Ann. Inst. Statist. Math.
Vol. 45, No. 1, 113–127 (1993)

A GENERAL RATIO ESTIMATOR AND ITS APPLICATION IN MODEL BASED INFERENCE

Z. OUYANG[1], J. N. SRIVASTAVA[2]* AND H. T. SCHREUDER[1]

[1] *240 W. Prospect St., Rocky Mountain Forest Experiment Station, Fort Collins, CO 80521, U.S.A.*
[2] *Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A.*

(Received October 29, 1990; revised April 16, 1992)

Abstract. A general ratio estimator of a population total is proposed as an approximation to the estimator introduced by Srivastava (1985, *Bull. Internat. Statist. Inst.*, **51**(10.3), 1–16). This estimator incorporates additional information gathered during the survey in a new way. Statistical properties of the general ratio estimator are given and its relationship to the estimator proposed by Srivastava is explored. A special kind of general ratio estimator is suggested and it turns out to be very efficient in a simulation study when compared to several other commonly used estimators.

Key words and phrases: Finite population sampling, sample weight function, linear model, regression estimator, the Horvitz-Thompson estimator.

1. Introduction

Assume a finite population U with index set $\{1, \ldots, N\}$. An unknown, but observable, variable y is defined on U with value y_i at unit i. Let Y be the population total of y and $\boldsymbol{y} = (y_1, \ldots, y_N)$. A sample ω of U is defined as a non-empty subset of U, which has probability $p(\omega)$ of being drawn.

Consider an estimator $\hat{Y}_{\text{sr1}}$ of the population total proposed by Srivastava (1985) which is given by

$$(1.1) \qquad \hat{Y}_{\text{sr1}} = r(\omega) \sum_{i\omega} y_i/\pi_r(i),$$

where

$$(1.2) \qquad \pi_r(i) = \sum_{\omega i} p(\omega) r(\omega),$$

* The work of this author was supported by AFOSR grant #830080.

and $\sum_{i\omega}$ denotes the sum over all i in sample ω, $\sum_{\omega i}$ denotes the sum over all ω containing the unit i. The $r(\cdot)$ is called the "sample weight function" defined on all nonempty ω, which is independent of the actually drawn sample, and may be chosen at will by the statistician. During a survey, the values y_i (for i belonging to a sample ω) are collected. Usually, extraneous information about the nature of the population is also available. For instance, it is well known in forestry that volumes of trees in a forest fit an exponential distribution quite well. With the help of previous information, an experienced cruiser in a forest survey might be able to give a quite good guess to the parameters of the exponential distribution. This kind of information should be utilized in the estimation stage. The function $r(\cdot)$ was introduced for the purpose of utilizing such information. If π_i, the inclusion probability of unit i, is positive for all i, $\hat{Y}_{\mathrm{sr1}}$ is an unbiased estimator of Y (Srivastava (1985)). Suppose the extraneous information is formally introduced as a vector $\boldsymbol{y}^*$ which represents a guess of $\boldsymbol{y}$, such that $\boldsymbol{y}^*$ is independent of the drawn sample. When the parameter space is $R_+^N = \{(y_1, \ldots, y_N)' : y_i \geq 0, i = 1, \ldots, N\}$, a necessary and sufficient condition for $\hat{Y}_{\mathrm{sr1}}$ to have zero variance at $\boldsymbol{y}^* \in R_+^N$ is that

$$\sum_{i\omega} 1/r_i = 1/r(\omega), \qquad \text{for all } \omega \text{ such that } p(\omega) > 0, \tag{1.3}$$

and

$$y_i^*/Y^* = \pi_r(i)/r_i, \qquad i = 1, \ldots, N, \tag{1.4}$$

where $r_i = r(\{i\}) > 0$ and Y^* is the total of y_i^*. For a sample weight function which satisfies equation (1.3), the corresponding $\hat{Y}_{\mathrm{sr1}}$ is admissible among all homogeneous linear unbiased estimators of Y and its variance is given by the following:

$$\mathrm{Var}(\hat{Y}_{\mathrm{sr1}}) = \sum_{i<j} \frac{\pi_r(i)\pi_r(j) - \pi_{r^2}(i,j)}{r_i r_j} \left(\frac{r_i y_i}{\pi_r(i)} - \frac{r_j y_j}{\pi_r(j)} \right)^2, \tag{1.5}$$

where

$$\pi_{r^2}(i,j) = \sum_{\omega ij} p(\omega)[r(\omega)], \tag{1.6}$$

and $\sum_{\omega ij}$ denotes the sum over all ω containing the units i and j. Details of the result listed above were given in Srivastava and Ouyang (1992).

The most important concept underlying estimator (1.1) is the sample weight function. This function relies on the availability of extraneous information about the nature of the population other than the information given by a sample. Such information does exist, for example, in social surveys. Srivastava and Ouyang (1992) give an example to demonstrate how to collect this information in social surveys. Indeed, questionnaires collected by a sample s provide much more information than the information given by $\{y_i, i \in s\}$. Without doubt, we should use as much information as we can in the estimation stage. A technique given in

Srivastava and Ouyang (1992) uses such information and provides a better result in the example used. Encouraged by this result, Ouyang and Schreuder (1992) tried to use this technique in forest survey. An important step in this technique is to "match" the y's in a sample to the y^*'s in the guessed population. Thus, the sample weight function obtained was based on the "match" from the sample. This can improve estimation as shown but makes estimation of precision more difficult. We need to either refine this technique or develop other theories to incorporate sample related "extraneous" information. In this paper we only demonstrate that if we have a model of the population, this model can provide improved estimation for the population. We do this by incorporating this additional information into estimator $\hat{Y}_{\text{sr1}}$ efficiently and simply as shown below.

