New Product-Type Estimators for the Population Mean Using Quartiles of the Auxiliary Variable

Amer Ibrahim Falah Al-Omari

Abstract—In this paper, we suggest new product-type estimators for the population mean of the variable of interest exploiting the first or the third quartile of the auxiliary variable. We obtain mean square error equations and the bias for the estimators. We study the properties of these estimators using simple random sampling (SRS) and ranked set sampling (RSS) methods. It is found that, SRS and RSS produce approximately unbiased estimators of the population mean. However, the RSS estimators are more efficient than those obtained using SRS based on the same number of measured units for all values of the correlation coefficient.

Keywords—Product estimator; auxiliary variable; simple random sampling, extreme ranked set sampling.

I. INTRODUCTION

Let \( X \) and \( Y \) denotes the auxiliary variable and the variable of interest respectively, with means \( \mu_X, \mu_Y \), variances \( \sigma_X^2, \sigma_Y^2 \) respectively and correlation coefficient \( \rho \). Let \( (X_i, Y_i), i = 1, 2, \ldots, m \) in the \( j \)th cycle \( j = 1, 2, \ldots, n \) denote the measurements of \( X \) and \( Y \). Assuming the population mean \( \mu_X \) of the auxiliary variable \( X \) is known, the classical product estimator \( \hat{Y}_{CSRS} \) for \( \mu_Y \) using SRS is defined as

\[
\hat{Y}_{CSRS} = \frac{\sum_{i=1}^{m} X_i Y_{i1j}}{mn},
\]

where,

\[
\overline{X}_{SRS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} X_{i1j} \quad \text{and} \quad \overline{Y}_{SRS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} Y_{i1j}.
\]

Singh and Tailor (2003) suggested a modified product estimator \( \hat{Y}_{MSRS} \) for \( \mu_Y \) using SRS from a sample of size \( m \) as given by

\[
\hat{Y}_{MSRS} = \frac{\overline{X}_{SRS}}{\mu_X} \left( \frac{\overline{X}_{SRS} + \rho}{\mu_X + \rho} \right).
\]


In this paper, we have suggested new product-type estimators for estimating the population mean \( \mu_Y \) assuming the knowledge of the first or third quartile of the auxiliary variable \( X \). We compare the properties of the estimators under SRS and RSS methods.

This paper is organized as follows: in Section 2.1, we present the product-type estimators using SRS with their properties. In Section 2.2 we present the RSS product-type estimators with their properties. In Section 3 simulation study is conducted to investigate the performance of the suggested estimators under RSS with respect to SRS based on the same number of measured units. Conclusion is presented in Section 4.

II. THE SUGGESTED ESTIMATORS

Let \( (X_1, Y_1), (X_2, Y_2), \ldots, (X_m, Y_m) \) be a bivariate random sample with pdf \( f(x, y) \), cdf \( F(x, y) \), with means \( \mu_X, \mu_Y \), variances \( \sigma_X^2, \sigma_Y^2 \) and correlation coefficient \( \rho \). In this paper we assume that the ranking is performed on the variable \( X \) to estimate the mean of the variable of interest \( Y \). Let \( (X_{11j}, Y_{11j}), (X_{12j}, Y_{12j}), \ldots, (X_{1mj}, Y_{1mj}), (X_{21j}, Y_{21j}), (X_{22j}, Y_{22j}), \ldots, (X_{2mj}, Y_{2mj}) \),  \( \ldots, (X_{mj1}, Y_{mj1}), (X_{m2j}, Y_{m2j}), \ldots, (X_{mjj}, Y_{mjj}) \), \( i = 1, 2, \ldots, m, j = 1, 2, \ldots, n \) be \( m \) independent random samples each of size \( m \).
A. The SRS product-type estimators

The product-type estimators of the population mean $\mu_Y$ using SRS are given by

$$\hat{Y}_{SRS1} = \bar{Y}_{SRS} \left( X_{SRS} + q_1 \right) / \mu_X + q_1,$$

and

$$\hat{Y}_{SRS3} = \bar{Y}_{SRS} \left( X_{SRS} + q_3 \right) / \mu_X + q_3,$$

where $q_1$ and $q_3$ are the first and third quartiles of the auxiliary variable $X$ respectively, and where $X_{SRS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} X_{ij}$ and $\bar{Y}_{SRS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} Y_{ij}$. Using Taylor expansion as

$$h\left( X_{SRS}, \bar{Y}_{SRS} \right) \approx h\left( \mu_X, \mu_Y \right) + \left( \bar{Y}_{SRS} - \mu_Y \right) \left( \frac{\partial h}{\partial Y} \right)_{\mu_X=\mu_Y} + \left( \bar{Y}_{SRS} - \mu_Y \right)^2 \left( \frac{\partial^2 h}{\partial Y^2} \right)_{\mu_X=\mu_Y},$$

