A Ratio Estimator Under General Sampling Design

Jehad Al-Jararha and Mohammed Al-Haj Ebrahem
Yarmouk University, Irbid, Jordan

Abstract: Recently, many authors introduced ratio-type estimators for estimating the mean, or the ratio, for a finite populations. Most of the articles are discussing this problem under simple random sampling design, with more assumptions on the auxiliary variable such as the coefficient of variation, and kurtosis are assumed to be known. Gupta and Shabbir (2008) have suggested an alternative form of ratio-type estimators and they assumed the coefficient of variation of the auxiliary variable must be known; this assumption is crucial for this estimator.

An estimator of the population ratio, under general sampling design, is proposed. Further, exact and an unbiased variance estimator of this estimator are obtained, and the Godambe-Joshi lower bound is asymptotically attainable for this estimator. The assumption on the coefficient of variation of the auxiliary variable is not needed for the proposed estimator. Simulation results from real data set and simulations from artificial population, show that the performance of the proposed estimator is better than Gupta and Shabbir (2008) and Hartley and Ross (1954) estimators.

Keywords: Asymptotic Results, General Sampling Design, Mean and Variance, Godambe-Joshi Lower Bound, Stratified Sampling Design.
1 Introduction

Consider a finite population $U$ of units $\{1, \ldots, N\}$. For the $i$th unit, let $y_i$ and $x_i$ be the values of the variable of interest and the auxiliary variable respectively. One of the interest is to estimate the population ratio $\theta = t_y/t_x$, where $t_y = \sum_{i \in U} y_i$, the population total for the variable of interest, and $t_x = \sum_{i \in U} x_i$, the population total for the auxiliary variable. Another interest is estimate the population total, $t_y$, by $\hat{\theta} \cdot t_x$, where $t_x$ is assumed to be known, and $\hat{\theta}$ is an estimator of $\theta$.

As it well known that Hartley and Ross (1954) estimator is an unbiased estimator under simple random sampling (srs) design without replacement for estimating the population ratio $\theta$. Under general sampling design, Al-Jararha (2008) obtained an exactly unbiased estimator for the population ratio $\theta$, this estimator gives the Hartley and Ross (1954) estimator under srs design. Further, the variance and unbiased estimator of the variance of such estimator were obtained. This estimator, also works well in stratified sampling designs.

Gupta and Shabbir (2008) showed that, under srs their estimator gives better results than the estimators given by Kadilar and Cingi (2004), Kadilar and Cingi (2006a), Kadilar and Cingi (2006b), Singh and Tailor (2003) and the regression estimator.

In this article, we will propose an estimator for the population ratio, $\theta$, under general sampling design. Through simulations from real data set and under srs design, we will compare the proposed estimator with the ratio estimators obtained by Gupta and Shabbir (2008) and Hartley and Ross (1954). Further, Hartley and Ross (1954) will be written under general sampling design and we will compare this form with the proposed estimator under proportional to size design.

Based on a measurable sampling design $p(\cdot)$, draw a random sample $s$ from $U$. An auxiliary variate $x_i$, correlated with $y_i$, is obtained for each unit in the sample $s$. Define $\pi_i$, the first order inclusion probability, by

$$\pi_i = \Pr(i \in s) = \sum_{s \ni i} p(s).$$

The Horvitz and Thompson (1952) estimator of the population total $t_y = \sum_{i \in U} y_i$ is defined by

$$\hat{t}_{y\pi} = \sum_{i \in U} y_i \frac{I_{\{i \in s\}}}{\pi_i},$$

where $I_{\{i \in s\}}$ is one if $i \in s$ and zero otherwise. It is an easy task to show that $\hat{t}_{y\pi}$ is an unbiased estimator for $t_y$. Further,

$$\bar{y}_s = \frac{1}{N} \hat{t}_{y\pi},$$

can be used to estimate the population mean $\bar{y}_U = t_y/N$. 


1.1 The Hartley and Ross Estimator

Under srs, Hartley and Ross (1954) have proposed the following estimator

$$\hat{\theta}_{HR} = \bar{r}_s + \frac{n(N - 1)}{N(n - 1)\bar{x}_U} (\bar{y}_s - \bar{r}_s \bar{x}_s)$$

(1)

to estimate the population ratio $\theta$, where

$$\bar{y}_s = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \bar{x}_s = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{r}_s = \frac{1}{n} \sum_{i=1}^{n} r_i,$$

and $r_i = \frac{y_i}{x_i}$.