2. An approximation to $\hat{Y}_{\text{sr1}}$

Suppose a guess of $\boldsymbol{y}$, $\boldsymbol{y}^*$, is available such that $\boldsymbol{y}^* \in R_+^N$ after a sample is drawn by using a sampling design, which may or may not depend on a covariate of the population. An iterative procedure to solve equations (1.3) and (1.4) is given in Srivastava and Ouyang (1992). For a set of given $\pi_r(i)$, say π_i' (we might let $\pi_i' = \pi_i$), the iterative procedure consists of using

$$\begin{aligned} &\text{(a)} \quad \pi_r^0(i) = \pi_i' \quad (i = 1, \ldots, N), \\ &\text{(b)} \quad r_i^m = \pi_r^m(i)/(y_i^*/Y^*) \quad (i = 1, \ldots, N), \\ &\text{(c)} \quad r^m(\omega) = \left[\sum_{i\omega} (r_i^m)^{-1}\right]^{-1}, \\ &\text{(d)} \quad \pi_r^{m+1}(i) = \sum_{\omega i} r_i^m(\omega) p(\omega) \quad (i = 1, \ldots, N) \end{aligned} \tag{2.1}$$

where iteration starts with $m = 1$. A numerical example given in Ouyang and Schreuder (1992) shows that the iterative procedure works well (for one, five, and hundred iterations). But it is not easy to express the estimator obtained analytically. This fact leads to the consideration of the following eatimator obtained by using a "half" step iteration procedure.

Suppose $s(\cdot)$ is a sample weight function with $s(\omega) > 0$ for all ω such that $s(\cdot)$ is an approximate solution to equations (1.3) and (1.4). Let $\pi_i' = \pi_s(i)$ in (2.1). Then (2.1) (b) and (c) provide a sample weight function

$$t(\omega) = \left\{\sum_{i\omega} [y_i^*/\pi_s(i)]\right\}^{-1} Y^*, \quad \text{for all} \quad \omega. \tag{2.2}$$

Obviously $t(\omega) > 0$ for all ω. This "half" step iteration gives an estimator

$$\hat{Y} = t(\omega) \sum_{i\omega} y_i/\pi_s(i) = Y^* \left\{\sum_{i\omega} y_i/\pi_s(i)\right\} \Big/ \left\{\sum_{i\omega} y_i^*/\pi_s(i)\right\}, \tag{2.3}$$

which will be called the "general ratio estimator" in this study.

3. Statistical properties of $\hat{Y}$

To study $\hat{Y}$ it is natural to consider the "model" based on the $\boldsymbol{y}^*$:

$$y_i = (y_i^*/Y^*)Y + \epsilon_i = \theta_i Y + \epsilon_i, \qquad i = 1, \ldots, N, \tag{3.1}$$

where

$$\sum_i \epsilon_i = 0, \quad \sum_i \theta_i = 1, \quad \text{and} \quad y_i^* = \theta_i Y^*, \qquad \text{for all } i. \tag{3.2}$$

THEOREM 3.1. *Let $\pi_t(i)$ be defined by (1.2) with $r(\cdot)$ replaced by $t(\cdot)$. Then*

$$|E(\hat{Y}) - Y| \le \max_i \left| \frac{\pi_t(i) - \pi_s(i)}{\pi_s(i)} \right| \left(\sum_i |\epsilon_i| \right). \tag{3.3}$$

PROOF. Straightforward.

Usually, $\hat{Y}$ is a biased estimator of Y. But when $\pi_s(i)$ is an exact solution of equations (1.3) and (1.4) or $\epsilon_i = 0$ for all i, $\hat{Y}$ is unbiased. The following gives a useful formula for the bias.

THEOREM 3.2. *Let $\gamma = \gamma(\omega) = s(\omega) \sum_{i\omega} \theta_i/\pi_s(i)$, then*

$$E(\hat{Y}) = Y - \mathrm{Cov}(\hat{Y}, \gamma). \tag{3.4}$$

PROOF. Similar to the proof of Theorem 5.1 in Raj (1972).

THEOREM 3.3. *Let*

$$\pi_{t^2}(i,j) = \sum_{\omega ij} p(\omega)[t(\omega)]^2. \tag{3.5}$$

Then

$$\mathrm{MSE}(\hat{Y}) = \sum_{i<j} \frac{\pi_s(i)\pi_s(j) - \pi_t(i)\pi_s(j) - \pi_t(j)\pi_s(i) + \pi_{t^2}(i,j)}{\pi_s(i)\pi_s(j)} \cdot y_i^* y_j^* \left\{ \frac{y_i}{y_i^*} - \frac{y_j}{y_j^*} \right\}^2. \tag{3.6}$$

PROOF. Let $a_{i\omega} = 1$ if $i \in \omega$ and $a_{i\omega} = 0$ if $i \notin \omega$. So $\hat{Y} = t(\omega) \sum_i a_{i\omega} y_i/\pi_s(i)$. Straightforward calculation gives

$$\begin{aligned}
&E\{[t(\omega)]^2 a_{i\omega} a_{j\omega}/[\pi_s(i)\pi_s(j)]\} = \pi_{t^2}(i,j)/[\pi_s(i)\pi_s(j)],\\
&E\{t(\omega) a_{i\omega}/\pi_s(i)\} = \pi_t(i)/\pi_s(i),\\
&E\{[t(\omega) a_{i\omega}/\pi_s(i) - 1][t(\omega) a_{j\omega}/\pi_s(j) - 1]\}\\
&\qquad = [\pi_s(i)\pi_s(j) - \pi_t(i)\pi_s(j) - \pi_t(j)\pi_s(i) + \pi_{t^2}(i,j)]/(\pi_s(i)\pi_s(j)).
\end{aligned}$$

When $y_i = \theta_i Y^*$ for all i, since $y_i^* = \theta_i Y^*$, then

$$\hat{Y} = Y^* \left[\sum_{i\omega} \theta_i/\pi_s(i)\right]^{-1} \left[\sum_{i\omega} \theta_i/\pi_s(i)\right] = Y^*, \tag{3.7}$$

hence $\mathrm{MSE}(\hat{Y}) = 0$ at $\boldsymbol{y}^*$. Then apply Rao's theorem on the mean square error of estimators that possess the ratio estimator property (Rao (1979)) to obtain (3.6).

