MEAN SQUARED ERROR OF ESTIMATORS USING TWO STAGE SAMPLING FOR STRATIFICATION AND PRIOR INFORMATION

Donald B. White, State University of New York at Buffalo
Department of Statistics, Buffalo, NY 14261

The problem considered is that of estimating the total of a stratified finite population. Vardeman and Meeden (Ann. Stat. 12, 1984, 675-684) have introduced estimators for this situation which employ prior information regarding stratum sizes, averages, and memberships. In a two-stage sampling model for the situation of prior information on stratum sizes, we analyze their estimator for bias and for mean squared error relative to the usual estimators. A natural question then considered is the following: for given levels of confidence in the values specified by our prior information, what weighting constants for these values do we use in the estimators so as to minimize mean squared error? The answer to this question is found to depend upon some knowledge regarding the error in the prior information.

1. INTRODUCTION

Various stratified sampling designs employ various types of prior information. For example, the usual stratification model assumes full prior knowledge of individual stratum memberships. Poststratification is useful when there is global information on stratum sizes but no information on individuals. Two stage sampling for stratification, on the other hand, assumes no prior information on strata.

In each of these methods, the knowledge regarding stratum memberships or stratum sizes is assumed to be complete knowledge or complete lack of knowledge. However, the information required in these models is most commonly available through previous censuses or surveys or from parallel studies on other populations possessing similar global characteristics. It is apparent that usually the information is either dated or at least has applicability to the current study which is open to question. When forming an estimate, one therefore might wish to combine this prior information with analogous information gleaned from the current sample. Vardeman and Meeden (1984) have introduced a pair of estimators of the population total which combine information on stratum memberships, stratum sizes, and stratum averages with analogous information gained from the current sampling situation.

Their two estimators apply to two essentially different situations. The first, to be explored in Section 2, is where the prior information is global only, i.e., only on stratum sizes and averages. Their second estimate applies to the more complex situation where there is some partial information on individual stratum memberships. Due to space limitations, this discussion is limited to the first type of estimator, and only for partial information on stratum sizes, not stratum averages. In Section 2, bias and variance are derived with a view towards the analysis in Section 3 of standard error relative to that of the usual estimate. In particular, it is shown that if some knowledge of the accuracy of the prior information is available, then use of the methods given here is virtually certain to give a reduction in standard error. In fact, if one has an estimate of the distribution of this error, one can select weighting constants for the prior information which minimize the expected standard error (averaged over the distribution of the error in the prior information). Section 4 concludes with a summary and some possible extensions.

2. BIAS AND VARIANCE

We now discuss details of the model and estimator for the first situation described by Vardeman and Meeden. Using their notation, we have a finite population $\mathcal{P}$ of N units labelled $1,2,\ldots,N$ with associated values $y_1,\ldots,y_N$ which are unknown. Denote the population total by $\tau = \Sigma y_i$, where here, and in the following, summations over i run from 1 to N. Unit i possesses stratum membership $j_i \in \{1,2,\ldots,J\}$. Unknown constants of interest are the stratum averages $\bar{Y}_j$, stratum sizes N_j and relative stratum sizes $\rho_j = N_j/N$, for $1 \le j \le J$. Prior guesses for relative stratum sizes are Π_j. In this section, we regard Π_j as constant for each j since it does not depend upon the current samples. The constant $M \in [0,\infty]$ reflects the confidence in the prior guesses and will be described below.

The sample quantities of interest are the following. First, the stage one sample s^* is a simple random sample without replacement (WOR) of size n^* from $\mathcal{P}$. Here, we observe only stratum memberships j_i for $i \in s^*$ and we let n_j^* denote the number of units in the stage one sample which fall in stratum j. We note that $(n_1^*,n_2^*,\ldots,n_J^*)$ is multivariate hypergeometric with parameters N (population size), n^* (sample size) and $N_1,\ldots,N_J$ (stratum or group sizes).

