AUSTRIAN JOURNAL OF STATISTICS
Volume 41 (2012), Number 2, 105–115

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.

Zusammenfassung: Jüngst führten einige Autoren quotientenartige Schätzer ein, um den Erwartungswert oder einen Quotienten davon für eine endliche Stichprobe zu schätzen. In den meisten dieser Artikeln wird dieses Problem unter einem einfachen Zufallsstichproben Design diskutiert mit weiteren Annahmen bezüglich der Hilfsvariablen wie die Bekanntheit deren Variationskoeffizienten und Kurtosis. Gupta and Shabbir (2008) schlugen eine alternative Form von quotientenartige Schätzer vor, und sie nahmen dazu an, dass Variationskoeffizient der Hilfsvariablen bekannt sei; derarige Annahmen sind für diesen Schätzer is kritisch.

Ein Schätzer des Populationsverhältnisses unter einem allgemeinen Stichprobendesign wird vorgeschlagen. Weiters werden ein exakter und ein unverzerrter Varianzschätzer dieses Schätzers erhalten, und die Godambe-Joshi untere Schranke ist asymptotisch erreichbar für diesen Schätzer. Die Annahme bezüglich des Variationskoeffizienten der Hilfsvariablen wird für den vorgeschlagenen Schätzer nicht benötigt. Simulationsergebnisse von realen Datensätzen und Simulationen von künstlich generierten Daten zeigen, dass die Eigenschaften des vorgeschlagenen Schätzers besser sind als die der Gupta and Shabbir (2008) und der Hartley and Ross (1954) Schätzer.

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 ith 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 θ.

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 θ. Under general sampling design, Al-Jararha (2008) obtained an exactly unbiased estimator for the population ratio θ, 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, θ, 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 π_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

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

to estimate the population ratio θ, where

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

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

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

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

$$\begin{aligned}\hat{\theta}_{HR} &\cong \text{constant} + \bar{r}_s + \frac{n(N-1)}{N(n-1)\bar{x}_U}\bar{y}_s - \frac{n(N-1)}{N(n-1)}\bar{r}_s - \frac{n(N-1)}{N(n-1)}\frac{\bar{r}_U}{\bar{x}_U}\bar{x}_s \\ &= \text{constant} + \sum_{i\in U} w_i\frac{I_{\{i\in s\}}}{\pi_i}\,, \end{aligned} \tag{2}$$

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

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

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

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

where

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

and π_{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

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

to estimate the population mean $\bar{y}_U$, where w_1 and w_2 are weights and $\eta \neq 0$ and λ 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

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

and

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

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

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

and

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

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

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

where $\text{MSE}(\bar{y}_{reg}) = \frac{1-f}{n}\bar{y}_U^2C_y^2(1-\rho_{yx}^2)$, $\tau = \frac{\eta\bar{x}_U}{\eta\bar{x}_U+\lambda}$, C_y is the coefficient of variation of y, ρ_{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 θ, divide equation (3) by $\bar{x}_U$, we have

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

with

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

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

$$\hat{\theta}_P = \bar{r}_s + \frac{1}{\bar{x}_U}(\bar{y}_s - \bar{r}_s \bar{x}_s)\,. \tag{7}$$

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

$$\begin{aligned}\hat{\theta}_P &\cong \bar{r}_s + \frac{\bar{y}_s}{\bar{x}_U} - \frac{1}{\bar{x}_U}[\bar{r}_U \bar{x}_U + \bar{x}_U(\bar{r}_s - \bar{r}_U) + \bar{r}_U(\bar{x}_s - \bar{x}_U)] \\ &= \frac{\bar{y}_s}{\bar{x}_U} + \frac{\bar{r}_U}{\bar{x}_U}(\bar{x}_U - \bar{x}_s)\,. \end{aligned} \tag{8}$$

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

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

Therefore,

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

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

$$\widehat{\mathrm{var}}_p(\hat{\theta}_P) = \frac{1}{N^2\bar{x}_U^2}\sum_{i\in s}\sum_{j\in s}\frac{\hat{Z}_i}{\pi_i}\frac{\hat{Z}_j}{\pi_j}\frac{\Delta_{ij}}{\pi_{ij}}\,, \tag{11}$$

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

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*

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

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

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

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

$$\frac{\hat{\theta}_P - \theta}{\sqrt{\widehat{avar}_{srs}(\hat{\theta}_P)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \qquad as \quad N \to \infty\,.$$