When $s(\omega) = t(\omega)$, (3.6) becomes the variance of $\hat{Y}_{\mathrm{sr1}}$. An unbiased estimator of $\mathrm{MSE}(\hat{Y})$ based on (3.6) can be easily proposed if π_{ij} is known. Now since

$$\begin{aligned} \hat{Y} &= t(\omega) \sum_{i\omega} (\theta_i Y)/\pi_s(i) + t(\omega) \sum_{i\omega} \epsilon_i/\pi_s(i) \\ &= Y + t(\omega) \sum_{i\omega} \epsilon_i/\pi_s(i), \end{aligned} \tag{3.8}$$

another expression for $\mathrm{MSE}(\hat{Y})$ can be obtained by the following equation

$$\mathrm{MSE}(\hat{Y}) = E\left(t(\omega) \sum_{i\omega} \epsilon_i/\pi_s(i)\right)^2. \tag{3.9}$$

THEOREM 3.4. *The* $\mathrm{MSE}(\hat{Y})$ *is given by*

$$\mathrm{MSE}(\hat{Y}) = \sum_i \frac{\pi_{t^2}(i) - [\pi_s(i)]^2}{[\pi_s(i)]^2} \epsilon_i^2 + \sum_{i \neq j} \frac{\pi_{t^2}(i,j) - \pi_s(i)\pi_s(j)}{\pi_s(i)\pi_s(j)} \epsilon_i \epsilon_j. \tag{3.10}$$

PROOF. Use the argument to obtain $\mathrm{Var}(\hat{Y}_{\mathrm{sr1}})$ given in Srivastava (1985) in (3.9) to obtain (3.10).

From (3.10), $\mathrm{MSE}(\hat{Y})$ looks very similar to $\mathrm{Var}(e)$, where

$$e = e(\omega) = s(\omega) \sum_{i\omega} \epsilon_i/\pi_s(i) \tag{3.11}$$

in an "estimate" of $0 = \epsilon_1 + \cdots + \epsilon_N$. The following theorem shows that in fact $\mathrm{Var}(e)$ is an approximation to $\mathrm{MSE}(\hat{Y})$.

THEOREM 3.5. *Suppose* $E\{|\gamma(\omega) - 1|\} < 1$. *Then*

$$\mathrm{MSE}(\hat{Y}) \cong \mathrm{Var}(e). \tag{3.12}$$

PROOF. Given in the Appendix.

Since $\gamma(\omega)$ is an unbiased estimator of the total of θ_i, i.e., 1, $|\gamma(\omega) - 1|$ should be small, in fact, usually much smaller than 1 (the total of θ_i). Hence it is not too hard to satisfy the condition given in Theorem 3.5. Finally, we compare $\hat{Y}$ and $\hat{Y}_{\mathrm{sr1}}$.

THEOREM 3.6. *As estimators of* Y, *the statistics* $\hat{Y}$ *and* $\hat{Y}_{\mathrm{sr1}}$ *are related as follows*:

(i) *If the* $s(\cdot)$ *is the exact solution of equations* (1.3) *and* (1.4), *then* $\mathrm{Var}(\hat{Y}_{\mathrm{sr1}}) = \mathrm{MSE}(\hat{Y})$.

(ii) *If the* $s(\cdot)$ *is not the exact solutions of equations* (1.3) *and* (1.4) *but* $y_i = \theta_i Y$ *for all* i, *then* $\mathrm{Var}(\hat{Y}_{\mathrm{sr1}}) > \mathrm{MSE}(\hat{Y}) = 0$.

(iii) *Suppose the condition in Theorem* 3.5 *holds and*

$$2|\epsilon_i| < \delta\theta_i Y, \qquad i = 1, \ldots, N, \tag{3.13}$$

for $\delta = \mathrm{Var}(\gamma(\omega))/[1 + \mathrm{Var}(\gamma(\omega))]$. *Then* $\mathrm{Var}(\hat{Y}_{\mathrm{sr1}})$ *is larger than* $\mathrm{MSE}(\hat{Y})$ *to the first order of approximation.*

$$\mathrm{Var}(\hat{Y}_{\mathrm{sr1}}) > \mathrm{MSE}(\hat{Y}).$$

PROOF. Given in the Appendix.

Thus, based on Theorem 3.6, if we have only an approximate solution to equation (2.4), $\hat{Y}$ should be preferred over $\hat{Y}_{\mathrm{sr1}}$.

