METRON - *International Journal of Statistics*
2003, vol. LXI, n. 1, pp. 75-90

HANI M. SAMAWI – MAHMOUD I. SIAM

Ratio estimation using stratified ranked set sample

Summary - Ratio estimation method is used to obtain a new estimator with higher precision for estimating the population mean or total. There are two methods for estimating ratios when the sampling design is simple stratified random sampling (SSRS), namely the combined and the separate ratio estimates. We study the performance of the combined and the separate ratio estimates using stratified ranked set sample (SRSS) introduced by Samawi (1996). Theoretical and simulation study as well as real data example are presented. It appears that using SRSS is more efficient than using SSRS for ratio estimation both in the case of combined and separate methods.

Key Words - Concomitant variables; Order statistics; Ranked set sample; Ratio estimators, Stratified sample.

1. Introduction

In many applications the quantity that is to be estimated from a random sample is the ratio of two variables both of which vary from unit to unit. The ratio of bilirubin level in jaundiced babies who stay in neonatal intensive care; to their weight at birth or the ratio of acres of wheat to total acres on a farm are two examples. Also, this method is used to obtain increased precision for estimating the population mean or total by taking advantage of the correlation between an auxiliary variable X and the variable of interest Y.

In the literature, ratio estimators are used in case of simple random sampling (SRS) as well as in case of SSRS. (See for example Cochran, 1977). There are two methods for estimating ratios that are generally used when the sampling design is SSRS, namely the combined ratio estimate and the separate ratio estimate.

McIntyre (1952) used the mean of n units based on a ranked set sample (RSS) to estimate the population mean. Samawi and Muttlak (1996) used RSS

Received September 2001 and revised November 2002.

to estimate the population ratio and showed that it provided a more efficient estimator compared to the one obtained by using SRS. Also, Samawi (1996) introduced the concept of stratified ranked set sampling (SRSS), to improve the precision of estimating the population mean.

In this paper we use the idea of SRSS instead of SSRS to improve the precision of the two methods for estimating ratios. Also, we study the properties of these estimators and compare them with other estimators.

In Section 2 and 3, we obtain the separate and combined ratio estimators using SRSS sample respectively. We also derive the asymptotic mean and variance of these estimators. The comparison between the two estimators (separate and combined) is discussed in terms of bias and efficiency in Section 4. Also, the results of our simulation study and the use of the two methods using real data about bilirubin level of baby's, who stay in neonatal intensive care, is discussed in Section 5 and 6 respectively.

1.1. Stratified ranked set sample

For the h-th stratum of the population, first choose r_h independent samples each of size r_h $h = 1, 2, \ldots, L$. Rank each sample, and use RSS scheme to obtain L independent RSS samples of size r_h, one from each stratum. Let $r_1 + r_2 + \ldots + r_L = r$. This complete one cycle of stratified ranked set sample. The cycle may be repeated m times until $n = mr$ elements have been obtained.

A modification of the above procedure is suggested here to be used for the estimation of the ratio using stratified ranked set sample. For the h-th stratum, first choose r_h independent samples each of size r_h of independent bivariate elements from the h-th subpopulation, $h = 1, 2, \ldots, L$. Rank each sample with respect to one of the variables say Y or X. Then use the RSS sampling scheme to obtain L independent RSS samples of size r_h one from each stratum. This complete one cycle of stratified ranked set sample. The cycle may be repeated m times until $n = mr$ bivariate elements have been obtained. We will use the following notation for the stratified ranked set sample when the ranking is on the variable Y. For the k-th cycle and the h-th stratum, the SRSS is denoted by

$$\{(Y_{h(1)k}, X_{h[1]k})(Y_{h(2)k}, X_{h[2]k}), \ldots, (Y_{h(r_h)k}, X_{h[r_h]k}) : k = 1, 2, \ldots, m;\ h = 1, 2, \ldots, L\},$$

where $Y_{h(i)k}$ is the i-th order statistic from the i-th set in the h-th stratum and $X_{h[i]k}$ is the corresponding concomitant variable (see Stokes, 1977).

1.2. Ratio estimate using SSRS

The parameter of interest to be estimated in this paper is $R = \frac{\mu_Y}{\mu_X}$. Using SSRS, we have two types of ratio estimators:

IN SEPARATE CASE. Following Levy and Lemeshow (1991), we have the following definition of separate ratio estimator. For the h-th stratum ($h = 1, 2, \ldots, L$), the ratio is defined by $\hat{R}_{hsrs} = \frac{\overline{Y}_h}{\overline{X}_h}$. Therefore, assuming that the subpopulations totals of the variable X are known, the separate ratio estimator using SSRS is given by

$$\hat{R}_{\mathrm{SSRS}(s)} = \sum_{h=1}^{L} \frac{T_h}{T} \hat{R}_{hsrs} = \sum_{h=1}^{L} \frac{\overline{Y}_h}{\overline{X}_h} \frac{T_h}{T} = \sum_{h=1}^{L} W_h \frac{\mu_{Xh}}{\mu_X} \frac{\overline{Y}_h}{\overline{X}_h} \tag{1.1}$$

