Bulletin of Business and Economics, 12(3), 573-579
https://bbejournal.com
https://doi.org/10.61506/01.00071

Comparative Analysis of Variance Estimation Methods in Two-Phase Sampling: A Focus on Regression-cum-Exponential Estimators with Multiple Auxiliaries

Amber Asghar[1], Aamir Sanaullah[2], Hina Khan[3], Muhammad Hanif[4]

Abstract
In this study, we introduce a regression-cum-exponential estimator designed for estimating population variance. Specifically, we focus on the estimation of unknown population variance in a two-phase sampling setup, considering the use of multiple auxiliary variables. We derive and discuss various cases pertaining to this estimation framework. Additionally, we compare the asymptotic properties of existing approaches with those of our proposed estimator. This allows us to assess the performance and efficiency of the different methods. Finally, we conduct a simulation study to evaluate the performance of our proposed estimator in finite samples, specifically utilizing multi-auxiliary variables. This empirical analysis provides insights into the practical effectiveness of the estimators.
Keywords: regression-cum-exponential estimator, two-phase sampling, multi-auxiliary variables, variance estimation, simulation study

1. Introduction
Survey sampling has gained predominate significance across various sectors, including academia, healthcare, and public and private industries. The execution of surveys cover both probability and non-probability sampling methods, with public surveys continuance a broad range of activities such as agriculture, industry, and healthcare. Undoubtedly, survey sampling has change into an vital tool for data collection in various feature of life.
As the prevalence of survey sampling continues to rise, the need for enhanced methods to interpret results becomes increasingly imperative. Variance estimation appears as a pivotal approach in navigating the intricacies of survey designs, facilitating the extraction of certain conclusions from the amassed data. The primary goal of survey statisticians is to achieve statistical efficiency by leveraging auxiliary information, a facet that consistently proves valuable in enhancing the efficacy of estimators.
In scenarios where auxiliary information remains unknown for the entire population, the two-phase sampling scheme is exploited. Neyman (1938) pioneered this concept, wherein a large sample is initially drawn from the population, with only the auxiliary information being observed. Subsequently, a smaller sample is taken from this initial selection to observe the study variable. For example, Suppose we want to estimate the average income of households in a city, and we don't have any auxiliary information available. We decide to use a two-phase sampling approach. In the first phase, we randomly select neighborhoods across the city and in the second phase, we randomly select households within the neighborhoods previously chosen. For each selected household, we collect the income information.
Building on this foundation, Sen (1972) introduced ratio estimators that leverage multi-auxiliary information under double sampling. Isaki (1983) proposed ratio and regression estimators employing a single auxiliary variable for population variance estimation. Prasad and Singh (1990) later enhanced Isaki's (1983) estimator. Subsequently, Arcos et al. (2005) proposed more efficient estimators compared to Isaki's (2008) original estimator.
Gupta and Shabbir (2008) contributed to the field by introducing hybrid estimators for both population mean and variance. Their f3indings revealed that this hybrid class of variance estimators exhibits less bias and greater efficiency compared to Isaki's (2008) and Kadilar and Cingi's (2006) estimators.
Singh et al. (2009) expanded the methodology by developing exponential ratio and product estimators for population variance in both single and two-phase sampling, contributing further to the evolution of variance estimation techniques. John et al. (2014) introduced a regression estimator that employ multiple auxiliary variables or attributes, along with mixture regression estimators, in the context of two-phase sampling for population mean estimation, considering both partial and no information cases. Additionally, Singh et al. (2015) developed a chain regression-type estimator for estimating population mean through successive sampling on two occasions, incorporating multi-auxiliary variables. Ahmad et al. (2022) put forth an enhanced variance estimator. This novel approach leverages dual auxiliary information to enhance the robustness and generalization of the proposed estimator.
Numerous researchers have contributed significantly to the exploration of regression, ratio, and exponential estimators employing both single and multiple auxiliary variables for the estimation of population variance. Among them, Olkin (1958), Cebrian and Garcia (2018), Upadhyaya and Singh (1983), Shabbir and Gupta (2015), Asghar et al. (2014), Dubey and Sharma (2008), Abid et al. (2020), Zaman et al. (2021), and Niaz et al. (2021) have all delved into the development of ratio estimators dedicated to the task of estimating population variance. In recent studies, Masood and Shabbir (2016), Singh and Pal (2017), and Adichwal and Singh (2018) have extended this exploration. Furthermore, Cekim and Kadilar (2020) proposed ln-type regression estimators specifically designed for variance estimation under simple random sampling (SRS).
Let $\bar{X}_1, \bar{X}_2, ..., \bar{X}_k$ and $S^2_{x1}, S^2_{x2}, ..., S^2_{xk}$ be the population means and variances of k auxiliary variables where, $k = 1, 2, ..., l$ and the population variance of the study variable is S^2_y.
Additionally, it is assumed that S_{Y2} denotes the sample variance of the study variable of size n_2 selected at the second phase.