This estimator can be extended to be used under general sampling design $p(\cdot)$ by redefining

$$\bar{y}_s = \frac{1}{N} \hat{t}_y, \quad \bar{x}_s = \frac{1}{N} \hat{t}_x, \quad \bar{r}_s = \frac{1}{N} \sum_{i \in U} \frac{I_{\{i \in s\}}}{\pi_i} r_i$$

in equation (1).

To find an approximate variance and an estimate for the approximate variance, by using Taylor expansion to first order, expand the righthand side of equation (1) we have

$$\hat{\theta}_{HR} \approx \text{constant} + \bar{r}_s + \frac{n(N - 1)}{N(n - 1)\bar{x}_U} \bar{y}_s - \frac{n(N - 1)}{N(n - 1)} \bar{r}_s - \frac{n(N - 1)}{N(n - 1)} \frac{\bar{r}_U}{\bar{x}_U} \bar{x}_s$$

$$= \text{constant} + \sum_{i \in U} w_i \frac{I_{\{i \in s\}}}{\pi_i},$$

(2)

where

$$w_i = \frac{n(N - 1)}{N^2(n - 1)\bar{x}_U} y_i - \frac{N - n}{N^2(n - 1)} r_i - \frac{n(N - 1)}{N^2(n - 1)} \frac{\bar{r}_U}{\bar{x}_U} x_i.$$ 

Take the variance of both sides of equation (2), we have

$$\text{var}(\hat{\theta}_{HR}) = \sum_{ij \in U} \frac{w_i w_j}{\pi_i \pi_j} \Delta_{ij}.$$ 

Therefore, an unbiased estimator for $\text{var}(\hat{\theta}_{HR})$ is

$$\hat{\text{var}}(\hat{\theta}_{HR}) = \sum_{ij \in s} \frac{w_i \hat{w}_j}{\pi_i \pi_j} \Delta_{ij},$$

where

$$\hat{w}_i = \frac{n(N - 1)}{N^2(n - 1)\bar{x}_U} y_i - \frac{N - n}{N^2(n - 1)} r_i - \frac{n(N - 1)}{N^2(n - 1)} \frac{\bar{r}_U}{\bar{x}_U} x_i, \quad \Delta_{ij} = \pi_{ij} - \pi_i \pi_j,$$

and $\pi_{ij}$ is the second order inclusion probability.
1.2 The Gupta and Shabbir Estimator

Under srs design, Gupta and Shabbir (2008) have proposed the estimator

$$\bar{y}_{GS} = \left[ w_1 \bar{y}_s + w_2 (\bar{x}_U - \bar{x}_s) \right] \left( \frac{\eta \bar{y}_U + \lambda}{\eta \bar{x}_s + \lambda} \right)$$  (3)

to estimate the population mean $\bar{y}_U$, where $w_1$ and $w_2$ are weights and $\eta \neq 0$ and $\lambda$ are either constants or functions of the known parameters such as standard deviation, variance, etc. The bias and the mean squares error (MSE), as corrected by Koyuncu and Kadilar (2010), of $\bar{y}_{GS}$ are

$$\text{bias}(\bar{y}_{GS}) = (w_1 - 1)\bar{y}_U + \gamma [w_1 \bar{y}_U (\tau^2 C_x^2 - \tau C_{yx}) + w_2 \bar{x}_U \tau C_x^2]$$

and

$$\text{MSE}(\bar{y}_{GS}) = (w_1 - 1)^2 \bar{y}_U^2 + w_1^2 \bar{y}_U^2 \gamma (C_y^2 - 4\tau C_{yx} + 3\tau^2 C_x^2) + w_2^2 \bar{x}_U^2 \gamma C_x^2$$

$$- 2w_1 \bar{y}_U \gamma (\tau^2 C_x^2 - \tau C_{yx}) - 2\bar{x}_U \bar{y}_U w_2 \tau \gamma C_x^2$$

$$- 2\bar{x}_U \bar{y}_U w_1 w_2 \gamma (C_{yx} - 2\tau C_x^2).$$  (4)