4. A special case of the general ratio estimator

A special case of $\hat{Y}_{\mathrm{sr1}}$ proposed by Srivastava (1985) is the Horvitz-Thompson estimator $\hat{Y}_{\mathrm{HT}} = \sum_{i\omega} y_i/\pi_i$. Let $\hat{Y}^*_{\mathrm{HT}}$ be the "Horvitz-Thompson estimator" of Y^* as if the y_i^*'s are unknown. A special case of $\hat{Y}$ is:

$$\hat{Y}_1 = (\hat{Y}_{\mathrm{HT}}/\hat{Y}^*_{\mathrm{HT}})Y^* = \left[\left(\sum_{i\omega} y_i/\pi_i\right) \bigg/ \left(\sum_{i\omega} y_i^*/\pi_i\right)\right] Y^*. \tag{4.1}$$

An obvious result of the estimator given by (4.1) is

THEOREM 4.1. *If* $\hat{Y}_{\mathrm{HT}}$ *and* $\hat{Y}^*_{\mathrm{HT}}$ *converge to* Y *and* Y^* *in probability respectively,* $\hat{Y}_1$ *converges to* Y *in probability.*

Under some mild conditions, $\hat{Y}_{\mathrm{HT}}$ and $\hat{Y}^*_{\mathrm{HT}}$ do converge to Y and Y^* in probability respectively. A summary of these results are given in Ouyang *et al.* (1991*a*). Note that $\hat{Y}$ converges to Y in probability implies $\hat{Y}$ is asymptotically unbiased. The form of $\hat{Y}_1$ looks similar to the ratio estimator given in Hájek (1981), but there is a conceptual difference between them since the y_i^* in $\hat{Y}_1$ can be decided using extraneous information. The estimator $\hat{Y}_1$ has a special application in the following situation.

Consider the Horvitz-Thompson estimator under fixed size sampling design. Suppose the π_i are proportional to the y_i^*. By replacing the r_i and $\pi_r(i)$ in (1.4) by n and π_i, we obtain an equation $ny_i^*/Y^* = \pi_i$. We state this result as a lemma.

LEMMA 4.1. *Suppose for a sampling design with fixed sample size* n, *the* π_i *are proportional to the* y_i^*. *Then* $r(\omega) \equiv 1$ *and* $r_i = n$ *constitute a solution to equations* (1.3) *and* (1.4).

Consider the following case. Suppose at the beginning of a sampling survey, guesses of y_i, say x_i, such that $x_i > 0$ for all i, are available. A sampling design is used such that the π_i are proportional to the x_i. Suppose after drawing a sample, new guess y_i^* of y_i, such that $y_i^*/Y^* = \theta_i$ which is close to x_i/X, are obtained, where X is the population total of x_i. (This situation happens, for example, when the x_i's are good estimates of the y_i's.) By Lemma 4.1, since $r(\omega) \equiv 1$ and $r_i \equiv 1/n$ are a solution to equations (1.3) and (1.4) when $y_i^* = x_i$ for all i, it follows that if $(y_1^*/Y^*, \ldots, y_N^*/Y^*)$ is "slightly" different from $(x_1/X, \ldots, x_N/X)$, $r(\omega) \equiv 1$ is an "approximate" solution of equations (1.3) and (1.4). Hence it is natural to consider estimator $\hat{Y}_1$ in this case. Note that all the results given in Section 3 apply to the estimator $\hat{Y}_1$.

5. Application to sampling with a regression model

For the estimator $\hat{Y}_{\text{sr}1}$ proposed by Srivastava (1985) and $\hat{Y}$ considered in this study, it is important that the introduced sample weight function be independent of the drawn sample. Consider a survey on a population which has the following linear model structure

$$(5.1) \qquad y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, \ldots, N,$$

where ϵ_i are uncorrelated random error with mean zero and variance $\sigma^2 x_i^\kappa$, and the κ is unknown but $1 < \kappa < 2$. In model (5.1), the x_i's are known, but α, β are unknown parameters. Suppose in the model, it is known that $\alpha \ll \beta x_i$. To use the $\hat{Y}_1$ given in (4.1), let x_i be the first guess of y_i. Then a pps sampling procedure of size n (i.e., sampling with inclusion probability proportional to size) is used to draw a sample ω. If after the survey, some idea about the α and β say α_0 and β_0, which does not depend on the sample, is available, we can simply use

$$(5.2) \qquad y_i^* = \alpha_0 + \beta_0 x_i, \qquad i = 1, \ldots, N,$$

as a second guess of y_i.

If such information about α and β is not available, α and β have to be estimated by the sample. Let $\boldsymbol{x} = (x_1, \ldots, x_N)'$, let $\mathbf{1}$ be a $N \times 1$ vector with 1 everywhere, then

$$(5.3) \qquad \begin{bmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{bmatrix} = [(\mathbf{1}, \boldsymbol{x})' \operatorname{diag}^{-1}(g_1, \ldots, g_N)(\mathbf{1}, \boldsymbol{x})]^{-1}(\mathbf{1}, \boldsymbol{x})' \operatorname{diag}^{-1}(g_1, \ldots, g_N)\boldsymbol{y}$$

is a "census fit" of α and β with weight $\operatorname{diag}^{-1}(g_1, \ldots, g_N)$, where $\operatorname{diag}^{-1}(g_1, \ldots, g_N)$ is a diagonal matrix with $g_1^{-1}, \ldots, g_N^{-1}$ appeared as the diagonal elements in order. Those $g_1, \ldots, g_N$ are decided upon by us, for instance, $g_i = x_i^k$ to get BLUE for α and β. An estimator of $\tilde{\alpha}$ and $\tilde{\beta}$ (with sample $\omega = \{1, \ldots, n\}$) is

$$(5.4)\quad \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \left[\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}' \operatorname{diag}^{-1}(\pi_1 g_1, \ldots, \pi_n g_n) \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}' \cdot \operatorname{diag}^{-1}(\pi_1 g_1, \ldots, \pi_n g_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Let

$$(5.5)\quad y_i^* = \hat{\alpha} + \hat{\beta} x_i, \qquad i = 1, \ldots, N$$

as a second guess of y_i. Replace them in (4.2) to get an estimator

$$(5.6)\quad \hat{Y}_{1e} = Y^* \left[\sum_{i\omega} y_i/\pi_i \right] \Big/ \left[\sum_{i\omega} y_i^*/\pi_i \right].$$

But now the second guess used in $\hat{Y}_{1e}$ depends on the sample. In order to show the unified theory developed in Section 3 can also be applied to $\hat{Y}_{1e}$ asymptotically, we need to show that

$$(5.7)\quad E(\hat{Y}_{1e} - Y)^2 - E(\hat{Y}_1 - Y)^2 \to 0,$$

where $\hat{Y}_1$ in (5.7) is the "estimator" obtained from (5.6) with $\hat{\alpha}$ and $\hat{\beta}$ replaced by $\tilde{\alpha}$ and $\tilde{\beta}$ in (5.5).