The second stage sample is a stratified sample with stratum sample sizes $n_j = \nu_j(n_j^*)$, where for each j, ν_j is a function on the non-negative integers such that $\nu_j(0) = 0$, $\nu_j(1) = 1$ and $2 \le \nu_j(i) \le i$ if $i \ge 2$. The set of $n = \Sigma n_j$ (summations over j run from 1 to J) units in this second stage sample is denoted by s and y-values are determined for each of these units. Thus, we let $\bar{y}_j$ denote the average of the units in s and stratum **j**. Technically,

$$\bar{y}_j = \begin{cases} \dfrac{1}{n_j} \sum\limits_{i\in s\cap\{i:j_i=j\}} y_i & \text{if } n_j^* \ge 1 \\ 0 & \text{if } n_j^* = 0 . \end{cases}$$

We are now in a position to consider the meaning of the confidence coefficient M. M represents the confidence in the collection of guesses of relative stratum sizes, $\Pi_1,\ldots,\Pi_J$, and should be considered relative to the confidence in the sample estimates, n_j^*/n^*, $1 \le j < J$, which is given by the sample size n^*. Thus, if the prior guesses are given weight equal to that of the current estimates, we take $M = n^*$. Extreme cases of no confidence in (and thus no use of) the prior guesses and total confidence in the prior guesses (and thus no use of the current estimates) correspond to $M = 0$ and $M = \infty$, respectively.

We can now construct an estimate $\hat{\Pi}_j$ for ρ_j which is a weighted average of the prior guess and the sample estimate. This is

$$\hat{\Pi}_j = \frac{M\Pi_j + n_j^*}{M + n^*} . \tag{2.1}$$

Finally, an estimate $\hat{\tau}$ of the population total τ is constructed by replacing, in the formula for τ, any unobserved quantity by its estimate. We thus note that, letting $\mathcal{P}_j$ denote the units in stratum j, the total can be written

$$\tau = \sum N_j\bar{Y}_j = \sum_{j=1}^{J}\Bigg\{\sum_{i\varepsilon s\cap\mathcal{P}_j} Y_i + \sum_{i\varepsilon(s^*\cap\mathcal{P}_j)\setminus(s\cap\mathcal{P}_j)} Y_i + \sum_{i\varepsilon\mathcal{P}_j\setminus(s^*\cap\mathcal{P}_j)} Y_i\Bigg\} . \tag{2.2}$$

For $i \ \varepsilon\ s \cap \mathcal{P}_j$, Y_i are observed and $\sum_{i\varepsilon s\cap\mathcal{P}_j} Y_i = n_j\bar{y}_j$. For $i\ \varepsilon\ (s^* \cap\mathcal{P}_j)\setminus(s\cap\mathcal{P}_j)$, Y_i are not observed, but we know that there are $n_j^* - n_j$ such units. Values Y_i for $i\ \varepsilon\ \mathcal{P}_j\setminus(s^*\cap\mathcal{P}_j)$ are likewise not observed; here, there are $N_j - n_j^*$ such units but this must be estimated by $(N - n^*)\hat{\Pi}_j$. Estimating all unobserved quantities, we obtain our estimate of the total,

$$\hat{\tau} = \sum[n_j^* + (N - n^*)\hat{\Pi}_j]\bar{y}_j. \tag{2.3}$$

We now seek the bias, variance and mean squared error of $\hat{\tau}$. We begin with the expectation, and thus bias, of $\hat{\tau}$. All moments are computed using the following two-step conditioning argument. Since the second stage sample depends upon the first stage sample, we condition first on s^*. The expectation or variance of the resulting function of s^* is then required. Since the first stage sample is a simple random sample without replacement, the units within a given stratum all have the same probability of selection. It is the number of units selected from each stratum (n_j^*) which is random. Thus, conditionally, given n_j^*, the first stage sample from $\mathcal{P}_j$ is a simple random sample without replacement of size n_j^*, and the second conditioning step is to condition on n_j^*.

Thus, proceeding piecewise, we have $E(n_j^*\bar{y}_j) = E(E(n_j^*\bar{y}_j|s^*)) = E(n_j^*\bar{y}_j^*)$ where $\bar{y}_j^*$ is the unobserved average of the n_j^* y-values of the units in $s^* \cap \mathcal{P}_j$. Any average of a vacuous set of values will be taken as zero. Finally, we have $E(n_j^*\bar{y}_j^*) = E(n_j^*E(\bar{y}_j^*|n_j^*)) = E(n_j^*\bar{Y}_j) = n^*\rho_j\bar{Y}_j$.