where, $h\left( X_{SRS}, \bar{Y}_{SRS} \right) = \hat{Y}_{SRS} - k$, and $h\left( \mu_X, \mu_Y \right) = \mu_Y$, the estimators in (3) to the first degree of approximation can be written as

$$\hat{Y}_{SRS} \approx \bar{Y}_{SRS} + T \left( X_{SRS} - \mu_X \right),$$

where $T = \frac{\mu_Y}{\mu_X + q_k}$. The bias and the MSE of the estimator, respectively, are

$$\text{Bias} \left( \hat{Y}_{SRS} \right) = E \left( \hat{Y}_{SRS} \right) - \mu_Y$$

$$\approx E \left( \bar{Y}_{SRS} + T \left( X_{SRS} - \mu_X \right) \right) - \mu_Y = 0,$$

and

$$\text{Var} \left( \hat{Y}_{SRS} \right) \approx T^2 \text{Var} \left( X_{SRS} \right) + \text{Var} \left( \bar{Y}_{SRS} \right) + 2T \text{Cov} \left( X_{SRS}, \bar{Y}_{SRS} \right).$$

Using the two relations

$$\text{Cov} \left( X_{SRS}, \bar{Y}_{SRS} \right) = \beta \text{Var} \left( X_{SRS} \right) \mu_X + q_k,$$

and

$$\text{Var} \left( \bar{Y}_{SRS} \right) \approx \frac{\sigma_Y^2}{m},$$

\begin{align*}
\text{Var} \left( \hat{Y}_{SRS} \right) & \approx T^2 \frac{\sigma_Y^2}{m} + \frac{\beta^2 \sigma_Y^2}{m} + \frac{2T \sigma_Y^2}{m} \left( \mu_X + q_k \right) + \frac{\sigma_Y^2 \mu_X}{m} \\
& \approx \frac{\beta^2 \sigma_Y^2}{m} + \frac{2T \sigma_Y^2}{m} \left( \mu_X + q_k \right) + \frac{\sigma_Y^2 \mu_X}{m},
\end{align*}

where

$$\beta = \frac{\sigma_Y}{\sigma_X}, \text{Var} \left( X_{SRS} \right) = \frac{\sigma_X^2}{m}, \text{Var} \left( \bar{Y}_{SRS} \right) = \frac{\sigma_Y^2}{m},$$

the variance of $\hat{Y}_{SRS}$ can be written as

$$\text{Var} \left( \hat{Y}_{SRS} \right) \approx \frac{\sigma_Y^2}{m} \left( T + \beta \right)^2 + \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right).$$

Note that the Variance and the MSE of the estimators are the same since the estimators are approximately unbiased.

B. RSS product estimator

Let $\left( X_{(1m)}, Y_{(1m)} \right), \left( X_{(2m)}, Y_{(2m)} \right), ...,$

$$\left( X_{(mn_m)}, Y_{(mn_m)} \right) \text{ be the order statistics of } X_{ij},$$

and the judgment order of $Y_{i1}, Y_{i2}, ..., Y_{im}$, $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. The RSS involves randomly selecting $m^2$ units from the population. These units are randomly allocated into $m$ sets, each of size $m$. The $m$ units of each sample are ranked visually or by any inexpensive method with respect to the variable of interest. From the first set of $m$ units, the smallest unit is measured. From the second set of $m$ units, the second smallest unit is measured. The process is continued until from the $m$th set of $m$ units the largest unit is measured. Repeating the process $n$ times yields a set of size $mn$ from the initial $mn^2$ units. Then

$$\left( X_{(1m)}, Y_{(1m)}, X_{(2m)}, Y_{(2m)}, ..., X_{(mn)}, Y_{(mn)} \right)$$

denote the RSS of size $m$. The RSS product-type estimator of the population mean of $Y$ from a sample of size $m$ is defined as

$$\hat{Y}_{RSS1} = \bar{Y}_{RSS} \left( X_{RSS} + q_1 \right) / \mu_X + q_1,$$

$$\hat{Y}_{RSS3} = \bar{Y}_{RSS} \left( X_{RSS} + q_3 \right) / \mu_X + q_3,$$

where

$$X_{RSS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} X_{ij}, \text{ and } \bar{Y}_{RSS} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} Y_{ij}.$$