where $W_h = \frac{N_h}{N}$, N_h is the h stratum size, N is the total population size, $T_h = N_h \mu_{Xh}, T = N\mu_X, \overline{Y}_h = \frac{\sum_{i=1}^{r_h} Y_{hi}}{r_h}, \overline{X}_h = \frac{\sum_{i=1}^{r_h} X_{hi}}{r_h}, \mu_{Xh}, h = 1, 2, \ldots, L$ are the known subpopulation specific means for the random variable X and $\mu_X = \sum_{h=1}^{L} W_h \mu_{Xh}$. Note that the subpopulations totals of the variable X need to be known in order to compute $\hat{R}_{\mathrm{SSRS}(s)}$ for estimating the population mean or total of the variable Y. It can be shown (see Hansen, 1953) that

$$E(\hat{R}_{\mathrm{SSRS}}(s)) = R + O\left(\frac{1}{\underset{h}{\mathrm{Min}}(mr_h)}\right)$$

and the variance of ratio estimator can be approximated by

$$\mathrm{Var}(\hat{R}_{\mathrm{SSRS}(s)}) \approx \sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2 R_h^2}{\mu_X^2 n_h} \times \left\{ C_{Xh}^2 + C_{Yh}^2 - 2\rho_{XhYh} C_{Xh} C_{Yh} \right\} \tag{1.2}$$

where $n_h = mr_h, C_{Xh} = \frac{\sigma_{Xh}}{\mu_{Xh}}, C_{Yh} = \frac{\sigma_{Yh}}{\mu_{Yh}}, \rho_{XhYh} = \frac{\sum_{i=1}^{Nh} (X_{hi} - \mu_{Xh})(Y_{hi} - \mu_{Yh})}{Nh \sigma_{Xh} \sigma_{Yh}}, R_h = \frac{\mu_{Yh}}{\mu_{Xh}}$, and σ_{Xh}, and σ_{Yh} are the standard deviations of the variable X and Y, respectively in the h-th stratum. Note that, equation (1.2) also can be derived easily from equation (6.45) in Cochran (1977) by dividing (6.45) by the population total of the variable X and simple algebra.

IN COMBINED CASE. Combined ratio estimator using SSRS is defined by

$$\hat{R}_{\mathrm{SSRS}(c)} = \frac{\overline{Y}_{\mathrm{SSRS}}}{\overline{X}_{\mathrm{SSRS}}} = \frac{\sum_{h=1}^{L} W_h \overline{Y}_h}{\sum_{h=1}^{L} W_h \overline{X}_h}. \tag{1.3}$$

It can be shown (see Hansen (1953) that

$$E(\hat{R}_{\mathrm{SSRS}(c)}) = R + O(\mathrm{Max}(n_h^{-1}))$$

and that the variance of the estimator is given by

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SSRS}(c)}) \approx R^2 \sum_{h=1}^{L} \frac{W_h^2}{n_h} \left\{ \frac{\sigma_{Yh}^2}{\mu_Y^2} + \frac{\sigma_{Xh}^2}{\mu_X^2} - 2\frac{\sigma_{Yh}}{\mu_Y} \frac{\sigma_{Xh}}{\mu_X} \rho_{XhYh} \right\}. \tag{1.4}$$

2. Separate ratio estimation using SRSS

2.1. Ratio estimation when ranking is on variable Y

The separate ratio estimate, requires knowledge of the stratum totals T_h (see Levy and Lemeshow, 1991) to estimate the population mean or the total. Using the notation introduced in Section (1.1) ratio is estimated as follows. Let

$$\overline{Y}_{h(r_h)} = \frac{1}{n_h} \sum_{k=1}^{m} \sum_{i=1}^{r_h} Y_{h(i)k},$$
$$\overline{X}_{h[r_h]} = \frac{1}{n_h} \sum_{k=1}^{m} \sum_{i=1}^{r_h} X_{h[i]k},$$

and $n_h = mr_h$. Then the separate ratio estimator using SRSS when the ranking on variable Y, is given by

$$\hat{R}_{\mathrm{SSRS1}(s)} = \sum_{h=1}^{L} W_h \frac{\mu_{Xh}}{\mu_X} \frac{\overline{Y}_{h(r_h)}}{\overline{X}_{h[r_h]}}\,. \tag{2.1}$$

It can be shown using Taylor expansion that

$$E(\hat{R}_{\mathrm{SSRS1}}(s)) = R + O\left(\frac{1}{\operatorname{Min}_h(mr_h)}\right)\,.$$

Also, the approximate variance of separate ratio is given by

$$\begin{aligned}\mathrm{Var}\,(\hat{R}_{\mathrm{SSRS1}(s)}) \approx \sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} \frac{R_h^2}{n_h} \Bigg\{ & C_{Xh}^2 + C_{Yh}^2 - 2\rho_{XhYh} C_{Xh} C_{Yh} \\ & - \frac{m}{n_h} \sum_{i=1}^{r_h} (M_{Xh[i]} - M_{Yh(i)})^2 \Bigg\}\end{aligned} \tag{2.2}$$

where, $M_{Xh[i]} = \frac{\mu_{Xh[i]} - \mu_{Xh}}{\mu_{Xh}}$, $M_{Yh(i)} = \frac{\mu_{Yh(i)} - \mu_{Yh}}{\mu_{Yh}}$, $E(Y_{h(i)}) = \mu_{Yh(i)}$ and $E(X_{h[i]}) = \mu_{Xh[i]}$. For evaluating the terms in the variances based on RSS see Stokes (1980). For evaluating the expectation and the variance of the concomitant variable of the order statistics see Stokes (1977).