Notice that when we are talking the asymptotic behavior of an estimator in sampling, we always suppose that there is a sequence of finite populations such that the sizes of the populations go to infinity. Also a sampling design is assigned to each population. Let $\tilde{\theta}_i = \tilde{y}_i/\tilde{Y}$, $\theta_i^* = y_i^*/Y^*$.

THEOREM 5.1. *Under some mild conditions,* $\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i$ *converges to zero in* 4*th mean i.e.,* $E\left[\sum_{i\omega}(\theta_i^* - \theta_i)/\pi_i\right]^4 \to 0$.

The conditions and proof of Theorem 5.1 are given in the Appendix. Using Theorem 5.1, we have

THEOREM 5.2. *If* $E\left(\sum_{i\omega} y_i/\pi_i\right) / \left[\left(\sum_{i\omega} \theta_i^*/\pi_i\right)\left(\sum_{i\omega} \theta_i^*/\pi_i\right)\right]$ *is bounded, then under the condition of Theorem* 5.1, $\hat{Y}_{1e} - \hat{Y}_1$ *converges to zero in* 2*nd mean, i.e.,* $E(\hat{Y}_{1e} - \hat{Y}_1)^2 \to 0$.

PROOF. Use Theorem 5.1 to obtain

$$
\begin{aligned}
E(\hat{Y}_{1e} - \hat{Y}_1)^2 \\
&= E\left\{\left[\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i^*)/\pi_i\right]\left(\sum_{i\omega} y_i/\pi_i\right)\Big/\left[\left(\sum_{i\omega}\theta_i^*/\pi_i\right)\left(\widetilde{\sum_{i\omega}}\theta_i/\pi_i\right)\right]\right\}^2 \\
&\leq \left\{E\left[\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i\right]^4\right\}^{1/2}\left\{E\left[\frac{\sum_{i\omega} y_i/\pi_i}{\left(\sum_{i\omega}\theta_i^*/\pi_i\right)\left(\sum_{i\omega}\tilde{\theta}_i/\pi_i\right)}\right]^4\right\} \to 0.
\end{aligned}
$$

By Theorem 5.2, the results of the bias of $\hat{Y}_1$ given in Section 3 can be applied to $\hat{Y}_{1e}$ asymptotically. Theorem 5.2 implies $\hat{Y}_{1e} - \hat{Y}_1$ converges in mean. Since

$$E(\hat{Y}_{1e} - Y)^2 = E(\hat{Y}_1 - Y)^2 + 2E(\hat{Y}_1 - Y)(\hat{Y}_1 - \hat{Y}_{1e}) + E(\hat{Y}_1 - \hat{Y}_{1e})^2,$$

we have $E(\hat{Y}_1 - \hat{Y}_{1e})^2 \to 0$ by Theorem 5.2 and, if $E(\hat{Y}_1 - Y)^2$ is bounded,

$$|E(\hat{Y}_1 - Y)(\hat{Y}_1 - \hat{Y}_{1e})| \leq [E(\hat{Y}_1 - Y)^2]^{1/2}[E(\hat{Y}_1 - \hat{Y}_{1e})^2]^{1/2} \to 0,$$

so we have proved the following theorem.

THEOREM 5.3. *Under the conditions of Theorem* 5.1 *and* 5.2, *if* $E(\hat{Y}_1 - Y)^2$ *is bounded, then* (5.7) *is true.*

Under the condition of Theorem 5.3, the mean square error of $\hat{Y}_1$ is asymptotically the mean square error of $\hat{Y}_{1e}$. Hence, the estimator of the mean square error of $\hat{Y}_1$ is asymptotically the estimator of the mean square error of $\hat{Y}_{1e}$.

6. Simulations

We compare $\hat{Y}_{1e}$ and other commonly used estimators numerically. The populations used in the numerical comparisons are forestry populations (Schreuder *et al.* (1987)) which generally satisfy model (5.1) with $1 < k < 2$. Let $\hat{Y}_{1e}$ denote the estimator given in (5.6) with $g_i = x_i^{1.5}/\pi_i$ in (5.4), hence the $\hat{\alpha}$ and $\hat{\beta}$ given in (5.4) are the 'best linear estimators' of α and β based on the drawn sample as if $k = 1.5$ in model (5.1). We consider the classical ratio estimator (ratio of mean, denoted by $\hat{Y}_r$), the Horvitz-Thompson estimator ($\hat{Y}_{\mathrm{HT}}$), and the ordinary linear regression estimator ($\hat{Y}_{lr}$) and the best linear model unbiased estimator ($\hat{Y}_{\omega r}$) in our comparison. The last two estimators are given in (6.24) and (6.23) of Sukhatme *et al.* (1984). Another estimator to be compared is the general regression estimator, denoted by $\hat{Y}_{gr}$, where

$$\hat{Y}_{gr} = \sum_{i\omega} y_i/\pi_i + N\hat{\alpha}_1 + X\hat{\beta}_1 - \hat{\alpha}_1\sum_{i\omega} 1/\pi_i - \hat{\beta}_1\sum_{i\omega} x_i/\pi_i, \tag{6.1}$$

and $\hat{\alpha}_1$ and $\hat{\beta}_1$ are obtained from (5.4) by letting $g_i = 1$ for all i. This estimator was proposed by Särndal (1980) and denoted by T_{PI}. Another general regression estimator (T_{BLU}) has the form (6.1) except that $\hat{\alpha}$ and $\hat{\beta}$ were obtained from (5.4) by letting $g_i = x_i^{1.5}/\pi_i$, i.e., the $\hat{\alpha}_1$ and $\hat{\beta}_1$ used in T_{BLU} were the 'best linear estimators' of α and β with $k = 1.5$ in model (5.1). Among these two estimators, Särndal (1980) preferred T_{PI} (i.e., $\hat{Y}_{gr}$ in this study).