Using (4), the estimator given in (11) can be written as

$$\hat{Y}_{RSS} \approx \bar{Y}_{RSS} + T \left( X_{RSS} - \mu_X \right).$$

The bias and the MSE of (12), respectively, are given by

$$\text{Bias} \left( \hat{Y}_{RSS} \right) \approx E \left( \hat{Y}_{RSS} \right) - \mu_Y$$

$$\approx E \left( \bar{Y}_{RSS} + T \left( X_{RSS} - \mu_X \right) \right) - \mu_Y = 0,$$
and
\[
\text{Var} \left( \hat{Y}_{RSS} \right) \cong T^2 \text{Var} \left( X_{RSS} \right) + \text{Var} \left( Y_{RSS} \right) + 2T \text{Cov} \left( X_{RSS}, Y_{RSS} \right).
\]  \hspace{1cm} (14)

where,
\[
\text{Var} \left( X_{RSS} \right) = \frac{\sigma_X^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} \left( \mu_{(im)} - \mu_X \right)^2
\]
and
\[
\text{Var} \left( Y_{RSS} \right) = \frac{\sigma_Y^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} \left( \mu_{(im)} - \mu_Y \right)^2.
\]

Using the two relations
\[
\text{Cov} \left( X_{RSS}, Y_{RSS} \right) = \beta \text{Var} \left( X_{RSS} \right),
\]  \hspace{1cm} (15)

and
\[
\text{Var} \left( Y_{RSS} \right) = \beta^2 \text{Var} \left( X_{RSS} \right) + \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right),
\]  \hspace{1cm} (16)

the variance is given by
\[
\text{Var} \left( \hat{Y}_{RSS} \right) \cong \text{Var} \left( X_{RSS} \right) (T + \beta)^2 + \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right)
\]
\[
= (T + \beta)^2 \left( \frac{\sigma_X^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} \left( \mu_{(im)} - \mu_X \right)^2 \right)
\]
\[
+ \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right).
\]  \hspace{1cm} (17)

The Efficiency of \( \hat{Y}_{RSS} \) with respect to \( \hat{Y}_{SRS} \) for estimating the population mean \( \mu_Y \) is defined as:
\[
\text{eff} \left( \hat{Y}_{SRS}, \hat{Y}_{RSS} \right) = \frac{\text{Var} \left( \hat{Y}_{SRS} \right)}{\text{Var} \left( \hat{Y}_{RSS} \right)}
\]
\[
\cong \frac{\sigma_X^2 (T + \beta)^2 + \sigma_Y^2 \left( 1 - \rho^2 \right)}{(T + \beta)^2 \left( \frac{\sigma_X^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} \left( \mu_{(im)} - \mu_X \right)^2 \right) + \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right)}
\]
\[
\cong \frac{\sigma_X^2 (T + \beta)^2 + \sigma_Y^2 \left( 1 - \rho^2 \right)}{(T + \beta)^2 \left( \frac{\sigma_X^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} \left( \mu_{(im)} - \mu_X \right)^2 \right) + \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right)}.
\]  \hspace{1cm} (18)

Since \( (T + \beta)^2 \frac{\beta^2}{(\mu_X + q)^2} \) and \( \frac{1}{m} \sigma_Y^2 \left( 1 - \rho^2 \right) \) are fixed in the numerator and denominator and since \( \text{Var} \left( X_{RSS} \right) < \text{Var} \left( X_{SRS} \right) \) (Takahasi and Wakimoto, 1968), therefore \( \text{eff} \left( \hat{Y}_{SRS}, \hat{Y}_{RSS} \right) > 1 \). Implies that \( \hat{Y}_{RSS} \) is more efficient than \( \hat{Y}_{SRS} \) based on the same number of measured units.

III. SIMULATION STUDY

In this section, simulation study is conducted to investigate the performance of \( \hat{Y}_{RSS} \) and \( \hat{Y}_{SRS} \) methods for estimating the population mean where the ranking was performed on the variable \( X \). The samples were generated from bivariate normal distribution with parameters \( \mu_X = 2 \), \( \mu_Y = 4 \), \( \sigma_X^2 = \sigma_Y^2 = 1 \), and \( \rho = \pm 0.99, \pm 0.90, \pm 0.80, \pm 0.70, \pm 0.50, \pm 0.25 \).

