.2})$ is not tractable since one has to evaluate the sum

$$E_2(\hat{Y}_{pst}) = \sum_t p(s_t | n_{.1}, n_{.2}) \hat{Y}_{pst}(t) \tag{7.3}$$

where s_t is a possible sample such that the observed sample sizes $\tilde{n}_{hi}$ satisfy $\tilde{n}_{1i} + \tilde{n}_{2i} = n_{.i}$ $(i = 1, 2)$, and $\hat{Y}_{pst}(t)$ is the value of (7.2) for the sample s_t, and $p(s_t | n_{.1}, n_{.2})$ is the conditional probability of observing s_t given $(n_{.1}, n_{.2})$:

$$p(s_t | n_{.1}, n_{.2}) = \left[\sum_{n_{11}=0}^{n_1} \binom{N_{11}}{n_{11}} \binom{N_{12}}{n_{1.} - n_{11}} \binom{N_{21}}{n_{.1} - n_{11}} \binom{N_{22}}{n_{2.} - n_{.1} + n_{11}} \right]^{-1}. \tag{7.4}$$

It is clear from (7.3) and (7.4), however, that $E_2(\hat{Y}_{pst}) \neq Y$ since $\hat{Y}_{pst}$ does not depend on the cell totals N_{hi} unlike $p(s_t | n_{.1}, n_{.2})$.

Turning to variance estimation, the usual formula for general designs is given by

$$v^*(\hat{Y}_{pst}) = v(z_t^*) \tag{7.5}$$

where $v(y_t) = v(\hat{Y})$ is the usual variance estimator of the estimated total $\hat{Y}$, and $v(z_t^*)$ is obtained from $v(\hat{Y})$ by replacing y_t by

$$z_t^* = y_t - \sum_i \frac{\hat{Y}_i}{M_i} a_t(i) \tag{7.6}$$

where $a_t(i) = 1$ if the t-th element belongs to the i-th post-stratum and $a_t(i) = 0$ otherwise (Williams 1962). In the SRS case, (7.5) reduces to

$$v^*(\hat{Y}_{pst}) \doteq N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \sum n_i s_{iy}^2 \tag{7.7}$$

(assuming $(n_i - 1)/(n-1) \doteq n_i/n$) which is not equal to (3.3) when multiplied by N^2. Hence, (7.5) does not behave well in the conditional framework, even in the SRS case. On the other hand, a new variance estimator

$$v(\hat{Y}_{pst}) = v(z_t), \tag{7.8}$$

where

$$z_t = \sum_i \frac{M_i}{\hat{M}_i} (y_t(i) - \frac{\hat{Y}_i}{\hat{M}_i} a_t(i)) \tag{7.9}$$

and $y_t(i) = y_t$ if the t-th element belongs to the i-th post-stratum and $y_t(i) = 0$ otherwise, might be preferable over $v^*(\hat{Y}_{pst})$ since in the SRS case it reduces to (3.3) when multiplied by N^2 and the finite population correction is ignored:

$$v(\hat{Y}_{pst}) = \sum_i \frac{N_i^2}{n_i} s_{iy}^2. \tag{7.10}$$

Some theory for ratio estimators under models also suggests that $v(\hat{Y}_{pst})$ might perform better conditionally than $v^*(\hat{Y}_{pst})$. In any case, there is no harm in switching to (7.8) since it is asymptotically equivalent to the customary variance estimator (7.5), unconditionally.

8. DISCUSSION

Our study clearly shows that conditional inference for complex designs involves formidable difficulties. Nevertheless, we should not use conventional procedures blindly. In those cases where conditional inference is feasible, as in the SRS case, we should certainly employ conditionally relevant methods as elaborated in Sections 2 - 6, while in the more complex cases we should at least make simple modifications to conventional methods, as in (7.8), so that they agree with known, conditionally correct results in special cases. Clearly, we need more research in this area.

ACKNOWLEDGEMENTS

This paper is based on my lectures given at a workshop on conditional inference in sample surveys. I am thankful to Mrs. Nanjamma Chinnappa for organizing the workshop at Statistics Canada. Constructive comments from colleagues at Statistics Canada and Professor D. Holt were helpful in preparing this paper.

REFERENCES

BANKIER, M. (1985). Conditionally unbiased estimators based on any number of independent stratified samples. Memorandum, Business Survey Methods Division, Statistics Canada.

BATTESE, G.E., and FULLER, W.A. (1981). Prediction of county crop areas using survey and satellite data. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 500-505.

BRYANT, E.C., HARTLEY, H.O., and JESSEN, R.J. (1960). Design and estimation in two-way stratification. *Journal of the American Statistical Association,* 55, 105-124.

CHINNAPPA, B.N. (1976). A preliminary note on methods of dealing with unusually large units in sampling from skew populations. Unpublished Technical Report, Institution and Agriculture Survey Methods Division, Statistics Canada.

COX, D.R., and HINKLEY, D.V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

DEMPSTER, A.P., RUBIN, D.B., and TSUTAKAWA, R.K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association,* 76, 341-353.

DOSS, D.C., HARTLEY, H.O., and SOMAYAJULU, G.R. (1979). An exact small sample theory for post-stratification. *Journal of Statistical Planning and Inference,* 3, 235-248.

DREW, J.H., SINGH, M.P., and CHOUDHRY, H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology,* 8, 17-47.

DURBIN, J. (1969). Inferential aspects of the randomness of sample size in survey sampling. In *New Developments in Survey Sampling* (Eds. N.L. Johnson and H. Smith), New York: Wiley - Interscience.

FISHER, R.A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd (5th Ed., 1934).

FULLER, W.A. (1966). Estimation employing post strata. *Journal of the American Statistical Association,* 61, 1172-1183.

HARTLEY, H.O., RAO, J.N.K., and KIEFER, G. (1969). Variance estimation with one unit per stratum. *Journal of the American Statistical Association,* 64, 841-851.

HIDIROGLOU, M.H., and SÄRNDAL, C.E. (1985). An empirical study of some regression estimators for small domains. *Survey Methodology,* 11, 65-77.

HIDIROGLOU, M.H., and SRINATH, K.P. (1981). Some estimators of the population total from simple random samples containing large units. *Journal of the American Statistical Association,* 76, 690-695.

HOLT, D., and SMITH, T.M.F. (1979). Post-stratification. *Journal of the Royal Statistical Society,* Ser. A, 142, 33-46.

LAHIRI, D.B. (1969). On the unique sample, the surveyed one. Unpublished Technical Report, Indian Statistical Institute.

OH, H.L., and SCHEUREN, F.J. (1983). Weighting adjustment for unit nonresponse. In *Incomplete Data in Sample Surveys,* Vol. 2, Academic Press, 142-184.

ROYALL, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika,* 57, 377-387.

SARNDAL, C.E. (1984). Design-consistent versus model-dependent estimators for small domains. *Journal of the American Statistical Association,* 79, 624-631.

WILLIAMS, W.H. (1962). The variance of an estimator with post-stratified weighting. *Journal of the American Statistical Association,* 57, 622-627.

YATES, F. (1984). Tests of significance for 2 $\times$ 2 contingency tables. *Journal of the Royal Statistical Society,* Ser. A, 147, 426-463.