The optimum values of $w_1$ and $w_2$, which minimize the MSE, are given by

$$w_1^* = \frac{1 - \gamma \tau^2 C_x^2}{1 + \gamma C_y^2 - \gamma \rho_x^2 C_y^2 - \gamma \tau^2 C_x^2},$$

and

$$w_2^* = \frac{\bar{Y}}{\bar{X}} \left( \tau + \frac{(1 - \gamma \tau^2 C_x^2) (C_{yx} - 2\tau C_x^2)}{C_x^2 + C_y^2 C_x^2 - \gamma C_{yx}^2 - \gamma \tau^2 C_x^4} \right).$$

Therefore, the optimum MSE of $\bar{y}_{GS}$ is

$$\text{MSE}(\bar{y}_{GS})_{\text{min}} = \frac{(1 - \gamma \tau^2 C_x^2) \text{MSE}(\bar{y}_{reg})}{(1 - \gamma \tau^2 C_x^2) + \text{MSE}(\bar{y}_{reg}) / \bar{y}_U^2},$$  (5)

where $\text{MSE}(\bar{y}_{reg}) = \frac{1 - f}{n} \bar{y}_U^2 C_y^2 (1 - \rho_{yx}^2)$, $\tau = \frac{\bar{y}_U}{\bar{y}_U + \lambda}$, $C_y$ is the coefficient of variation of $y$, $\rho_{yx}$ is the correlation coefficient between $y$ and $x$, which can be estimated from the sample, $C_x$ is the coefficient of variation of $x$ is assumed to be known, $f = n/N$ and $\gamma = (N - n)/(nN)$.

Since our goal is to estimate the population ratio $\theta$, divide equation (3) by $\bar{x}_U$, we have

$$\hat{\theta}_{GS} = \left[ w_1 \frac{\bar{y}_s}{\bar{x}_U} + w_2 \left( 1 - \frac{\bar{x}_s}{\bar{x}_U} \right) \right] \left( \frac{\eta \bar{y}_U + \lambda}{\eta \bar{x}_s + \lambda} \right),$$  (6)

with

$$\text{MSE} \left( \hat{\theta}_{GS} \right)_{\text{min}} = \text{MSE} \left( \bar{y}_{GS}/\bar{x}_U \right)_{\text{min}}.$$
2 The Proposed Estimator

Assume that $x_i > 0$ for all $i = 1, \ldots, N$ and $\bar{x}_U$ is known. Under general sampling design, $p(\cdot)$, the following estimator is proposed

$$\hat{\theta}_P = \bar{r}_s + \frac{1}{\bar{x}_U}(\bar{y}_s - \bar{r}_s \bar{x}_s).$$

(7)

**Remark 2.1** $\hat{\theta}_P$ is not the Hartley and Ross (1954) estimator especially for small sample size $n$.

By using the Taylor expansion, expand $\hat{\theta}_P$ to first order, we have

$$\hat{\theta}_P \approx \bar{r}_U + \frac{1}{\bar{x}_U}[\bar{x}_U (\bar{r}_s - \bar{r}_U) + \bar{r}_U (\bar{x}_s - \bar{x}_U)]$$

$$= \bar{y}_n + \bar{r}_U (\bar{x}_U - \bar{x}_s).$$

(8)

Hence, $E_p(\hat{\theta}_P) = \bar{y}_U/\bar{x}_U = \theta$, i.e. to first order, $\hat{\theta}_P$ is an unbiased estimator for $\theta$. From equation (8) rewrite $\hat{\theta}_P$ as

$$\hat{\theta}_P = \bar{r}_U + \frac{1}{N\bar{x}_U} \sum_{i \in U} (y_i - \bar{r}_U x_i) \frac{1}{\pi_i}.\tag{9}$$

Therefore,

$$\text{var}_p(\hat{\theta}_P) = \frac{1}{N^2\bar{x}_U^2} \sum_{i \in U} \sum_{j \in U} Z_i Z_j \frac{1}{\pi_i} \frac{1}{\pi_j} \Delta_{ij},\tag{10}$$

where $Z_i = y_i - \bar{r}_U x_i$. This variance can be estimated by

$$\hat{\text{var}}_p(\hat{\theta}_P) = \frac{1}{N^2\bar{x}_U^2} \sum_{i \in s} \sum_{j \in s} \tilde{Z}_i \tilde{Z}_j \frac{1}{\pi_i} \frac{1}{\pi_j} \Delta_{ij},$$

(11)

where $\tilde{Z}_i = y_i - \bar{r}_s x_i$.