In the comparison, πps sampling is used with estimators $\hat{Y}_{\mathrm{HT}}$, $\hat{Y}_{gr}$, $\hat{Y}_{1e}$. For $\hat{Y}_r$, $\hat{Y}_{lr}$, $\hat{Y}_{wr}$, simple random sampling without replacement (SRSWOR) is applied. The reason for doing this is that estimators $\hat{Y}_r$, $\hat{Y}_{lr}$, $\hat{Y}_{wr}$ are usually used under SRSWOR, but estimators $\hat{Y}_{\mathrm{HT}}$, $\hat{Y}_{gr}$, $\hat{Y}_{1e}$ are developed based on πps sampling. The estimator $\hat{Y}_{gr}$ will become $\hat{Y}_{1l}$ when πps sampling is replaced by SRSWOR. Real forestry data sets described in Schreuder *et al.* (1987) are used, selected with the expectation that they would show strong linear, weak linear, or well defined curvilinear relationships between the variable of interest and covariate. The simulation biases of all estimators given in Table 1 are satisfactorily small. Table 1 shows that $\hat{Y}_{1e}$ is more efficient than $\hat{Y}_r$ and $\hat{Y}_{\mathrm{HT}}$ in general (better than $\hat{Y}_r$ in 12 cases out of 16, much better than $\hat{Y}_r$ in 8 cases). $\hat{Y}_{1e}$ is as efficient as $\hat{Y}_{gr}$ in all cases and they are close. $\hat{Y}_{1e}$ is also more efficient than $\hat{Y}_{lr}$ in 8 cases, much more so in 5 cases and is less efficient than $\hat{Y}_{lr}$ in 8 cases but only slightly so. $\hat{Y}_{wr}$ is usually not as good as $\hat{Y}_{lr}$.

$\hat{Y}_{lr}$ and $\hat{Y}_{gr}$ can also be compared in terms of robustness. Three forestry data sets are used here using Poisson sampling and Poisson-Poisson sampling (Ouyang *et al.* (1991*b*)). All three tree data consist of three variables: y, net volume of trees; x_1, diameter (at breast high) squared times height of trees; and x_2, ocular estimate of net volume of trees by experienced cruiser. Both x_1 and x_2 are highly correlated with y, but x_2 is more expensive to get. Some trees with large volume have $x_2 = 0$ because it is sometimes difficult to assess the true usable volume of a tree ocularly. Two data sets ($BLM1$ and $BLM2$) are seriously affected by these bad data points. They provide a good check on the robustness of the estimators. Besides $\hat{Y}_{1e}$, $\hat{Y}_{lr}$, $\hat{Y}_{gr}$, we also consider the unadjusted estimator (i.e., $\hat{Y}_{\mathrm{HT}}$) and the adjusted estimator in Poisson sampling (Brewer and Hanif (1983)) and three natural generalizations of the unadjusted and adjusted estimators for Poisson-Poisson sampling. These three estimators are denoted by $\hat{Y}_{\mathrm{I}}$, $\hat{Y}_{\mathrm{II}}$, $\hat{Y}_{\mathrm{III}}$ respectively, and they are totally unadjusted, adjusted (for sample size) in both stages and adjusted (for sample size) in the second stage only (Ouyang *et al.* (1991*b*)). For Poisson sampling, samples of expected size 30 were drawn with inclusive probability proportional to x_2 (with a small positive value added to the $x_2 = 0$ values). For Poisson-Poisson sampling, first stage Poisson sampling samples of expected size 50 were drawn with inclusion probability proportional to x_1, then second stage Poisson sampling samples of expected size 30 were drawn from the first stage samples with inclusion probability proportional to x_2. The simulation biases of all the estimators are acceptable except $\hat{Y}_{lr}$ has 6% bias in Poisson-Poisson sampling for population $BLM1$. The numerical comparison results are given in Table 2 respectively.

Both $\hat{Y}_{gr}$ and $\hat{Y}_{1e}$ are more efficient than the other estimators in almost all cases (except $\hat{Y}_{\mathrm{II}}$ is better than $\hat{Y}_{gr}$ and $\hat{Y}_{1e}$ for the 'good' data BFV in Poisson-Poisson sampling). $\hat{Y}_{1e}$ might be somewhat more robust than $\hat{Y}_{gr}$ for the two 'bad'

Table 1. Relative efficiency expressing the simulation mean square error of the ratio of means estimator as a fraction of the simulation mean square error for each estimator (C.L. = curvilinearity, W.L. = weak linearity, L = linearity, L.S. = linearity with considerably scatter, S.L. = strong linearity).