[1] Virtual University of Pakistan Lahore, Pakistan, zukhruf10@gmail.com
[2] COMSATS University Islamabad Lahore Campus, Pakistan, chaamirsanaullah@yahoo.com
[3] Corresponding Author, Government College University, Lahore, Pakistan, hinakhan@gcu.edu.pk
[4] NCBA&E, Lahore, Pakistan, drmianhanif@gmail.com

An unbiased estimator of population variance using simple random sampling is given by,

$$t_0 = s_y^2 \ , \tag{1}$$

The variance of the unbiased estimator is given by,

$$Var(t_0) = \gamma S_y^4 \left[\beta_2(\mathrm{y}) - 1 \right]. \tag{2}$$

where,

$\gamma = \dfrac{1}{n}$ and $\beta_2(\mathrm{y}) = \text{Kurtosis of study variable}$.

For the no information case, Abu-Dayyeh & Ahmend (2005) and Sanaullah et al. (2016) proposed the ratio, regression and the exponential estimators for the variance estimation using two and several auxiliary variables. Our motivation is to develop a univariate regression-cum-exponential estimator using multi-auxiliary variables for the no information case under two-phase sampling. As we are dealing with the univariate estimators using multi-auxiliary variables, therefore we adopt the literature estimators for the univariate case using $k = 1, 2, ..., l$ multi-auxiliary variables for the no information case, i.e. when there is no information about the auxiliary variables in the population before sampling.

We modify the regression estimator of Isaki (1983), which uses multi-auxiliary variables for the no information case as,

$$t_{reg} = s^2_{y_{(2)}} + \sum_{k=1}^{l} \beta_k \left(s^2_{x_{(1)k}} - s^2_{x_{(2)\mathrm{k}}} \right). \tag{3}$$

The mean square error (MSE) of t_{reg} is given,

$$MSE(t_{reg}) = S_y^2 \left[\gamma_2 \Sigma_{y_{(1\times 1)}} - (\gamma_2 - \gamma_1) \Sigma_{yx_{(1\times l)}} \Sigma^{-1}_{x_{(l\times l)}} \Sigma_{xy_{(l\times 1)}} \right]. \tag{4}$$

Another regression-type estimator using multi-auxiliary variables is developed,

$$t_{rg} = s^2_{y_{(2)}} + \sum_{k=1}^{l} \alpha_k \left(\bar{x}_{(1)\mathrm{k}} - \bar{x}_{(2)\mathrm{k}} \right), \ k = 1, 2, ..., l. \tag{5}$$

The MSE of t_{rg} is given,

$$t_{rg} = S_y^2 \left[\gamma_2 \Sigma_{y_{(1\times 1)}} - (\gamma_2 - \gamma_1) \Sigma_{yx_{d(1\times l)}} \Sigma^{-1}_{x_{d(l\times l)}} \Sigma_{x_d y_{(l\times 1)}} \right]. \tag{6}$$

Following Sunaullah et al. (2016) we also developed the regression-cum-exponential estimator,

$$t_a = \left(s^2_{y_{(2)}} + \sum_{k=1}^{l} \alpha_k \left(\bar{x}_{(1)\mathrm{k}} - \bar{x}_{(2)\mathrm{k}} \right) \right) \left(\exp \sum_{k=1}^{l} d_k \left(\frac{\bar{x}_{(1)\mathrm{k}} - \bar{x}_{(2)\mathrm{k}}}{\bar{x}_{(1)\mathrm{k}} + \bar{x}_{(2)\mathrm{k}}} \right) \right), \tag{7}$$

Where α_k and d_k are the known quantities we take the positive values for both

The MSE of t_a is given,

$$t_{rg} = S_y^2 \left[\gamma_2 \Sigma_{y_{(1\times 1)}} - (\gamma_2 - \gamma_1) \Sigma_{yx_{d(1\times l)}} \Sigma^{-1}_{x_{d(l\times l)}} \Sigma_{x_d y_{(l\times 1)}} \right]. \tag{8}$$