**Remark 2.2** Under mild conditions, asymptotic results for $\hat{\theta}_P$ can be established. As an example, a central limit theorem for $\hat{\theta}_P$ can be established. Under srs, it can be shown that

$$\text{avar}_{srs}(\hat{\theta}_P) = \frac{1}{\bar{x}_U n} \left(1 - \frac{n}{N}\right) \frac{1}{n-1} \sum_{i \in U} (Z_i - \bar{Z})^2$$

is the asymptotic variance of $\hat{\theta}_P$ and

$$\hat{\text{avar}}_{srs}(\hat{\theta}_P) = \frac{1}{\bar{x}_U n} \left(1 - \frac{n}{N}\right) \frac{1}{n-1} \sum_{i \in s} (\tilde{Z}_i - \bar{Z})^2$$

is a consistent estimator for $\text{avar}_{srs}(\hat{\theta}_P)$. Therefore,

$$\frac{\hat{\theta}_P - \theta}{\sqrt{\hat{\text{avar}}_{srs}(\hat{\theta}_P)}} \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{as} \quad N \to \infty.$$
Now, consider the model \( \theta_i = \beta x_i + \nu \varepsilon_i \), where \( \varepsilon_i \) are independent with mean zero and variance \( \sigma_i^2 \). Let \( \hat{\theta} \) be any estimator for the population ratio \( \theta \), the estimation error \( \hat{\theta} - \theta \) can be examined, jointly under the model, \( \xi \), and the sampling design, \( p(\cdot) \). The anticipated variance (Särndal, Swensson, and Wretman, 1992) of \( \hat{\theta} - \theta \) is

\[
E_\xi E_p \left[ (\hat{\theta} - \theta)^2 \right] = \left[ E_\xi E_p (\hat{\theta} - \theta) \right]^2.
\]

If \( E_\xi E_p (\hat{\theta} - \theta) = 0 \), the anticipated variance is

\[
E_\xi E_p \left[ (\hat{\theta} - \theta)^2 \right].
\]

The Godambe-Joshi lower bound (Godambe and Joshi, 1965) is defined by

\[
E_\xi E_p \left( \hat{\theta} - \theta \right)^2 \geq \frac{1}{N^2 \bar{x}_U^2} \sum_{i \in U} \left( \frac{1}{\pi_i} - 1 \right) \sigma_i^2.
\]

Assume that \( \pi_i \geq \pi_{ij} \geq \pi^* > 0 \), for all \( i, j \in U \). The Godambe-Joshi lower bound (GJLB) is of order \( O \left( \left( N \pi^* \right)^{-1} \right) \).

Under the model \( \xi \)

\[
Z_i = \varepsilon_i - \frac{x_i}{N} \sum_{j \in U} \varepsilon_j x_j
\]

are independent with mean zero and variance

\[
\sigma_i^2 - \frac{2\sigma_i^2}{N} + \frac{x_i^2}{N^2} \sum_{j \in U} \frac{\sigma_j^2}{x_j^2}.
\]

Hence, we can show that

\[
E_\xi E_p \left( \hat{\theta}_P - \theta \right)^2 = \text{GJLB} + \text{terms of order } O \left( \left( N^2 \pi^* \right)^{-1} \right).
\]

Therefore, the GJLB is asymptotically attainable for \( \hat{\theta}_P \).

### 3 Simulation Studies and Conclusions

Consider the real data set, USPOP: a summary of the United States population from the 2000 Census. This data is obtained from Scheaffer, Mendenhall, and Ott (2006). The percent in poverty for US was 11.9 %, as reported in the data set or as computed from the data. In this section, our main goal is to estimate this number based on different estimators.