					Design					
					SRSWOR			PPS sampling		
Population	Variable	Size	R^2	Relationship	$\hat{Y}_r$	$\hat{Y}_{lr}$	$\hat{Y}_{wr}$	$\hat{Y}_{\mathrm{HT}}$	$\hat{Y}_{gr}$	$\hat{Y}_{1e}$
DF	H vs age	1703	.64	C.L.	1	1.037	1.104	1.200	1.255	1.268
$LOBM3$	DBT vs D	5163	.60	W.L.	1	1.250	1.254	1.072	1.271	1.274
$LOBM4$	DBT vs D	3989	.55	W.L.	1	1.085	1.066	1.012	1.049	1.053
$LOBM5$	DBT vs D	2618	.53	W.L.	1	1.068	1.042	0.974	0.986	0.987
$LOBM3$	H vs D	5163	.57	L.	1	1.834	1.776	1.056	1.766	1.789
$LOBM4$	H vs D	3989	.56	L.	1	1.355	1.286	1.024	1.318	1.321
$LOBM5$	H vs D	2618	.58	L.	1	1.293	1.233	0.918	1.112	1.118
$LOBM3$	SL vs D	3153	.53	L.S.	1	0.926	0.905	0.977	0.935	0.942
$LOBM4$	SL vs D	2019	.52	L.S.	1	0.993	1.059	0.975	1.070	1.071
$LOBM5$	SL vs D	1794	.47	L.S.	1	1.046	1.046	0.998	1.054	1.054
$LOBM3$	BL vs D	5039	.65	C.L.	1	0.970	0.907	1.085	0.928	0.955
$LOBM4$	BL vs D	3757	.62	C.L.	1	0.996	0.958	1.015	0.988	0.995
$LOBM5$	BL vs D	2464	.58	C.L.	1	1.000	0.957	0.996	0.923	0.924
$LOBM3$	$GBFV$ vs D^2BL	3153	.96	S.L.	1	1.460	1.626	1.388	1.976	2.066
$LOBM4$	$GBFV$ vs D^2BL	2019	.93	S.L.	1	1.529	1.530	1.280	1.843	1.860
$LOBM5$	$GBFV$ vs D^2BL	1794	.91	S.L.	1	1.378	1.377	1.199	1.488	1.512

Table 2. Iterated mean square error as % of population total.

			Poisson sampling					Poisson-Poisson sampling					
Population	Size	Iter. #	$\hat{Y}_u$	$\hat{Y}_a$	$\hat{Y}_{lr}$	$\hat{Y}_{gr}$	$\hat{Y}_{1e}$	$\hat{Y}_{\mathrm{I}}$	$\hat{Y}_{\mathrm{II}}$	$\hat{Y}_{\mathrm{III}}$	$\hat{Y}_{lr}$	$\hat{Y}_{gr}$	$\hat{Y}_{1e}$
$BLM1$	331	50,000	33.03	29.57	20.91	6.68	6.85	49.94	43.58	48.05	62.91	19.10	18.34
$BLM2$	510	200,000	31.21	25.36	23.68	6.78	6.41	66.97	63.78	65.91	55.91	18.33	17.59
BFV	5525	10,000	17.26	3.80	17.57	2.94	3.05	25.92	10.72	17.52	28.99	15.16	15.14

data sets $BLM1$ and $BLM2$. This may be due to the ratio estimator structure of $\hat{Y}_{1e}$. Based on these simulations, there is little to choose one estimator over another.

Appendix

PROOF OF THEOREM 3.5. Let $f(\theta) = E\{[e(\omega)]^2[1+\theta(\gamma(\omega)-1)]^{-2}\}$. Since the number of possible samples is finite, the order of E and the differentiating operator can be interchanged. Straightforward calculation gives

$$\{1/[n!]\}f^{(n)}(0) = (-1)^n(n+1)E\{[e(\omega)]^2(\gamma(\omega)-1)^n\}.$$

Let $1/\rho = \lim_{n\to\infty} |1/[n!]f^{(n)}(0)|^{1/n} = E\{|\gamma(\omega)-1|\}$. Since $E\{|\gamma(\omega)-1|\} < 1$, $\rho > 1$, the first order approximation in Taylor series expansion yields

$$E\{[e(\omega)]^2/[\gamma(\omega)]^2\} = f(1) \cong E\{[e(\omega)]^2\}. \tag{A.1}$$

Hence from (3.10) and (A.1),

$$E[\hat{Y}-Y]^2 = E\{[e(\omega)/\gamma(\omega)]^2\} \cong E[e(\omega)]^2 = \mathrm{Var}(e(\omega)).$$

PROOF OF THEOREM 3.6.

(i) In this case $s(\omega) = t(\omega)$ for all ω with $p(\omega) > 0$. So $\hat{Y}_{sr1} = \hat{Y}$.

(ii) When $y_i = \theta_i Y$ for all i, $\hat{Y} = Y$ for all ω with $p(\omega) > 0$ as shown in the proof of Theorem 3.4. On the other hand, under the given condition, we have $\mathrm{Var}(\hat{Y}_{sr1}) > 0$.

(iii) Because $y_i = \theta_i Y + \epsilon_i$ for all i,

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{sr1}) &= E(\gamma Y - Y + e)^2 \\ &= Y^2\,\mathrm{Var}(\gamma) + \mathrm{Var}(e) + 2YE[(\gamma-1)e].\end{aligned} \tag{A.2}$$

Now $E(e) = E(s(\omega)\sum_{i\omega}\epsilon_i/\pi_s(i)) = \sum_i \epsilon_i = 0$, and

$$\begin{aligned}E(\gamma e) &= E\left\{[s(\omega)]^2\sum_{i\omega}\theta_i\epsilon_i/[\pi_s(i)]^2\right\} + E\left\{[s(\omega)]^2\sum_{\substack{ij\omega\\ i\neq j}}\theta_i\epsilon_j/[\pi_s(i)\pi_s(j)]\right\} \\ &= \sum_i\{\theta_i\epsilon_i\pi_{s^2}(i)/[\pi_s(i)]^2\} + \sum_{i\neq j}\{\theta_i\epsilon_j\pi_{s^2}(i,j)/[\pi_s(i)\pi_s(j)]\}.\end{aligned}$$