2.2. Ratio estimation when ranking is on variable X

By changing the notation ( ) of perfect ranking by [ ] of imperfect ranking for X and Y (i.e. changing the role of the order statistics with the role of the concomitant variable of the order statistics) ratio is given by

$$\hat{R}_{\mathrm{SSRS2}(s)} = \sum_{h=1}^{L} W_h \frac{\mu_{Xh}}{\mu_X} \frac{\overline{Y}_{h[r_h]}}{\overline{X}_{h(r_h)}} \tag{2.3}$$

where

$$\overline{Y}_{h[r_h]} = \frac{1}{n_h} \sum_{k=1}^{m} \sum_{i=1}^{r_h} Y_{h[i]k} \quad \text{and} \quad \overline{X}_{h(r_h)} = \frac{1}{n_h} \sum_{k=1}^{m} \sum_{i=1}^{r_h} X_{h(i)k}\,.$$

Again we get the following results:

$$E(\hat{R}_{\mathrm{SSRS2}(s)}) = \frac{\mu_Y}{\mu_X} + O\left(\frac{1}{\operatorname*{Min}_h(mr_h)}\right)$$

and

$$\operatorname{Var}(\hat{R}_{\mathrm{SSRS2}(s)}) \approx \sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} \frac{R_h^2}{n_h} \left\{ C_{Xh}^2 + C_{Yh}^2 - 2\rho_{XhYh} C_{Xh} C_{Yh} - \frac{m}{n_h} \sum_{i=1}^{r_h} (M_{Xh(i)} - M_{Yh[i]})^2 \right\} \tag{2.4}$$

where, $M_{Xh(i)} = \frac{\mu_{Xh(i)} - \mu_{Xh}}{\mu_{Xh}}$, $M_{Yh[i]} = \frac{\mu_{Yh[i]} - \mu_{Yh}}{\mu_{Yh}}$, $E(Y_{h[i]}) = \mu_{Yh[i]}$ and $E(X_{h(i)}) = \mu_{Xh(i)}$.

Theorem 2.1. *Assume that the approximation to the variance of the ratio estimators in* (1.2), (2.2) *and* (2.4) *are valid and the bias of the estimators can be ignored for large m. Then*

$$\operatorname{Var}(\hat{R}_{\mathrm{SRSS1}(s)}) \le \operatorname{Var}(\hat{R}_{\mathrm{SSRS}(s)})$$

and

$$\operatorname{Var}(\hat{R}_{\mathrm{SRSS2}(s)}) \le \operatorname{Var}(\hat{R}_{\mathrm{SSRS}(s)})\,.$$

Proof. Take

$$\operatorname{Var}(\hat{R}_{\mathrm{SSRS}(s)}) - \operatorname{Var}(\hat{R}_{\mathrm{SRSS1}(s)}) = m\left\{\sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} \frac{R_h^2}{n_h^2} \left[\sum_{i=1}^{r_h} \left(M_{Xh[i]} - M_{Yh(i)}\right)^2\right]\right\} > 0.$$

Therefore, $\operatorname{Var}(\hat{R}_{\mathrm{SRSS1}(s)}) \le \operatorname{Var}(\hat{R}_{\mathrm{SSRS}(s)})$. Similarly, $\operatorname{Var}(\hat{R}_{\mathrm{SRSS2}(s)}) \le \operatorname{Var}(\hat{R}_{\mathrm{SSRS}(s)})$.

2.3. Which variable to rank?

Since we can not rank on both variables at the same time and some time it is easier to rank on one variable than the other, we need to decide to rank on variable X or Y.

Theorem 2.2. *Let us assume that there are L linear relationships between* Y_h *and* X_h*, i.e.,* $|\rho_h| > 0$ *and it is easy to rank on the variable* X*. Also, assume that the approximation to the variance of the ratio estimators* $\hat{R}_{\mathrm{SRSS1}(s)}$ *and* $\hat{R}_{\mathrm{SRSS2}(s)}$ *as given in equations* (2.2) *and* (2.4) *respectively are valid and the bias of the estimators can be ignored. Then*

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(s)}) \leq \mathrm{Var}\,(\hat{R}_{\mathrm{SSRS1}(s)})\,.$$

Proof. By looking at the two variances in equations (2.2) and (2.4) we need oly to compare $\sum_{i=1}^{r_h}(M_{Xh[i]} - M_{Yh(i)})^2$ with $\sum_{i=1}^{r_h}(M_{Xh(i)} - M_{Yh[i]})^2$. At this end, we note firstly that

$$\mu_{Xh[i]} = \begin{cases} \mu_{Xh(i)} & \text{for perfect ranking of } X \text{ in the } i\text{-th set} \\ \mu_{Xh} & \text{for imperfect ranking of } X \text{ in the } i\text{-th set}. \end{cases} \tag{2.5}$$

Consider the simple linear regression model of Y_h on X_h

$$Y_{hi} = \alpha_h + \beta_h X_{hi} + \varepsilon_{hi}, \tag{2.6}$$

$$\mu_{Yh} = \alpha_h + \beta_h \mu_{Xh} \tag{2.7}$$

where α_h and β_h are parameters and ε_{hi} is a random error with $E(\varepsilon_{hi}) = 0$, $\mathrm{Var}\,(\varepsilon_{hi}) = \sigma_h^2$, and $\mathrm{Cov}(\varepsilon_{hi}, \varepsilon_{hj}) = 0$ for $i \neq j, i = 1, 2, \ldots, r_h$. Also, ε_{hi} and X_{hi} are independent.