Based on 60,000 replications, the efficiency and bias values of \( \hat{Y}_{RSS} \) and \( \hat{Y}_{SRS} \) are obtained and the results for \( m = 2, 3, 4, 5, 6 \) based on the knowledge of \( q_i \) are presented in Table 1 and the results are summarized in Table 2 when \( q_j \) is known. The efficiency of \( \hat{Y}_{RSS} \) with respect to \( \hat{Y}_{SRS} \) is obtained using Equation (18). In this section the ranking is performed on the variable \( X \), however the whole process can be repeated while the ranking is performed on the variable \( Y \).
### TABLE I

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Eff</th>
<th>( m = 2 )</th>
<th>( m = 3 )</th>
<th>( m = 4 )</th>
<th>( m = 5 )</th>
<th>( m = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>Bias RSS</td>
<td>0.089</td>
<td>0.051</td>
<td>0.033</td>
<td>0.024</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.151</td>
<td>0.098</td>
<td>0.075</td>
<td>0.059</td>
<td>0.048</td>
</tr>
<tr>
<td>0.90</td>
<td>Bias RSS</td>
<td>0.093</td>
<td>0.048</td>
<td>0.028</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.148</td>
<td>0.084</td>
<td>0.067</td>
<td>0.050</td>
<td>0.041</td>
</tr>
<tr>
<td>0.80</td>
<td>Bias RSS</td>
<td>0.123</td>
<td>0.089</td>
<td>0.063</td>
<td>0.045</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.179</td>
<td>0.147</td>
<td>0.112</td>
<td>0.094</td>
<td>0.075</td>
</tr>
<tr>
<td>0.70</td>
<td>Bias RSS</td>
<td>0.253</td>
<td>0.200</td>
<td>0.157</td>
<td>0.119</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.322</td>
<td>0.269</td>
<td>0.225</td>
<td>0.181</td>
<td>0.137</td>
</tr>
</tbody>
</table>

### TABLE II

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Eff</th>
<th>( m = 2 )</th>
<th>( m = 3 )</th>
<th>( m = 4 )</th>
<th>( m = 5 )</th>
<th>( m = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>Bias RSS</td>
<td>0.089</td>
<td>0.051</td>
<td>0.033</td>
<td>0.024</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.151</td>
<td>0.098</td>
<td>0.075</td>
<td>0.059</td>
<td>0.048</td>
</tr>
<tr>
<td>0.90</td>
<td>Bias RSS</td>
<td>0.093</td>
<td>0.048</td>
<td>0.028</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.148</td>
<td>0.084</td>
<td>0.067</td>
<td>0.050</td>
<td>0.041</td>
</tr>
<tr>
<td>0.80</td>
<td>Bias RSS</td>
<td>0.123</td>
<td>0.089</td>
<td>0.063</td>
<td>0.045</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.179</td>
<td>0.147</td>
<td>0.112</td>
<td>0.094</td>
<td>0.075</td>
</tr>
<tr>
<td>0.70</td>
<td>Bias RSS</td>
<td>0.253</td>
<td>0.200</td>
<td>0.157</td>
<td>0.119</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Bias SRS</td>
<td>0.322</td>
<td>0.269</td>
<td>0.225</td>
<td>0.181</td>
<td>0.137</td>
</tr>
</tbody>
</table>

---

World Academy of Science, Engineering and Technology  
International Journal of Mathematical and Computational Sciences  
Vol:3, No:1, 2009
Based on the results in Table 1 and 2, the following remarks can be concluded:

1). For small values of $\rho$, a gain in efficiency is obtained based on the suggested estimators using RSS with respect to SRS for all cases considered in this study.

2). The efficiency of the RSS estimators is increasing in the sample size. For example, for $\rho = 0.80$ assuming $q_1$ is known, for $m = 2,3,4,5,6$ the efficiencies are 1.429, 1.816, 2.162, 2.486 and 2.724 respectively.

3). The efficiencies obtained when $q_1$ is known are seems to be larger than those obtained when $q_3$ is known. For example for $m = 4$ and $\rho = 0.70$, the efficiency for $q_1$ is 2.044 while for $q_3$ the efficiency is 1.944.

4). When $\rho < 1$, the bias is negative while for $\rho > 1$ the bias is positive.

The efficiency of $\hat{Y}_{RSS1}$ and $\hat{Y}_{RSS3}$ is decreasing in $\rho$. For example, for $m = 3$ and $q_3$, the efficiency for $\rho = 0.99, 0.90, 0.80, 0.70, 0.50, 0.25$, respectively are 1.918, 1.826, 1.747, 1.651, 1.508 and 1.406.

IV. CONCLUSIONS

In this paper, we suggested new product-type estimators for the population mean are suggested using SRS and RSS methods providing that the first or third quartile of the auxiliary variable is known. We obtained the bias and MSE equations of the suggested estimators. The results showed that the estimators are approximately unbiased for the first degree of approximation and the RSS estimators are more efficient than SRS estimators. However, for small values of the correlation coefficient the efficiency still greater than 1.

REFERENCES