In a similar fashion, we have $E(\hat{\Pi}_j\bar{y}_j) = E(\hat{\Pi}_j E(\bar{y}_j|s^*)) = E(\hat{\Pi}_j\bar{y}_j^*) = E(\hat{\Pi}_j E(\bar{y}_j^*|n_j^*)) = E(\hat{\Pi}_j\bar{Y}_j I_{[n_j^*\neq 0]}) = \bar{Y}_j E(\hat{\Pi}_j I_{[n_j^*\neq 0]})$.

Combining these results, we discover $E(\hat{\tau}) = \Sigma[n^*\rho_j + (N - n^*)E(\hat{\Pi}_j I_{[n_j^*\neq 0]})]\bar{Y}_j$. But, since $N_j = N\rho_j$, we note that $\tau = \Sigma[n^*\rho_j + (N - n^*)\rho_j]\bar{Y}_j$ and the bias, $B(\hat{\tau}) = E(\hat{\tau}) - \tau$, is therefore given by $(N - n^*)\Sigma\bar{Y}_j E(\hat{\Pi}_j I_{[n_j^*\neq 0]} - \rho_j)$. But $E(\hat{\Pi}_j I_{[n_j^*\neq 0]} - \rho_j) = E[(\hat{\Pi}_j - \rho_j - M\Pi_j I_{[n_j^*\neq 0]})/(M + n^*)] = M(\Pi_j P[n_j^* \neq 0] - \rho_j)/(M + n^*)$, and

$$B(\hat{\tau}) = [M(N - n^*)\Sigma\bar{Y}_j(\Pi_j' - \rho_j)]\ /(M + n^*) \tag{2.4}$$

where $\Pi_j' = \Pi_j P[n_j^*\neq 0] \approx \Pi_j$.

We now turn to the derivation of the variance of $\hat{\tau}$. As before, we condition on s^*. Letting $g_j(n_j^*) = n_j^* + (N - n^*)\hat{\Pi}_j$, we first obtain the random variables

$$E(\hat{\tau}|s^*) = \sum g_j(n_j^*)\bar{y}_j^* . \tag{2.5}$$

and

$$\mathrm{var}(\hat{\tau}|s^*) = \sum g_j(n_j^*)^2\mathrm{var}(\bar{y}_j|s^*). \tag{2.6}$$

The latter formula follows from the fact that if $j \neq j'$, $\bar{y}_j$ and $\bar{y}_{j'}$ are conditionally independent given s^*. Also, conditioned on the first stage sample, $\bar{y}_j$ is a sample average based on a simple random sample without replacement from units in $s^* \cap \mathcal{P}_j$. If we denote the finite population variance of these n_j^* units by s_{yj}^{2*}, then finally we obtain

$$\mathrm{var}(\hat{\tau}|s^*) = \sum g_j(n_j^*)^2(1/n_j - 1/n_j^*)s_{yj}^{2*}. \tag{2.7}$$

When $n_j^* \le 1$, we define the coefficient of s_{yj}^{2*} to be zero.

The second step is to condition on $\underset{\sim}{n}^* = (n_1^*,$

$\ldots n_J^*$). We proceed by deriving $\mathrm{var}(E(\hat{\tau}|s^*)) = \Sigma \mathrm{var}(g_j(n_j^*)\bar{y}_j^*) + \sum_{j\neq j'} \mathrm{cov}(g_j(n_j^*)\bar{y}_j^*, g_{j'}(n_{j'}^*)\bar{y}_{j'}^*)$. These terms shall be evaluated separately. First, we have $\mathrm{var}(g_j(n_j^*)\bar{y}_j^*) = \mathrm{var}(g_j(n_j^*)E(\bar{y}_j^*|\underset{\sim}{n}^*)) + E(g_j(n_j^*)^2 \mathrm{var}(\bar{y}_j^*|\underset{\sim}{n}^*)) = \bar{Y}_j^2 \mathrm{var}(g_j(n_j^*)I_{[n_j^*\neq 0]}) + S_{yj}^2 E(g_j(n_j^*)^2(1/n_j^* - 1/N_j)I_{[n_j^*\neq 0]})$ where S_{yj}^2 is the finite population variance for the N_j units in stratum j. We derive the covariances in an exactly analogous fashion. Here, for $j \neq j'$, $\mathrm{cov}(g_j(n_j^*)\bar{y}_j^*, g_{j'}(n_{j'}^*)\bar{y}_{j'}^*) = \bar{Y}_j\bar{Y}_{j'}\mathrm{cov}(g_j(n_j^*) I_{[n_j^*\neq 0]}, g_{j'}(n_{j'}^*)I_{[n_{j'}^*\neq 0]})$ since the two variables are conditionally independent given $\underset{\sim}{n}^*$.