CASE 1. If we are ranking on the Y_h variable we get the following model from equation (2.6)

$$Y_{h(i)} = \alpha_h + \beta_h X_{h[i]} + \varepsilon_{h[i]}, \tag{2.8}$$

where $\varepsilon_{h[i]}$ is a random error with, $\mathrm{Var}\,(\varepsilon_{h[i]}) = \sigma^2_{[r_h]i}$ and $\mathrm{Cov}(\varepsilon_{h[i]}, \varepsilon_{h[j]}) = 0$ for $i \neq j, i = 1, 2, \ldots, r_h$ also $\varepsilon_{h[i]}$ and $X_{h[i]}$ are independent. Then,

$$\mu_{Yh(i)} = \alpha_h + \beta_h \mu_{Xh[i]}. \tag{2.9}$$

From (2.7) and (2.9) we get

$$M_{Yh(i)} = \frac{\beta_h M_{Xh[i]}}{R_h} \tag{2.10}$$

and now (2.2) can be written as

$$\operatorname{Var}(\hat{R}_{\text{SRSS1}(s)}) \approx \sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} \frac{R_h^2}{n_h} \left\{ C_{Xh}^2 + C_{Yh}^2 - 2\rho_{XhYh} C_{Xh} C_{Yh} - \frac{m}{n_h} \sum_{i=1}^{r_h} \mu_{Xh}^2 M_{Xh[i]}^2 \left(\frac{1}{\mu_{Xh}} - \frac{\beta_h}{\mu_{Yh}} \right)^2 \right\}.$$

CASE 2. If we are ranking on the variable X we get the following model:

$$Y_{h[i]} = \alpha_h + \beta_h X_{h(i)} = \varepsilon_{h[i]}\,. \tag{2.11}$$

The expected value of $Y_{h[i]}$ is

$$\mu_{Yh[i]} = \alpha_h + \beta_h \mu_{Xh(i)} + E(\varepsilon_{h[i]})\,. \tag{2.12}$$

Similarly, we can show that

$$M_{Yh[i]} = \frac{\beta_h M_{Xh(i)}}{R_h} \tag{2.13}$$

and now (2.4) can be written as

$$\operatorname{Var}(\hat{R}_{\text{SRSS2}(s)}) \approx \sum_{h=1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} \frac{R_h^2}{n_h} \left\{ C_{Xh}^2 + C_{Yh}^2 - 2\rho_{XhYh} C_{Xh} C_{Yh} - \frac{m}{n_h} \sum_{i=1}^{r_h} \mu_{Xh}^2 M_{Xh(i)}^2 \left(\frac{1}{\mu_{Xh}} - \frac{\beta_h}{\mu_{Yh}} \right)^2 \right\}.$$

Therefore, from (2.5) it is clear that $\operatorname{Var}(\hat{R}_{\text{SRSS2}(s)}) \le \operatorname{Var}(\hat{R}_{\text{SRSS1}(s)})$.

3. COMBINED RATIO ESTIMATION USING SRSS

3.1. Ratio estimation when ranking on variable Y

The combined ratio estimate using SRSS is defined by

$$\hat{R}_{\text{SRSS1}(c)} = \frac{\overline{Y}_{(\text{SRSS})}}{\overline{X}_{[\text{SRSS}]}} \tag{3.1}$$

where

$$\overline{Y}_{(\mathrm{SRSS})} = \sum_{h=1}^{L} W_h \overline{Y}_{h(r_h)}$$

and

$$\overline{X}_{[\mathrm{SRSS}]} = \sum_{h=1}^{L} W_h \overline{X}_{h[r_h]}$$

Therefore,

$$\hat{R}_{\mathrm{SRSS1}(c)} = \frac{\overline{Y}_{(\mathrm{SRSS})}}{\overline{X}_{[\mathrm{SRSS}]}} = \frac{\sum_{h=1}^{L} W_h \overline{Y}_{h(r_h)}}{\sum_{h=1}^{L} W_h \overline{X}_{h[r_h]}}. \tag{3.2}$$

For fixed r, assume that we have finite second moments for X and Y. Since the ratio is a function of the means of X and Y, i.e., $R = \frac{\mu_Y}{\mu_X}$, and hence R has at least bounded second order derivatives of all types in some neighborhood of (μ_Y, μ_X) provided that $\mu_X \neq 0$. Then, assuming large m, and by using the multivariate Taylor series expansion, we can approximate the variance and get the order of the bias of the ratio estimator as follows:

$$E(\hat{R}_{\mathrm{SRSS}(c)}) = R + O(\operatorname*{Max}_h(n_h^{-1}))$$

and

$$\begin{aligned} \mathrm{Var}\,(\hat{R}_{\mathrm{SRSS1}(c)}) \approx R^2 \sum_{h=1}^{L} \frac{W_h^2}{n_h} \Bigg\{ & \frac{\sigma_{Yh}^2}{\mu_Y^2} + \frac{\sigma_{Xh}^2}{\mu_X^2} - 2\frac{\sigma_{Xh}\sigma_{Yh}}{\mu_X \mu_Y} \rho_{XhYh} \\ & - \frac{m}{n_h} \sum_{i=1}^{r_h} \left(D_{Yh(i)} - D_{Xh[i]} \right)^2 \Bigg\} \end{aligned} \tag{3.3}$$

where, $D_{Xh[i]} = \frac{\mu_{Xh[i]} - \mu_{Xh}}{\mu_X}$ and $D_{Yh(i)} = \frac{\mu_{Yh(i)} - \mu_{Yh}}{\mu_Y}$.