Now, consider the model $\xi\ :\ y_i\ =\ \beta x_i +\ varepsilon_i$, where ε_i are independent with mean zero and variance σ_i^2. Let $\hat{\theta}$ be any estimator for the population ratio θ, the estimation error $\hat{\theta} - \theta$ can be examined, jointly under the model, ξ, and the sampling design, $p(\cdot)$. The anticipated variance (Särndal, Swensson, and Wretman, 1992) of $\hat{\theta} - \theta$ is

$$\mathrm{E}_\xi \mathrm{E}_p \left[\left(\hat{\theta} - \theta \right)^2 \right] - \left[\mathrm{E}_\xi \mathrm{E}_p \left(\hat{\theta} - \theta \right) \right]^2 .$$

If $\mathrm{E}_\xi \mathrm{E}_p (\hat{\theta} - \theta) = 0$, the anticipated variance is

$$\mathrm{E}_\xi \mathrm{E}_p \left[\left(\hat{\theta} - \theta \right)^2 \right] .$$

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

$$\mathrm{E}_\xi \mathrm{E}_p \left(\hat{\theta} - \theta \right)^2 \geq \frac{1}{N^2 \bar{x}_U^2} \sum_{i \in U} \left(\frac{1}{\pi_i} - 1 \right) \sigma_i^2 .$$

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

Under the model ξ

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

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

$$\mathrm{E}_\xi \mathrm{E}_p \left(\hat{\theta}_P - \theta \right)^2 = \text{GJLB} + \text{terms of order } O\left(\left(N^2 \pi^* \right)^{-1} \right) .$$

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 := 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 := 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 η and λ:

η	0	1	1	1	$\beta_{2(x)}$	C_x
λ	1	ρ_{yx}	C_x	$\beta_{2(x)}$	C_x	$\beta_{2(x)}$
$\hat{\theta}_{GS}$	$\hat{\theta}_{GS(0)}$	$\hat{\theta}_{GS(1)}$	$\hat{\theta}_{GS(2)}$	$\hat{\theta}_{GS(3)}$	$\hat{\theta}_{GS(4)}$	$\hat{\theta}_{GS(5)}$

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 $\theta = 11.9\%$.

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, ε_i. For given x_i, define $y_i = 8x_i + \varepsilon_i$. We will simulate ε_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

$$\text{EM}(\hat{\theta}) = \frac{1}{1500} \sum_{k=1}^{1500} \hat{\theta}^{(k)} , \tag{12}$$

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

$$\text{ERB}(\hat{\theta}) = \frac{\text{EM}(\hat{\theta}) - \theta}{\theta} \times 100\% . \tag{13}$$

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

$$\text{EMSE}(\hat{\theta}) = \frac{1}{1500} \sum_{k=1}^{1500} \left(\hat{\theta}^{(k)} - \theta\right)^2 , \tag{14}$$

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

$$\text{ERMSE}(\hat{\theta}) = \frac{\text{EMSE}(\hat{\theta})}{\text{EMSE}(\hat{\theta}_P)} . \tag{15}$$

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 θ 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 $\hat{\theta}_{GS(i)}$, $i = 0, \dots, 5$, under srs and based on 1500 simulations.

		$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_{GS(0)}$	$\hat{\theta}_{GS(1)}$	$\hat{\theta}_{GS(2)}$	$\hat{\theta}_{GS(3)}$	$\hat{\theta}_{GS(4)}$	$\hat{\theta}_{GS(5)}$
$n = 2$	EM	0.1168	0.1191	0.0341	0.1140	0.1140	0.1140	0.1140	0.1140
	ERB	−1.84	0.07	−71.39	−4.20	−4.20	−4.22	−4.20	−4.22
	ERMSE		2.04	93.06	68.99	68.99	68.98	68.99	68.98
$n = 5$	EM	0.1179	0.1187	0.0739	0.1044	0.1043	0.1044	0.1044	0.1044
	ERB	−0.99	−0.28	−37.94	−12.31	−12.31	−12.31	−12.31	−12.31
	ERMSE		1.19	298.82	110.21	110.21	110.24	110.21	110.24
$n = 10$	EM	0.1182	0.1185	0.0980	0.1098	0.1098	0.1098	0.1098	0.1098
	ERB	-0.72	−0.45	−17.71	−7.75	−7.75	−7.75	−7.75	−7.75
	ERMSE		1.07	11.14	4.59	4.59	4.59	4.59	4.59
$n = 15$	EM	0.1187	0.1189	0.1071	0.1135	0.1135	0.1135	0.1135	0.1135
	ERB	−0.31	−0.14	−10.02	−4.61	−4.61	−4.61	−4.61	−4.61
	ERMSE		1.04	6.90	4.06	4.06	4.06	4.06	4.06
$n = 20$	EM	0.1186	0.1187	0.1105	0.1148	0.1148	0.1148	0.1148	0.1148
	ERB	−0.35	−0.24	−7.16	−3.55	−3.55	−3.55	−3.55	−3.55
	ERMSE		1.02	4.83	3.21	3.21	3.21	3.21	3.21
$n = 25$	EM	0.1188	0.1189	0.1136	0.1163	0.1163	0.1163	0.1163	0.1163
	ERB	−0.18	−0.10	−4.54	−2.27	−2.27	−2.27	−2.27	−2.27
	ERMSE		1.02	4.26	2.91	2.91	2.91	2.91	2.91