The variables of our interest are \( X := \text{Total} \): total resident population for each state in US, and Percent in Poverty: percentage of the population estimated to live with income under the poverty line. To produce the variable \( Y := \text{number of resident with income under the poverty line} \), multiply the variable Total by the variable Percent in Poverty. Under
srs, we will compare the three estimators, namely Hartley and Ross (1954), different versions of Gupta and Shabbir (2008), and the proposed estimator which is given by equation (7). As suggested by Koyuncu and Kadilar (2010), in equation (6), consider the following choices of $\eta$ and $\lambda$:

\[
\begin{array}{|c|c|c|c|c|}
\hline
\eta & 0 & 1 & 1 & 1 \\
\lambda & 1 & \rho_{yx} & C_x & \beta_{2(x)} \\
\theta_{GS} & \theta_{GS(0)} & \theta_{GS(1)} & \theta_{GS(2)} & \theta_{GS(3)} & \theta_{GS(4)} & \theta_{GS(5)} \\
\beta_{2(x)} & C_x \\
\hline
\end{array}
\]

Here $\lambda = \beta_{2(x)}$ is the kurtosis of the auxiliary variable $X$. From the data USPOP, under srs, draw a random sample of size $n$ by using procedure `surveyselect` of SAS Institute.

Our purpose is to estimate the percent in poverty $= 11.9\%$.

Consider an artificial population of $N = 200$ units. For $i = 1, \ldots, 200$, simulate $x_i$ from exp(1) and independently from the random error, $\varepsilon_i$. For given $x_i$, define $y_i = 8x_i + \varepsilon_i$. We will simulate $\varepsilon_i$ from $N(0, x_i)$ and another case from $N(0, x_i^2)$. When $\varepsilon_i \sim N(0, x_i)$, we have $t_x = 190.7164$, $t_y = 1508.4788$, and $\theta = 7.9095$; further, when $\varepsilon_i \sim N(0, x_i^2)$, we have $t_x = 190.7164$, $t_y = 1492.4845$, and $\theta = 7.8257$.

Define the following: the empirical mean of the estimator $\hat{\theta}$ is defined by

\[
EM(\hat{\theta}) = \frac{1}{1500} \sum_{k=1}^{1500} \hat{\theta}^{(k)},
\]

where $\hat{\theta}^{(k)}$ is the estimate of $\theta$ based on the $k$th simulation. The empirical relative bias (ERB) of $\hat{\theta}$ is defined by

\[
ERB(\hat{\theta}) = \frac{EM(\hat{\theta}) - \theta}{\theta} \times 100\%.
\]

The empirical mean squares error of $\hat{\theta}$ is defined by

\[
EMSE(\hat{\theta}) = \frac{1}{1500} \sum_{k=1}^{1500} (\hat{\theta}^{(k)} - \theta)^2,
\]

and the empirical relative mean squares error (ERMSE) of the estimator $\hat{\theta}$ to the EMSE of the estimator $\hat{\theta}_P$ is defined by

\[
ERMSE(\hat{\theta}) = \frac{EMSE(\hat{\theta})}{EMSE(\hat{\theta}_P)}.
\]

From the described populations, under srs sampling design and by using procedure `surveyselect` of the SAS Institute, simulate 1500 samples when the sample size $n = 2, 5, 10, 15, 20, 25$. For a given sample size $n$, and based on each sample, estimate $\theta$ by using $\hat{\theta}_{HR}$, $\hat{\theta}_{GS(i)}$, $i = 0, \ldots, 5$, and $\hat{\theta}_P$. Further, compute EM, ERB, and ERMSE as defined by equations (12), (13), and (15), respectively. Results are given in Tables 1, 2, and 3.

It is not an easy task to extend $\hat{\theta}_{GS}$ to be used under general sampling design. However, the proposed estimator $\hat{\theta}_P$ can be used under a general sampling design. Further, the
Table 1: US population. Comparison between $\hat{\theta}_P$, $\hat{\theta}_{HR}$ and $\hat{\theta}_{GS(i)}$, $i = 0, \ldots, 5$, under srs and based on 1500 simulations.