3.2. Ratio estimation when ranking on variable X

In this case the estimate is given by:

$$\hat{R}_{\mathrm{SRSS2}(c)} = \frac{\overline{Y}_{[\mathrm{SRSS}]}}{\overline{X}_{(\mathrm{SRSS})}}$$

where

$$\overline{Y}_{[\mathrm{SRSS}]} = \sum_{h=1}^{L} W_h \overline{Y}_{h(r_h)}$$

and

$$\overline{X}_{(\mathrm{SRSS})} = \sum_{h=1}^{L} W_h \overline{X}_{h[r_h]} \,.$$

Therefore, in combined case, we get

$$\hat{R}_{\mathrm{SRSS2}(c)} = \frac{\overline{Y}_{[\mathrm{SRSS}]}}{\overline{X}_{(\mathrm{SRSS})}} = \frac{\sum_{h=1}^{L} W_h \overline{Y}_{h[r_h]}}{\sum_{h=1}^{L} W_h \overline{X}_{h(r_h)}} \,. \tag{3.4}$$

Using the same argument as in Section (3.1), $E(\hat{R}_{\mathrm{SRSS2}(c)}) = R + O(\underset{h}{\mathrm{Max}}(n_h^{-1}))$ and

$$\begin{aligned} \mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(c)}) \approx R^2 \sum_{h=1}^{L} \frac{W_h^2}{n_h} \Bigg\{ & \frac{\sigma_{Yh}^2}{\mu_Y^2} + \frac{\sigma_{Xh}^2}{\mu_X^2} - 2\frac{\sigma_{Xh}\sigma_{Yh}}{\mu_X \mu_Y}\rho_{XhYh} \\ & - \frac{m}{n_h} \sum_{i=1}^{r_h} \Big(D_{Yh[i]} - D_{Xh(i)} \Big)^2 \Bigg\}, \end{aligned} \tag{3.5}$$

where, $D_{Xh(i)} = \frac{\mu_{Xh(i)} - \mu_{Xh}}{\mu_X}$ and $D_{Yh[i]} = \frac{\mu_{Yh[i]} - \mu_{Yh}}{\mu_Y}$.

Theorem 3.1. *Assume that the approximations to the variance of the ratio estimators in* (1.4), (3.3) *and* (3.5) *are valid and the bias of the estimators can be ignored for large m. Then*

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS1}(c)}) \leq \mathrm{Var}\,(\hat{R}_{\mathrm{SSRS}(c)})$$

and

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(c)}) \leq \mathrm{Var}\,(\hat{R}_{\mathrm{SSRS}(c)}).$$

Proof. Similar to that for Theorem 2.1.

3.3. Ranking on which variable?

Again, since we cannot rank both variables at the same time we need to decide which variable we should rank. Therefore, we need to compare the variance of $\hat{R}_{\mathrm{SRSS1}(c)}$ in equation (3.3) and variance $\hat{R}_{\mathrm{SRSS2}(c)}$ in equation (3.5).

Theorem 3.2. *Assume that there are L linear relationships between Y_h and X_h, i.e., $|\rho_h| > 0$ and it is easy to rank variable X. Also assume that the approximation to the variance of the ratio estimators $\hat{R}_{\mathrm{SRSS1}(c)}$ and $\hat{R}_{\mathrm{SRSS2}(c)}$ given in equations* (3.3) *and* (3.5) *respectively are valid and the bias of the estimators can be ignored for large m. Then*

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(c)}) \leq \mathrm{Var}\,(\hat{R}_{\mathrm{SRSS1}(c)}).$$

Proof. Is similar to that of Theorem (2.2).

4. Comparison of the combined and separate estimates

Consider the case when ranking is on variable X. Equations (2.4) and (3.5) can be written respectively as

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(s)}) = \sum_{h=1}^{L} \frac{W_h^2 m}{n_h^2 \mu_X^2} \left[R_h^2 \sum_{i=1}^{r_h} \sigma^2_{Xh(i)} + \sum_{i=1}^{r_h} \sigma^2_{Yh[i]} - 2R_h \sum_{i=1}^{r_h} \sigma_{Xh(i)Yh[i]} \right]$$

and

$$\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(c)}) = \sum_{h=1}^{L} \frac{W_h^2 m}{n_h^2 \mu_X^2} \left[R^2 \sum_{i=1}^{r_h} \sigma^2_{Xh(i)} + \sum_{i=1}^{r_h} \sigma^2_{Yh[i]} - 2R \sum_{i=1}^{r_h} \sigma_{Xh(i)Yh[i]} \right]$$

where

$$\sigma^2_{Xh(i)} = \mathrm{Var}\left(X_{h(i)}\right), \;\; \sigma^2_{Yh[i]} = \mathrm{Var}\left(Y_{h[i]}\right) \quad \text{and} \quad \sigma_{Xh(i)Yh[i]} = \mathrm{Cov}\left(X_{Xh(i)}, Y_{Yh[i]}\right).$$

Thus,

$$\begin{aligned}\mathrm{Var}(\hat{R}_{\mathrm{SRSS2}(c)}) - \mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(s)}) &= \frac{m}{\mu_X^2} \sum_{h=1}^{L} \frac{W_h^2}{n_h^2} \Bigg[(R^2 - R_h^2) \sum_{i=1}^{r_h} \sigma^2_{Xh(i)} \\ &\quad - 2(R - R_h) \sum_{i=1}^{r_h} \sigma_{Xh(i)Yh[i]} \Bigg] \\ &= \frac{m}{\mu_X^2} \sum_{h=1}^{L} \frac{W_h^2}{n_h^2} \Bigg[(R - R_h)^2 \sum_{i=1}^{r_h} \sigma^2_{Xh(i)} + 2(R_h - R) \\ &\quad \times \left(\sum_{i=1}^{r_h} \sigma_{Xh(i)Yh[i]} - R_h \sum_{i=1}^{r_h} \sigma^2_{Xh(i)} \right) \Bigg].\end{aligned}$$

As in the case of SSRS (see Cochran, 1977), if the ratio estimate is valid, the last term on the right is usually small. (It vanishes if within each stratum the relationship between $Y_{h[i]}$ and $X_{h(i)}$ is a straight line through the origin).