We finally note that since, conditioned on n_j^*, the first stage sample from stratum j has the same distribution as a simple random without replacement, $E(\mathrm{var}(\hat{\tau}|s^*)) = \Sigma E(g_j(n_j^*)^2(1/n_j - 1/n_j^*)s_{yj}^{*2}) = \Sigma E(g_j(n_j^*)^2(1/n_j - 1/n_j^*)E(s_{yj}^{*2}|n_j^*)) = \Sigma S_{yj}^2 E(g_j(n_j^*)^2(1/n_j - 1/n_j^*))$. Combining these results, we obtain

$$\mathrm{var}(\hat{\tau}) = \sum\{S_{Yj}^2 E(g_j'(n_j^*)^2(1/n_j - 1/N_j)) + \bar{Y}_j^2 \mathrm{var}(g_j'(n_j^*))\} + \sum_{j\neq j'} \bar{Y}_j\bar{Y}_{j'}\mathrm{cov}(g_j'(n_j^*), g_{j'}'(n_{j'}^*)). \quad (2.8)$$

where $g_j'(n_j^*) = g_j(n_j^*)I_{[n_j^*\neq 0]}$.

In the usual two stage sampling for stratification situation, where there is no prior information and we take M = 0, this formula can be reduced to

$$\mathrm{var}(\hat{\tau}) = \frac{N^2}{n^*}(1 - \frac{n^*}{N})S_y^2 + \frac{N^2}{n^{*2}} \sum S_{yj}^2 E[n_j^*(\frac{n_j^*}{n_j} - 1)],$$

where S_y^2 is the finite population variance for all of $\mathcal{P}$. This formula is the exact version of Cochran's formula (12.8) (Cochran (1977), p. 329) and can be found in Tucker (1981), p. 144.

3. CHOICE OF WEIGHTING CONSTANTS

In this section, we shall examine the mean squared error of $\hat{\tau}$ and determine optimal values of the weighting constant M according to the following two rules. First, we wish to minimize mean squared error and second, we wish to insure a small likelihood of having mean squared error larger than would be obtained by ignoring the prior information. In the following, we shall explore the general principles of how to obtain a gain in precision and how much gain to expect. To carry out such an exploration, we take J = 2. The only other assumptions we shall make are that the stratum sizes are much larger than one, that $n_j = cn_j^*$, j=1,2, c is large enough so that $P[n_j^* = 0] \approx 0$, and that $S_{Y_1}^2 = S_{Y_2}^2 = S^2$. We shall also use standardized and scaled y-values so that we can take $\bar{Y}_1 = 1$ and $\bar{Y}_2 = 0$. From the above, 2.4 and 2.8 we have

$$\mathrm{MSE}(\hat{\tau}) = \left[\frac{n^*M(\frac{1}{f^*} - 1)}{M + n^*}\right]^2 (\Pi_1 - \rho_1)^2 + \left[\frac{M + \frac{n^*}{f^*}}{M + n^*}\right]^2 \times \left\{\frac{S^2}{c}\sum_{j=1}^{2} E\left[\frac{(n_j^* + a)^2}{n_j^*}\right] + n^*\rho_1(1 - \rho_1)(1 - f^*)\right\} \quad (3.1)$$