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

		$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_{GS(0)}$	$\hat{\theta}_{GS(1)}$	$\hat{\theta}_{GS(2)}$	$\hat{\theta}_{GS(3)}$	$\hat{\theta}_{GS(4)}$	$\hat{\theta}_{GS(5)}$
$n = 2$	EM	7.8848	7.8765	2.9993	5.1570	5.1973	4.5827	7.5436	4.6098
	ERB	−0.31	−0.42	−62.08	−34.80	−34.29	−42.06	−4.63	−41.72
	ERMSE		1.98	381.39	196.87	181.37	168.99	271.92	166.56
$n = 5$	EM	7.93	7.92	6.56	6.53	6.09	6.92	7.75	6.93
	ERB	0.12	0.07	−17.07	−17.42	−22.98	−12.52	−2.05	−12.41
	ERMSE		1.15	249.04	3972.53	10308.0	239.20	52.56	241.99
$n = 10$	EM	7.9174	7.9125	7.7984	7.3250	7.2977	8.3402	7.7542	8.3820
	ERB	0.10	0.04	−1.41	−7.39	−7.74	5.45	−1.96	5.97
	ERMSE		1.04	2280.8	1158.4	1304.9	7514.3	90.49	8194.3
$n = 15$	EM	7.9026	7.9007	7.4872	7.7206	7.7160	7.5939	7.8457	7.5965
	ERB	−0.09	−0.11	−5.34	−2.39	−2.45	−3.99	−0.81	−3.96
	ERMSE		1.04	13.08	8.38	8.44	10.34	7.46	10.28
$n = 20$	EM	7.9083	7.9059	7.6335	7.7987	7.7954	7.7079	7.8891	7.7097
	ERB	−0.02	−0.05	−3.49	−1.40	−1.44	−2.55	−0.26	−2.53
	ERMSE		1.01	8.92	6.18	6.22	7.35	5.70	7.31
$n = 25$	EM	7.9018	7.9008	7.7083	7.8287	7.8263	7.7620	7.8960	7.7633
	ERB	−0.10	−0.11	−2.54	−1.02	−1.05	−1.87	−0.17	−1.85
	ERMSE		1.01	7.55	5.57	5.59	6.41	5.29	6.38

Table 3: 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 $\varepsilon_i \sim N(0, x_i^2)$.

		$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_{GS(0)}$	$\hat{\theta}_{GS(1)}$	$\hat{\theta}_{GS(2)}$	$\hat{\theta}_{GS(3)}$	$\hat{\theta}_{GS(4)}$	$\hat{\theta}_{GS(5)}$
$n = 2$	EM	7.8664	7.7853	2.9941	5.2532	5.2686	4.5617	8.2338	4.5886
	ERB	0.52	−0.52	−61.74	−32.87	−32.68	−41.71	5.21	−41.37
	ERMSE		2.05	757.36	309.19	295.23	328.02	1242.0	323.08
$n = 5$	EM	7.8544	7.83	6.3771	8.5598	8.3352	6.6944	7.4650	6.6986
	ERB	0.37	0.01	−18.51	9.38	6.51	−14.46	−4.61	−14.40
	ERMSE		1.25	449.10	3180.1	2183.1	484.70	58.25	495.35
$n = 10$	EM	7.8352	7.8212	−0.9800	7.3812	7.3679	6.4771	7.6864	6.5205
	ERB	0.12	−0.06	−112.52	−5.68	−5.85	−17.23	−1.78	−16.68
	ERMSE		1.11	511547	282.03	300.80	6304.7	54.19	5767.6
$n = 15$	EM	7.8172	7.8082	7.3894	7.6211	7.6161	7.4952	7.7443	7.4977
	ERB	−0.11	−0.22	−5.58	−2.61	−2.68	−4.22	−1.04	−4.19
	ERMSE		1.07	8.72	5.55	5.60	6.87	4.93	6.84
$n = 20$	EM	7.8239	7.8174	7.5405	7.7039	7.7003	7.6139	7.7926	7.6157
	ERB	−0.02	−0.11	−3.65	−1.56	−1.60	−2.71	−0.42	−2.68
	ERMSE		1.05	5.90	4.22	4.24	4.94	3.92	4.92
$n = 25$	EM	7.8193	7.8143	7.6166	7.7355	7.7328	7.6695	7.8013	7.6708
	ERB	−0.08	−0.14	−2.67	−1.15	−1.19	−2.00	−0.31	−1.98
	ERMSE		1.04	4.90	3.79	3.81	4.26	3.66	4.24