Also, as in Cochran (1977), unless R_h is constant from stratum to stratum, the use of a separate ratio estimate in each stratum is likely to be more precise if the sample in each stratum is large enough so that the approximate formula for $\mathrm{Var}\,(\hat{R}_{\mathrm{SRSS2}(s)})$ is valid and the cumulative bias that can effect the ratio estimate is negligible. With only a small sample in each stratum, the combined estimate is to be recommended.

Similarly, we can show that similar conclusions hold in the case when ranking on Y is perfect and ranking of X is not perfect.

5. Simulation study

5.1. Design of the simulation study

We did computer simulation to gain insight in the properties of the ratio estimator. Bivariate random observations were generated from a bivariate normal distribution with parameters $\mu_{Xh}, \mu_{Yh}, \sigma_{Xh}, \sigma_{Yh}$, $h = 1, 2, \dots, L$ and correlation coefficient ρ. Also we divide the data in the sample into three strata and in some cases into four strata. The simulation was performed with $r = 10, 20, 30$ and with $m = 1$ for the SRSS, SSRS, RSS and SRS data sets. The ratio of the population means were estimated for these sampling methods. Using 2000 replications, estimates of the means and mean square errors were computed.

We considered ranking on either variable Y or X. Results of these simulations are summarized by the relative efficiencies of the estimators of the population ratio and by the bias of estimation for different values of the correlation coefficient ρ. In order to reduce the size of this paper we present two tables only. Tables 1 gives efficiency when ranking is perfect on variable X and Y respectively. Tables 2 gives the bias in estimation when ranking is perfect on variable X and Y respectively.

5.2. Results of the simulation study

We conclude that the largest gain in efficiency is obtained by ranking the variable X and with large values of negative ρ. (For example in Table 1, the relative efficiency when $\rho = 0.90$ and $r = 30$ is 1.97 while it is 4.16 when $\rho = -.90$ and $r = 30$. The results of simulation indicates that, when ranking is on the variable X or Y, the efficiency will decrease with decreasing values of ρ from 0.99 to 0, and start to increase as ρ decreases from 0 to -.99. However, in the separate case this conclusion may be changed when $r = 10$ as we indicated in Section 4, when the sample size is small we cannot use the separate ratio estimate. Moreover, the efficiency will increase when the set size (r) is increased.

Also, there will be no change in the efficiency if the sample size is increased by increasing the number of cycles. Also, we note that in combined case for any values of r or ρ

$$MSE(\hat{R}_{\mathrm{SRSS}}) \leq MSE(\hat{R}_{\mathrm{RSS}}) \leq MSE(\hat{R}_{\mathrm{SSRS}}) \leq MSE(\hat{R}_{\mathrm{SRS}})$$

when $R = 1.45$, $W_1 = 0.3$, $W_2 = 0.3$ and $W_3 = 0.4$ and have equals variances within strata. This is not completely true for different cases, e.g., when $R = 1.17$ or when variances within strata are not equal.

TABLE 1. *Relative effeciency of ratio estimators using SRSS relative to SSRS.*

W_h: .3 / .3 / .4 $R = 1.45$		μ_{Xh}: 2 / 3 / 4 μ_{Xh}: 3 / 4 / 6		σ_{Xh}: 1 / 1 / 1 σ_{Yh}: 1 / 1 / 1	
ρ	r	Ranking on Variable X		Ranking on Variable Y	
		Combined	Separate	Combined	Separate
.99	10	1.97	13.76	1.75	258.47
	20	2.98	3.66	2.17	2.81
	30	3.35	3.78	2.77	3.29
.90	10	1.52	8.18	2.00	0.43
	20	1.81	1.92	1.17	12.34
	30	1.97	2.18	1.27	1.37
.70	10	1.33	764.91	1.07	1.28
	20	1.54	1.82	0.98	1.05
	30	1.69	1.83	1.01	1.06
.50	10	1.43	28.96	1.06	0.88
	20	1.61	1.90	1.03	1.01
	30	1.76	2.13	1.03	0.99
.25	10	1.56	1.33	1.17	14.54
	20	1.77	2.13	1.14	1.38
	30	1.87	2.01	1.27	1.13
−.25	10	1.49	2.85	1.26	0.70
	20	2.32	2.66	1.42	1.29
	30	2.47	2.54	1.61	1.43
−.50	10	1.97	105.26	1.56	6.02
	20	2.46	2.97	1.69	1.62
	30	2.75	3.22	2.00	1.97
−.70	10	1.82	17.51	1.88	3.02
	20	2.84	3.26	2.27	2.22
	30	3.38	3.61	2.58	2.70
−.90	10	2.28	106.16	1.95	4.95
	20	3.37	3.52	2.74	2.92
	30	4.16	4.58	3.88	4.06
−.99	10	2.07	8.16	2.09	9.61
	20	3.29	4.04	3.11	3.62
	30	4.20	4.23	4.79	5.21

From Tables 2 it appears that the bias of $\hat{R}_{\text{SRSS}}$ is higher when ρ is negative than when it is positive. For example, the bias when $\rho = 0.99$ and $r = 30$ is 0.0016 while the bias when $\rho = -0.99$ and $r = 30$ is 0.0065. However, in most cases the bias is less than 0.01 but for small r the bias in separate case exceeds 0.01.