where $f^* = n^*/N$, the first stage sampling fraction and $a = M\Pi_j(N - n^*)/(M + N)$. The remaining moment can be expanded to $E(n_j^*) + 2a + a^2E(n_j^{*-1})$. Letting $g(x) = 1/x$, using the linear approximation to g(x) expanded about $E(n_j^*) = n^*\rho_j$ yields $E(n_j^{*-1}) = E(g(n_j^*)) \approx (n^*\rho_j)^{-1}$. The error to this approximation is bounded above by $(n^*\rho_j)^{-2} \mathrm{var}(n_j^*)^{1/2} \approx (n^*\rho_j)^{-3/2}(1 - \rho_j)^{1/2}(1 - f^*)^{1/2}$, and the relative error, which is bounded by $((1 - \rho_j)(1 - f^*)/n^*\rho_j)^{1/2}$, is small if $n^*\rho_j$ is large enough. With this approximation, $E((n_j^* + a)^2/n_j^*)$ takes the particularly simple form $n^*\rho_j(1 + a/(n^*\rho_j))^2$. Using this approximation, and substituting $M' = M/n^*$ for M, we have

$$\mathrm{MSE}(\hat{\tau}) = \left[\frac{n^*M'(\frac{1}{f^*} - 1)}{M' + 1}\right]^2(\Pi_1 - \rho_1^*)^2 + n^*\left[\frac{M' + \frac{1}{f^*}}{M' + 1}\right]^2 \times \left[\frac{S^2}{c}\sum_{j=1}^{2}\rho_j\left(1 + \frac{a}{n^*\rho_j}\right)^2 + \rho_1(1 - \rho_1)(1 - f^*)\right]. \quad (3.2)$$

If we regard the above as a function of M', say h(M'), then h(0) is the mean squared error (actually variance) of $\hat{\tau}$ in the usual case where there is no consideration of prior information.

Since we wish to perform an analysis in terms of standard error, $SE(\hat{\tau}) = MSE(\hat{\tau})^{1/2}$, then our current goal is analyze $r(M') = (h(M')/h(0))^{1/2}$ for minimum and $\{M' : r(M') \le 1\}$. The function r is given by

$$r(x) = \frac{(Ax^2 + Bx + 1)^{1/2}}{x + 1} \tag{3.3}$$

where

$$A = [n^*(1 - f^*)^2(\Pi_1 - \rho_1)^2 + \frac{S^2}{c}\sum_{j=1}^{2} \rho_j \left(\frac{\Pi_j}{\rho_j}(1 - f^*) + f^*\right)^2 + f^{*2}\rho_1(1 - \rho_1)(1 - f^*)] \div \left(\frac{S^2}{c} + \rho_1(1 - \rho_1)(1 - f^*)\right)$$

and

$$B = \frac{2[\frac{S^2}{c} + f^*\rho_1(1 - \rho_1)(1 - f^*)]}{\frac{S^2}{c} + \rho_1(1 - \rho_1)(1 - f^*)}.$$

Key features of r(x) are the following:

1) $r(0) = 1$ and $r'(0) < 0$, implying that for at least some small values of M', we will obtain a decrease in $SE(\hat{\tau})$ from that of the usual estimate,

2) r is minimized at $x_0 = (2-B)/(2A-B)$; if $2A \le B$, then r is strictly decreasing,

3) $r(x) = 1$ at $x_1 = (2-B)/(A-1)$ as well as at $x = 0$, and

4) $r(\infty) = \lim_{x\to\infty} r(x) = A^{1/2}$.

Except in a few extreme cases, r(x) is not very sensitive to f^* or the actual values of ρ_1 and ρ_2 (only relative to the guesses Π_1 and Π_2). For our future analysis, we therefore fix an example: $f^* = .10$, $\rho_1 = .25$, and $\rho_2 = .75$. Since Π_j/ρ_j should be near unity for j=1,2, the sum in the middle term of the numerator of A should also be near unity. Thus, the key variables which dominate r(x) are $n^{*1/2}|\Pi_1 - \rho_1|$ and $c^{-1/2}S$ which shall be denoted e and v, respectively. Figures