<table>
<thead>
<tr>
<th>n</th>
<th>$\theta_P$</th>
<th>$\theta_{HR}$</th>
<th>$\theta_{GS(0)}$</th>
<th>$\theta_{GS(1)}$</th>
<th>$\theta_{GS(2)}$</th>
<th>$\theta_{GS(3)}$</th>
<th>$\theta_{GS(4)}$</th>
<th>$\theta_{GS(5)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.1168</td>
<td>0.1191</td>
<td>0.0341</td>
<td>0.1140</td>
<td>0.1140</td>
<td>0.1140</td>
<td>0.1140</td>
<td>0.1140</td>
</tr>
<tr>
<td></td>
<td>-1.84</td>
<td>0.07</td>
<td>-71.39</td>
<td>-4.20</td>
<td>-4.20</td>
<td>-4.22</td>
<td>-4.20</td>
<td>-4.22</td>
</tr>
<tr>
<td></td>
<td>2.04</td>
<td>93.06</td>
<td>68.99</td>
<td>68.99</td>
<td>68.99</td>
<td>68.99</td>
<td>68.99</td>
<td>68.99</td>
</tr>
<tr>
<td>5</td>
<td>0.1179</td>
<td>0.1187</td>
<td>0.0739</td>
<td>0.1044</td>
<td>0.1043</td>
<td>0.1044</td>
<td>0.1044</td>
<td>0.1044</td>
</tr>
<tr>
<td></td>
<td>-0.99</td>
<td>-0.28</td>
<td>-37.94</td>
<td>-12.31</td>
<td>-12.31</td>
<td>-12.31</td>
<td>-12.31</td>
<td>-12.31</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>298.82</td>
<td>110.21</td>
<td>110.21</td>
<td>110.24</td>
<td>110.21</td>
<td>110.24</td>
<td>110.24</td>
</tr>
<tr>
<td>10</td>
<td>0.1182</td>
<td>0.1185</td>
<td>0.0980</td>
<td>0.1098</td>
<td>0.1098</td>
<td>0.1098</td>
<td>0.1098</td>
<td>0.1098</td>
</tr>
<tr>
<td></td>
<td>-0.72</td>
<td>-0.45</td>
<td>-17.71</td>
<td>-7.75</td>
<td>-7.75</td>
<td>-7.75</td>
<td>-7.75</td>
<td>-7.75</td>
</tr>
<tr>
<td></td>
<td>1.07</td>
<td>11.14</td>
<td>4.59</td>
<td>4.59</td>
<td>4.59</td>
<td>4.59</td>
<td>4.59</td>
<td>4.59</td>
</tr>
<tr>
<td>15</td>
<td>0.1187</td>
<td>0.1189</td>
<td>0.1071</td>
<td>0.1135</td>
<td>0.1135</td>
<td>0.1135</td>
<td>0.1135</td>
<td>0.1135</td>
</tr>
<tr>
<td></td>
<td>-0.31</td>
<td>-0.14</td>
<td>-10.02</td>
<td>-4.61</td>
<td>-4.61</td>
<td>-4.61</td>
<td>-4.61</td>
<td>-4.61</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>6.90</td>
<td>4.06</td>
<td>4.06</td>
<td>4.06</td>
<td>4.06</td>
<td>4.06</td>
<td>4.06</td>
</tr>
<tr>
<td>20</td>
<td>0.1186</td>
<td>0.1187</td>
<td>0.1105</td>
<td>0.1148</td>
<td>0.1148</td>
<td>0.1148</td>
<td>0.1148</td>
<td>0.1148</td>
</tr>
<tr>
<td></td>
<td>-0.35</td>
<td>-0.24</td>
<td>-7.16</td>
<td>-3.55</td>
<td>-3.55</td>
<td>-3.55</td>
<td>-3.55</td>
<td>-3.55</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>4.83</td>
<td>3.21</td>
<td>3.21</td>
<td>3.21</td>
<td>3.21</td>
<td>3.21</td>
<td>3.21</td>
</tr>
<tr>
<td>25</td>
<td>0.1188</td>
<td>0.1189</td>
<td>0.1136</td>
<td>0.1163</td>
<td>0.1163</td>
<td>0.1163</td>
<td>0.1163</td>
<td>0.1163</td>
</tr>
<tr>
<td></td>
<td>-0.18</td>
<td>-0.10</td>
<td>-4.54</td>
<td>-2.27</td>
<td>-2.27</td>
<td>-2.27</td>
<td>-2.27</td>
<td>-2.27</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>4.26</td>
<td>2.91</td>
<td>2.91</td>
<td>2.91</td>
<td>2.91</td>
<td>2.91</td>
<td>2.91</td>
</tr>
</tbody>
</table>

Table 2: Artificial population. Comparison between $\hat{\theta}_P$, $\hat{\theta}_{HR}$, and $\hat{\theta}_{GS(i)}$, $i = 0, \ldots, 5$, under srs and based on 1500 simulations when $\varepsilon_i \sim N(0, x_i)$.