TABLE 2. *Bias of ratio estimators using SRSS and SSRS.*

W_h: .3 / .3 / .4 $R = 1.45$		μ_{Xh}: 2 / 3 / 4 μ_{Xh}: 3 / 4 / 6				σ_{Xh}: 1 / 1 / 1 σ_{Yh}: 1 / 1 / 1			
ρ	r	Ranking on Variable X				Ranking on Variable Y			
		Combined		Separate		Combined		Separate	
		SRSS	SRSS	SRSS	SRSS	SRSS	SRSS	SRSS	SRSS
	10	.0042	.0050	.0114	.0295	−.0034	−.0112	−.0160	−.0395
.99	20	.0018	.0024	.0027	.0095	−.0013	−.0029	−.0056	−.0146
	30	.0016	.0019	.0016	.0062	−.0020	−.0041	−.0052	−.0133
	10	.0060	.0068	.0146	.0336	−.0001	−.0031	−.0044	.0327
.90	20	.0025	.0032	.0040	.0106	−.0009	−.0027	−.0018	−.0118
	30	.0010	.0019	.0006	.0065	.0003	−.0029	−.0006	−.0090
	10	.0082	.0060	.0212	.1130	.0023	−.0013	.0148	−.0022
.70	20	.0037	.0048	.0071	.0171	.0045	−.0044	.0137	−.0056
	30	.0005	.0034	.0011	.0110	.0005	.0006	.0056	−.0006
	10	.0061	.0100	.0223	.0668	.0053	.0099	.0331	.0304
.50	20	.0058	.0019	.0096	.0194	.0066	.0043	.0203	.0105
	30	−.0006	.0024	.0004	.0120	.0057	.0074	.0154	.0101
	10	.0129	.0186	.0388	.0714	.0077	.0054	.0690	.0542
.25	20	.0015	.0015	.0069	.0254	.0075	−.0027	.0316	.0129
	30	.0007	−.0004	.0023	.0119	.0088	.0104	.0214	.0203
	10	.0211	.0107	.0517	.0517	.0069	.0264	.0810	.0732
−.25	20	.0032	.0084	.0101	.0376	.0061	.0080	.0	.0467
	30	.0021	.0046	.0061	.0220	.0045	.0083	.0238	.0357
	10	.0173	.0227	.0565	.1588	.0186	.0402	.1078	.1617
−.50	20	.0102	.0120	.0196	.0499	−.0036	.0124	.0209	.0601
	30	.0048	.0028	.0096	.0268	.0003	.0043	.0148	.0354
	10	.0177	.0176	.0601	.1395	−.0007	.0453	.0653	.2204
−.70	20	.0097	.0145	.0206	.0526	.0023	.0166	.0255	.0706
	30	.0041	.0083	.0089	.0313	.0060	.0152	.0193	.0516
	10	.0183	.0265	.0630	.0733	.0088	.0442	.0772	.2112
−.90	20	.0101	.0166	.0230	.0592	.0055	.0189	.0272	.0832
	30	.0040	.0054	.0073	.0325	.0033	.0220	.0129	.0646
	10	.0309	.0336	.0776	.1302	−.0002	.0455	.0601	.2400
−.99	20	.00	.0107	.0159	.0549	−.0028	.0288	.0142	.1024
	30	.0065	.0112	.0126	.0381	−.0038	.0103	.0029	.0532

Also, the bias will decrease when the sample size is increased by increasing r. The bias in the combined case is always less than the corresponding bias in the separate case. Similar conclusions can be drawn when ranking on the variable Y. However, the bias when ranking is on Y is slightly lower than when ranking is on X. Moreover, it is clearly from equation (1.2) and (1.4) that for negativethe ρ the variance of the ratio estimators, in case of separate and combied methods, are larger than for positive ρ. However, our simulation indicated that the efficiency of using SRSS, for ratio estimation, is higher when ρ is negative. This may be due that the correlation between the order statistics is always positive.

6. Ratio of bilirubin level to weight at birth

We give a real life example about Bilirubin level in jaundiced babies who stay in neonatal intensive care. Most birth surveys on live newborns showed that jaundice is common. Jaundice in new born can be pathological and physiological. It start on second day of life and it has relationship with race, method of feeding and gestational age.

On the other hand if the total serum bilirubin in blood is above 1.5 mg/dl then we classify it as hyper-bilirubin. Neonatal jaundice is define as yellowish discoloration of skin and sclera and it occurs if bilirubin level is more than 5mg/dl. (see Nelson *et al.*, 1994).

Neonatal jaundice usually appears on the second day of life. Most of normal newborn babies leave the hospital after 24 hours of life. Therefore, our primary concern will be on babyies staying in neonatal intensive care. Physicians are interested in the jaundice, because of the risk on the hearing, brain and death. We will focus on the ratio of the level of bilirubin to the weight at birth for the newborn babies.