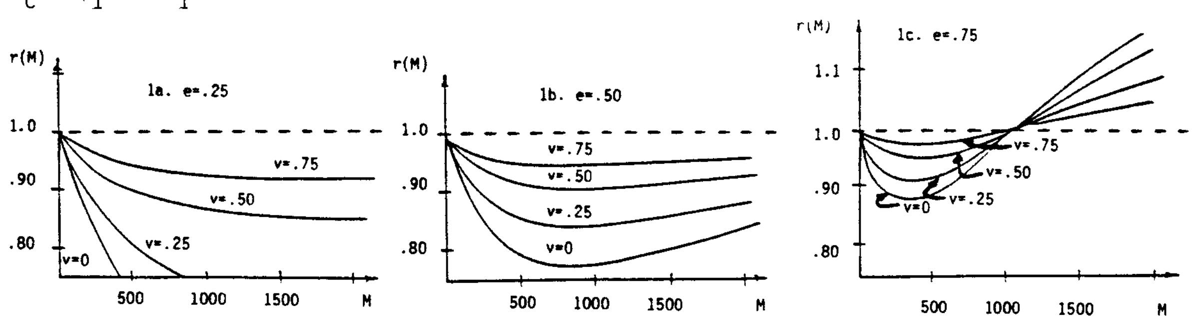

Figure 1. Standard error ratios for various e and v.

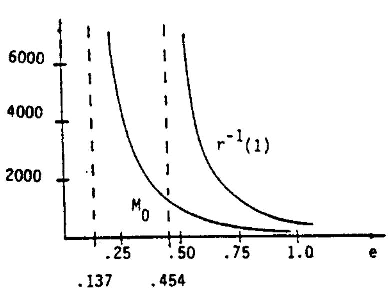

Figure 2.

$M = M_0$ yields the minimum value of r.

$M > r^{-1}(1)$ yields

$$[S.E.(\hat{\tau})](M) > [S.E.(\hat{\tau})](0).$$

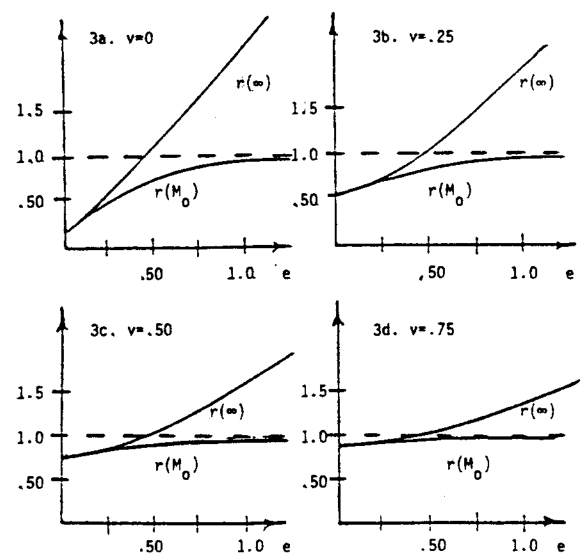

Figure 3. Minimum and limiting standard error ratios.

1a, b, and c show $r(x)$ for various values of these parameters. The following principles emerge from these graphs: 1) as e increases, the optimal value of M decreases, as does the value of M below which the ratio is less than one, and 2) as v increases, the curve $r(x)$ flattens around one, thus, decreasing both the potential gain and the risk of drastically increased standard error obtained by using the prior information. We then see that the ratio $r(x)$ is most sensitive to the two known constants n^* and c and to the two unknown quantities $|\Pi_1 - \rho_1|$ and S, but also that $r(x)$ only depends on these four through the two unknown quantities e and v. We thus set $n^* = 1000$ and c plays no role apart from v. Figure 2 shows the relationships between x_0 and e and x_1 and e; these quantities depend only inperceptably on v. The two asymptotes, .137 for M_0 and .454 for $r^{-1}(1)$, carry the following interpretations. Approximating Π_j/ρ_j by one, if $e < \sqrt{f^*}\rho_1\rho_2 = .137$ then M_0 is infinity and the usual poststratification estimate is best. Further, if $e < \sqrt{\rho_1\rho_2(1 + f^*)} = .454$, then $r^{-1}(1) = \infty$ and the standard error using the methods given here is always lower than when the prior information is ignored. Figures 3a, b, c, and d show for various v how $r(M_0)$ and $r(\infty)$ depend on e.