Since $E[(\gamma-1)e] = E(\gamma e) - E(e)$, from (A.2), we have

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{sr1}) = Y^2\,\mathrm{Var}(\gamma) + \mathrm{Var}(e) &+ 2\sum_i\{\theta_i Y\epsilon_i\pi_{s^2}(i)/[\pi_s(i)]^2\} \\ &+ 2\sum_{i\neq j}\{\theta_i Y\epsilon_j\pi_{s^2}(i,j)/[\pi_s(i)\pi_s(j)]\}\end{aligned} \tag{A.3}$$

(Replace ϵ_i by $-|\epsilon_i|$ to obtain)

$$\geq Y^2 \operatorname{Var}(\gamma) + \operatorname{Var}(e) - 2\sum_i \{\theta_i Y|\epsilon_i|\pi_{s^2}(i)/[\pi_s(i)]^2\}$$
$$- 2\sum_{i\neq j}\{\theta_i Y|\epsilon_j|\pi_{s^2}(i,j)/[\pi_s(i)\pi_s(j)]\}$$

(Use (3.13) to obtain)

$$> Y^2 \operatorname{Var}(\gamma) + \operatorname{Var}(e)$$
$$- \delta\left\{\sum_i \{(\theta_i Y)^2 \pi_{s^2}(i)/[\pi_s(i)]^2\} - \sum_{i\neq j}\{\theta_i Y \theta_j Y \pi_{s^2}(i,j)/[\pi_s(i)\pi_s(j)]\}\right\}$$

(Use $\theta_1 + \cdots + \theta_N = 1$ to obtain)

$$= Y^2 \operatorname{Var}(\gamma) + \operatorname{Var}(e)$$
$$- \delta\left\{\sum_i \{(\theta_i Y)^2 \pi_{s^2}(i)/[\pi_s(i)]^2\} - \sum_{i\neq j}\{\theta_i Y \theta_j Y \pi_{s^2}(i,j)/[\pi_s(i)\pi_s(j)]\}\right\}$$
$$+ \delta(\theta_1 Y + \cdots + \theta_N Y)^2 - \delta Y^2$$
$$= (1-\delta)Y^2 \operatorname{Var}(\gamma) + \operatorname{Var}(e) - \delta Y^2.$$

Using the definition of δ, (A.3) implies

$$\operatorname{Var}(\hat{Y}_{sr1}) > \operatorname{Var}(e). \tag{A.4}$$

Now Theorem 3.5 implies

$$\operatorname{Var}(\hat{Y}_{sr1}) - \operatorname{MSE}(\hat{Y}) > \operatorname{Var}(e) - \operatorname{MSE}(\hat{Y}) \approx 0.$$

PROOF OF THEOREM 5.1. We assume that we have $\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i$ converges to 0 in probability. Under some mild conditions we do have this result. Convergence in probability of the Horvitz-Thompson estimator and regression estimator has been studied widely in literatures. A summary and a list of references are given by Ouyang *et al.* (1991*a*). A common condition in most of the references is that there are two constants a and b such that

$$0 < a \leq \pi_i \leq b \leq 1 \tag{A.5}$$

holds for all the π_i under each sampling design. Now we also suppose (A.5) is true. Then we have

(i) $\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i$ goes to zero in probability,

(ii) $|\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i| \leq |\sum_{i\omega}\theta_i^*/\pi_i| + |\sum_{i\omega}\tilde{\theta}_i/\pi_i| \leq |\sum_{i\omega}\theta_i^*|/a + |\sum_{i\omega}\tilde{\theta}_i|/a \leq 2/a$, for all the populations and all the designs.

By Theorem 1.3.6 in Serfling (1980), (i) and (ii) imply that $\sum_{i\omega}(\theta_i^* - \tilde{\theta}_i)/\pi_i$ converges to zero in 4th moment.

References

Brewer, K. R. W. and Hanif, M. (1983). Sampling with unequal probabilities, *Lecture Notes in Statistics*, **15**, Springer, New York.

Hájek, J. (1981). *Sampling from a Finite Population*, Marcel Dekker, New York.

Ouyang, Z. (1989). Investigation on some estimators and strategies in sampling, proposed by Srivastava, Ph.D. Thesis, Colorado State University, Fort Collins.

Ouyang, Z. and Schreuder, H. T. (1992). Srivastava estimation in forestry, *Forest Science* (submitted).

Ouyang, Z., Schreuder, H. T. and Li, J. (1991*a*). Regression estimation under sampling with one unit per stratum, *Comm. Statist. Theory Methods*, **20**, 2431–2449.

Ouyang, Z., Schreuder, H. T., Max, T. and Williams, M. (1991*b*). Poisson-Poisson and binomial-Poisson sampling in forestry, *Survey Methodology* (submitted).

Raj, D. (1972). *Sampling Theory*, McGraw-Hill, New York.

Rao, J. N. K. (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling, *J. Indian Statist. Assoc.*, **17**, 125–136.

Särndal, C. E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling, *Biometrika*, **67**, 639–650.

Schreuder, H. T., Li, H. G. and Hazard, J. W. (1987). PPS and random sampling estimation using some regression and ratio estimators for underlying linear and curvilinear models, *Forest Science*, **33**, 997–1009.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Srivastava, J. N. (1985). On a general theory of sampling, using experiment design, Concepts I: Estimation, *Bull. Internat. Statist. Inst.*, **51**(10.3), 1–16.

Srivastava, J. N. and Ouyang, Z. (1992). Studies on the general estimator in sampling, based on sample weight function, *J. Statist. Plann. Inference*, **31**, 177–196.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Survey with Application*, Iowa State University Press, Ames.