The data was collected from five hospitals in Jordan. The jaundice is measured by the level of bilirubin in the blood. This level is determined according to a blood test (TSB), which takes nearly 30 minutes. Moreover, ranking on the level of bilirubin in the blood can be done visually by observing the following:

(i) Color of the face.
(ii) Color of the chest.
(iii) Color of the lower parts of the body and
(iv) the color of the terminal parts of the whole body.

Then as the yellowish goes from (i) to (iv) the bilirubin level in the blood goes higher. We present below the analysis of the collected data for 120 babies according to their weight at birth, sex and bilirubin level.

For illustration, assume that the collected data of 120 babies from the five hospitals is the study population. Denote the bilirubin level by Y and the weight at birth by X. Since there are two strata, $L = 2$, $m = 2$ and $r = 10$, then $n = r.m = 20$, $W_1 = \frac{72}{120} = 0.6$ and $W_2 = \frac{48}{120} = 0.4$. Therefore, for male babies

$n_1 = mr_1 = 0.6 \times 20 = 12$ and or for female babies $n_1 = mr_2 = 0.4 \times 20 = 8$. Also, the parameter of interest to be estimated is $R = 3.90$.

Using the sampling schemes of SRSS and SSRS, Table 3 contains the two selected samples. Note that the ranking was on variable X (weight).

TABLE 3. *The selected samples.*

	SRSS sample				SSRS sample			
	Female		Male		Female		Male	
	X kg	Y mg/dl	X kg	Y Mg/dl	X kg	Y mg/dl	X kg	Y mg/dl
cycle 1	2.80	9.30	2.43	10.80	3.00	5.90	3.60	9.50
	3.00	5.50	2.60	7.70	2.85	13.10	3.15	1.41
	2.85	13.10	3.20	6.12	3.15	7.80	2.60	10.94
	3.15	7.80	2.95	9.41	1.55	8.82	3.10	23.41
			3.85	15.76			3.70	12.82
			4.15	21.29			2.70	15.47
cycle 2	1.55	8.82	1.40	10.94	2.60	9.24	2.45	8.71
	2.10	20.41	1.90	11.88	1.50	8.51	3.65	16.20
	2.60	9.24	2.50	13.60	2.53	11.50	1.85	9.20
	3.00	12.55	3.15	29.24	2.65	5.40	2.80	7.06
			3.10	12.30			3.10	12.30
			3.70	5.50			2.20	7.60

Based on the SRSS and SSRS, Table 4 contains the results of the illustration.

TABLE 4. *Summary of the results of the illustration using Bilirubin data.*

	$\hat{R}_{\text{SSRS}(s)}$	$\hat{R}_{\text{SSRS}(C)}$	$\hat{R}_{\text{SRSS}(s)}$	$\hat{R}_{\text{SRSS}(C)}$
Estimate	3.6	3.68	4.32	4.32
Estimated Variance	0.14	0.13	0.13	0.12

Finally, we get $\text{eff}(\hat{R}_{\text{SRSS2}(s)}, \hat{R}_{\text{SRSS2}(c)}) = 1.08$, $\text{eff}(\hat{R}_{\text{SSRS2}(s)}, \hat{R}_{\text{SSRS2}(c)}) = 1.07$, $\text{eff}(\hat{R}_{\text{SSRS}(c)}, \hat{R}_{\text{SRSS2}(c)}) = 1.1$, $\text{eff}(\hat{R}_{\text{SSRS}(s)}, \hat{R}_{\text{SRSS2}(s)}) = 1.1$.

ACKNOWLEDGMENTS

The authors would like to thank the referees for their comments which were helpful in improving the paper.

REFERENCES

COCHRAN, W. G. (1977) *Sampling Techniques*, Third edition, John Wiley & Sons.

DELL, T. R. and CLUTTER, J. L. (1972) Ranked set sample theory with order statistics background, *Biometrics*, 28, 545-555.

HANSEN, M. H., HURWITZ., W. N., and MADOW, W. G. (1953) *Sampling survey methods and theory*, Vol. 2. John Wiley & Sons, New York.

LEVY P. S. and LEMESHOW S. (1991) *Sampling of populations methods and applications*, John Wiley & Sons, New York.

MCINTYRE, G. A. (1952) A method of unbiased selective sampling using ranked sets. Australian, *J. Agricultural Research*, 3, 385-390.

NELSON, W. E., BEHRMAN, R. E., KLIEGMAN, R. M., and VAUGHAN, V. C. (1994) *Textbook of Pediatrics*, 4-th edn, W.B. Saunders Company Harcourt Barace Jovanovich, Inc.

SAMAWI, H. M. (1996) Stratified ranked set sample, *Pakistan J. of Stat.*, Vol. 12 (1), 9 - 16.

SAMAWI, H. M. and MUTTLAK, H. A. (1996) Estimation of ratio using rank set sampling, *Biom. Journal*, 38, 753-764.

STOKES S. L. (1977) Ranked set sampling with concomitant variables, *Comm. Statist. -Theor. Meth.*, 12 (6), 1207-1211.

STOKES S. L. (1980) Estimation of the variance using judgment order ranked set samples, *Biometrics*, 36, 35-42.

HANI M. SAMAWI
Department of Mathematics & Statistics
Sultan Qaboos University
P.O.Box 36
Al-khod 123, Sultanate of Oman
hsamawi@squ.edu.om

MAHMOUD I. SIAM
Department of Mathematics & Statistics
Sultan Qaboos University
P.O.Box 36
Al-khod 123, Sultanate of Oman