<table>
<thead>
<tr>
<th>n</th>
<th>$\theta_P$</th>
<th>$\theta_{HR}$</th>
<th>$\theta_{GS(0)}$</th>
<th>$\theta_{GS(1)}$</th>
<th>$\theta_{GS(2)}$</th>
<th>$\theta_{GS(3)}$</th>
<th>$\theta_{GS(4)}$</th>
<th>$\theta_{GS(5)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.31</td>
<td>-0.42</td>
<td>-62.08</td>
<td>-34.80</td>
<td>-34.29</td>
<td>-42.06</td>
<td>-4.63</td>
<td>-41.72</td>
</tr>
<tr>
<td></td>
<td>1.98</td>
<td>381.39</td>
<td>196.87</td>
<td>181.37</td>
<td>168.99</td>
<td>271.92</td>
<td>166.56</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.93</td>
<td>7.92</td>
<td>6.56</td>
<td>6.53</td>
<td>6.09</td>
<td>6.92</td>
<td>7.75</td>
<td>6.93</td>
</tr>
<tr>
<td></td>
<td>0.12</td>
<td>0.07</td>
<td>-17.07</td>
<td>-17.42</td>
<td>-22.98</td>
<td>-12.52</td>
<td>-2.05</td>
<td>-12.41</td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>249.04</td>
<td>3972.53</td>
<td>10308.0</td>
<td>239.20</td>
<td>52.56</td>
<td>241.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.04</td>
<td>-1.41</td>
<td>-7.39</td>
<td>-7.74</td>
<td>5.45</td>
<td>-1.96</td>
<td>5.97</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>2280.8</td>
<td>1158.4</td>
<td>1304.9</td>
<td>7514.3</td>
<td>90.49</td>
<td>8194.3</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>7.9026</td>
<td>7.9007</td>
<td>7.4872</td>
<td>7.7206</td>
<td>7.7160</td>
<td>7.5939</td>
<td>7.8457</td>
<td>7.5965</td>
</tr>
<tr>
<td></td>
<td>-0.09</td>
<td>-0.11</td>
<td>-5.34</td>
<td>-2.39</td>
<td>-2.45</td>
<td>-3.99</td>
<td>-0.81</td>
<td>-3.96</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>13.08</td>
<td>8.38</td>
<td>8.44</td>
<td>10.34</td>
<td>7.46</td>
<td>10.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>-0.05</td>
<td>-3.49</td>
<td>-1.40</td>
<td>-1.44</td>
<td>-2.55</td>
<td>-0.26</td>
<td>-2.53</td>
</tr>
<tr>
<td></td>
<td>1.01</td>
<td>8.92</td>
<td>6.18</td>
<td>6.22</td>
<td>7.35</td>
<td>5.70</td>
<td>7.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.10</td>
<td>-0.11</td>
<td>-2.54</td>
<td>-1.02</td>
<td>-1.05</td>
<td>-1.87</td>
<td>-0.17</td>
<td>-1.85</td>
</tr>
<tr>
<td></td>
<td>1.01</td>
<td>7.55</td>
<td>5.57</td>
<td>5.59</td>
<td>6.41</td>
<td>5.29</td>
<td>6.38</td>
<td></td>
</tr>
</tbody>
</table>
For all values of \( n \), \( \hat{\theta}_P \) has lowest empirical relative mean squares error (ERMSE) compared with other estimators. Further, ERMSE(\( \hat{\theta}_P \)) and ERMSE(\( \hat{\theta}_{HR} \)) are approximately the same for large sample size \( n \).