The reader should observe at this point that, although the sampler may have an idea of the value of S and thus v, by the definition of the method being presented, he has little knowledge of the value of $|\Pi_1 - \rho_1|$ except that he hopes it is near zero. This problem can, however, be studied from the following point of view. If the sampler has a notion of Π_1 as an estimate based on a previous sample of size, say, m, then one might ask what value of M is appropriate, relative to m and to the known behavior of $r(x)$ as a function of $|\Pi_1 - \rho_1|$? Since $r(x)$ actually depends upon $(\Pi_1 - \rho_1)^2$, we note the following: If the true population fraction in stratum 1 at the time of the previous sample was also ρ_1, if m/N is not too large and if Π_1 is the usual estimate of ρ_1, then $m\Pi_1$ is approximately binomial, $E(\Pi_1 - \rho_1)^2 = \rho_1(1 - \rho_1)/m$ and $sd((\Pi_1 - \rho_1)^2) \approx \sqrt{2}E(\Pi_1 - \rho_1)^2$ (see Johnson and Kotz (1969) p. 51). A reasonable value to use for e^2 then is its expectation under this model, $\hat{e}^2 = (n^*/m)\rho_1(1 - \rho_1)$. Approximating Π_j/ρ_j by one and f^* by zero we thus select $x = \hat{x}_0 = \rho_1(1 - \rho_1)/\hat{e}^2 = m/n^*$; thus the ratio is minimized by taking $M = m$. This result reflects our original intuition in regarding M as a value reflecting the confidence in Π_1 relative to the sample estimate, n_1^*/n^*. If this guess at e^2 is correct, then $SE(\hat{\tau})$ is as small as we can make it. If this guess is an overestimate of e^2, then M is smaller than the optimal value and $x < x_0$. This is relatively painless; at least here, we are guaranteed that $r(x) < 1$. The danger comes in underestimating e^2 by so much that $x > x_1$ and the estimate $\hat{\tau}$ has higher standard error than does the usual estimate. Say that $e^2 = \alpha\,\hat{e}^2$, $\alpha > 1$. Then we select

$$x = \hat{x}_0 = \frac{\rho_1(1 - \rho_1)}{\hat{e}^2} = \alpha\,\frac{\rho_1(1 - \rho_1)}{e^2} \approx \alpha\, x_0.$$

Noting that $x_1 \gtrapprox 2x_0$ since $B < 2$, we are sure that $x < x_1$ if $\alpha < 2$, i.e., if $(\Pi_1 - \rho_1)^2 < 2E(\Pi_1 - \rho_1)^2 = E(\Pi_1 - \rho_1)^2 + (1/\sqrt{2})\, sd(\Pi_1 - \rho_1)^2$. We are thus safe if the true squared error of Π_1 is less than .71 standard deviations higher than its expected value. This fact, when combined with possible problems in the applicability of the previous sample, may lead the sampler to select M somewhat smaller than m in order to insure against a higher standard error than when the prior information is ignored.

4. CONCLUSIONS AND EXTENSIONS

The results in Section 3 strongly indicate that there are situations where the use of this estimate is a viable option. In any poststratified model where there is some information on stratum sizes, one should put some effort into either establishing bounds on the error in this prior information or in approximating its distribution. With such knowledge, even if it is only a gross approximation, one can use the methods and results given here to make an educated choice between 1) total trust in the prior information (standard poststratification) if the error is small enough, 2) no use of the prior information (standard two stage sampling for stratification) if the error is likely to be large, and 3) use of the weighted average estimates discussed here if the error in the prior guess for the distribution of units among strata is thought to be moderate. If the latter route is chosen, one can also use these methods to determine a weighting constant M which is close to optimal.

There are many areas for further study. Some extensions of these results to the situations when there is prior information available on individual stratum memberships and stratum averages as well as relative stratum sizes is in preparation for publication. Also, the model, as it stands, defies variance estimation. Adaptation of the model to allow for variance estimation is an important extension. Further work on how to determine optimal weighting constants is also needed. A primary need is to apply these methods to current sampling situations. Finally, on the theoretical side, further work in establishing the Bayesian foundations to these methods is required.

REFERENCES

Cochran, W.G. (1977), Sampling techniques, New York, John Wiley and Sons, Inc.

Johnson, N.L. and Kotz, S. (1969), Discrete distributions, Boston: Houghton Mifflin Company.

Tucker, H.G. (1981), An introduction to the mathematical theory of sample surveys, unpublished manuscript.

Vardeman, S. and Meeden, G. (1984), "Admissable estimators for the total of a stratified population that employ prior information," Annals of Statistics, 12, 675-684.