estimator $\hat{\theta}_{HR}$ can be used under general sampling and this can be done by using equation (1) with suggested extensions. Therefore, we will compare the two estimator $\hat{\theta}_P$ and $\hat{\theta}_{HR}$ under proportional to size and without replacement (πps) sampling design.

For the USPOP population, consider the variable $X :=$ Total as the size variable. Under πps, draw a random sample of size $n = 2, 4, 6, 8$ by using procedure `surveyselect` of the SAS Institute. With the same number of simulations (i.e. 1500) and from each simulation, estimate $\theta = 11.9\%$ by $\hat{\theta}_{HR}$ and by $\hat{\theta}_P$. Based on 1500 simulations, compute EM, ERB, and ERMSE. Due to the sampling limitation (the relative size of each sampling unit should not exceed 1/n), we can not take n greater than 8. Further, repeat the same ideas for the artificial population when X is the size variable. The results are summarized in Tables 4, 5, and 6.

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 πps sampling design and based on 1500 simulations.

	$n=2$		$n=4$		$n=6$		$n=8$	
	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$
EM	0.1191	0.1279	0.1193	0.1215	0.1188	0.1197	0.1186	0.1191
ERB	0.04	7.46	0.22	2.04	0.23	0.57	−0.35	0.02
ERMSE		26.08		3.84		2.09		4.74

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

	$n=2$		$n=4$		$n=6$		$n=8$	
	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$
EM	7.9014	7.7544	7.9192	7.9600	7.9151	7.8977	7.9273	7.9214
ERB	−0.10	−1.96	0.12	0.64	0.07	−0.15	0.22	0.15
ERMSE		210.94		21.40		8.88		4.45

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

	$n=2$		$n=4$		$n=6$		$n=8$	
	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$	$\hat{\theta}_P$	$\hat{\theta}_{HR}$
EM	7.832	7.6643	7.8360	7.8463	7.8349	7.7943	7.8396	7.8277
ERB	0.08	−2.06	0.13	0.26	0.12	−0.40	0.18	0.03
ERMSE		181.16		20.64		9.12		4.78

- 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

Al-Jararha, J. (2008). *Unbiased Ratio Estimation For Finite Populations*. Unpublished doctoral dissertation, Colorado State University, Fort Collins, Co.

Godambe, V. P., and Joshi, V. M. (1965). Admissibility and Bayes etimation in sampling finite population. *Annals of Mathematical Statistics*, *36*, 1707-1722.

Gupta, S., and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics*, *35*, 559-566.

Hartley, H. O., and Ross, A. (1954). Unbiased ratio estimates. *Nature*, *174*, 270-271.

Horvitz, D. G., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663-685.

Kadilar, C., and Cingi, H. (2004). Ratio estimator in simple random sampling. *Applied Mathematics and Computation*, *151*, 893-902.

Kadilar, C., and Cingi, H. (2006a). Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters*, *19*, 75-79.

Kadilar, C., and Cingi, H. (2006b). New ratio estimators using correlation coefficient. *Interstat*, *4*, 1-11.

Koyuncu, N., and Kadilar, C. (2010). On improvement in estimating population mean in stratified random sampling. *Journal of Applied Statistics*, *37*, 999-1013.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Scheaffer, R. L., Mendenhall, W., and Ott, R. L. (2006). *Elementary Survey Sampling* (6th ed.). Belmont, CA: Duxbury.

Singh, H. P., and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, *6*, 555-560.

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`