### 3.1 Results and Conclusions

From Tables 1, 2, and 3, we can conclude the following:

- The proposed estimator \( \hat{\theta}_P \) has a negligible relative bias, especially for small values of \( n \) and approaches zero with increasing \( n \).
- For all values of \( n \), \( \hat{\theta}_P \) has lowest empirical relative mean squares error (ERMSE) compared with other estimators. Further, ERMSE(\( \hat{\theta}_P \)) and ERMSE(\( \hat{\theta}_{HR} \)) are approximately the same for large sample size \( n \).
Table 4: US population. Comparison between \( \hat{\theta}_P \) and \( \hat{\theta}_{HR} \) under \( \pi ps \) sampling design and based on 1500 simulations.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.1191</td>
<td>0.1279</td>
<td>0.1193</td>
<td>0.1215</td>
<td>0.1188</td>
<td>0.1197</td>
<td>0.1186</td>
<td>0.1191</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>7.46</td>
<td>0.22</td>
<td>2.04</td>
<td>0.23</td>
<td>0.57</td>
<td>-0.35</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>26.08</td>
<td>3.84</td>
<td>2.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Artificial population. Comparison between \( \hat{\theta}_P \) and \( \hat{\theta}_{HR} \) under \( \pi ps \) sampling design and based on 1500 simulations when \( \varepsilon_i \sim N(0, x_i) \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7.9014</td>
<td>7.7544</td>
<td>7.9192</td>
<td>7.9600</td>
<td>7.9151</td>
<td>7.8977</td>
<td>7.9273</td>
<td>7.9214</td>
</tr>
<tr>
<td>4</td>
<td>-0.10</td>
<td>-1.96</td>
<td>0.12</td>
<td>0.64</td>
<td>0.07</td>
<td>-0.15</td>
<td>0.22</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>210.94</td>
<td>21.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Artificial population. Comparison between \( \hat{\theta}_P \) and \( \hat{\theta}_{HR} \) under \( \pi ps \) sampling design and based on 1500 simulations when \( \varepsilon_i \sim N(0, x_i^2) \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
<th>( \hat{\theta}_P )</th>
<th>( \hat{\theta}_{HR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.08</td>
<td>-2.06</td>
<td>0.13</td>
<td>0.26</td>
<td>0.12</td>
<td>-0.40</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>6</td>
<td>181.16</td>
<td>20.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The assumption that the coefficient of variation for the auxiliary variable \( C_x \) plus other conditions are crucial for \( \hat{\theta}_{GS} \) and can give worst results if \( C_x \) is estimated from samples especially for small values of \( n \). \( C_x \) is computed from the population in our calculations.

From Tables 4, 5, and 6 we notice that the two estimators have a negligible relative bias. However, the proposed estimator \( \hat{\theta}_P \) do much better than the \( \hat{\theta}_{HR} \) estimator in term of ERMSE for \( n = 2, 4, 6, 8 \).

The Natural Resources Inventory (NRI) is a real survey conducted by the US Department of Agriculture’s Natural Resources Conservation Service (NRCS), in cooperation with Iowa State University’s Center for Survey Statistics and Methodology. The sample design is based on a stratified two stage area sample of all US lands (http://www.nrcs.usda.gov/). In stratified sampling design, usually we are drawing a small sample size (NRI as an example). In such situations, one can apply \( \hat{\theta}_P \) to each strata since the estimator \( \hat{\theta}_P \) has negligible relative bias and has the smallest empirical relative mean squares error among all other estimators discussed in this paper.

From the above discussions, we can conclude that the estimator \( \hat{\theta}_P \) can be used under general sampling design and has the smallest empirical relative mean squares error among all other estimators discussed in this paper especially when the sample size is small. Since
\( \hat{\theta}_P \) has negligible bias and to avoid accumulation of bias from strata to strata, the estimator \( \hat{\theta}_P \) can be used in stratified sampling design, by applying \( \hat{\theta}_P \) to each strata.

**Acknowledgements**

The authors would like to thank the referee for the valuable comments and suggestions that improved the content and the style of this paper.

**References**


Authors’ address:

Jehad Al-Jararha and Mohammed Al-Haj Ebrahem

Department of Statistics

Yarmouk University

Irbid, Jordan

E-Mails: jjararha@gmail.com, m_hassanb@hotmail.com