Assuming simple random sampling without replacement *(SRSWOR)*, the expectation results for the relative error are derived in Section II. The regression-cum-exponential type estimator of the variance and the special cases of our proposed estimator are discussed in Section III. The proposed estimator depends on unknown constants and, by finding their optimum values, our estimator is shown to possess smaller variance than the unbiased variance estimator, regression estimator, and the other modified existing estimators are also discussed in section III. We present a simulation study in Section IV where the gamma distribution is used to demonstrate the usefulness of the proposed multi-auxiliary variable estimator. Finally, the paper concludes with a summary and discussions in Section V. The objective of this paper is to use the multi- auxiliary information to decrease the mean square error of our proposed estimator.

2. Some Useful Quantities Under Two-Phase Sampling

In this section, we present some useful quantities to be used to derive the expectation results for the mean square error of proposed and existing estimators. Using SRSWOR, the following results for two-phase sampling (nested) design hold, Let,

$$s_{y_{(2)}}^2 = S_y^2 \ (1+\varepsilon_{y_{(2)}}), s_{x(1)k}^2 = S_{x_k}^2 \ (1+\varepsilon_{x_{(1)k}})$$

$$\bar{x}_{x(1)k} = \bar{X}_{x_k} (1+e_{x_{d(1)k}}), \bar{x}_{x(2)k} = \bar{X}_{x_k} (1+e_{x_{d(2)k}})$$

$$E\left(\varepsilon_{y_2}\right) = E\left(\varepsilon_{x_{(2)k}}\right) = E\left(\bar{e}_{x_{d(1)k}}\right) = E\left(\bar{e}_{x_{d(2)k}}\right) = 0$$

$$E_2\left(\varepsilon_{y_{(2)}}^2\right) = \mathrm{A}_y = \gamma_2 [\beta_2(y)-1],$$

$$E_2\left(\varepsilon_{x_{(2)}}^2\right) = \mathrm{A}_x = \gamma_2 [\beta_2(\mathrm{x})-1] ,$$

$$E_1E_{2/1}\left(\varepsilon_{y_2}\varepsilon_x\right) = \mathrm{A}_{yx} = \gamma_1 [\lambda_{220}-1]$$

$$E_1\left(\bar{e}_{x_{d(1)k}}^2\right) = \mathrm{A}_{x_{d(1)k}} = \gamma_1 C_{x_{dk}}^2 , {}_1\left(\bar{e}_{x_{d(2)k}}^2\right) = \mathrm{A}_{x_{d(2)k}} = \gamma_2 C_{x_{dk}}^2$$

$$E_1E_{2/1}\left(\bar{e}_{x_{d(1)k}}\varepsilon_{y_2}\right) = \mathrm{A}_{yx_{d(1)k}} = \gamma_1 \lambda_{210} C_{x_{dk}} , E_1E_{2/1}\left(\bar{e}_{x_{d(1)k}}\varepsilon_{y_2}\right) = \mathrm{A}_{yx_{d(2)k}}$$

where,

$$\lambda_{p1....pmq1....ql} = \frac{\mu_{p_1......p_{mq1.....q_l}}}{\mu_{200}^{p_1......p_m/2}\ \mu_{020}^{q_1.......q_l/2}}$$

and $\mu_{p_1......p_{mq1.....q_l}} = \dfrac{1}{N}\sum_{i}^{N}(Y_{i1}-\bar{Y}_1)^{p_1}(Y_{i2}-\bar{Y}_2)^{p_2}....(Y_{ij}-\bar{Y}_1)^{p_m}(X_{i1}-\bar{X})^{q_1}(X_{i2}-\bar{X})^{q_2}...(X_{ik}-\bar{X})^{q_l}$.

and $\gamma_1 = \dfrac{1}{n_1}, \gamma_2 = \dfrac{1}{n_2}$.

Using multi-auxiliary variables for univariate estimators, the following results are used for finding the mean square error (MSE) expressions,

Let,

$$\Delta_y = \left[\varepsilon_{y_{(2)}}\right]_{(1\times 1)}$$

$$\Delta_x = \left[\varepsilon_{x_{(1)1}}-\varepsilon_{x_{(2)1}} \quad \varepsilon_{x_{(1)2}}-\varepsilon_{x_{(2)2}} \quad . \quad . \quad \varepsilon_{x_{(1)l}}-\varepsilon_{x_{(2)1}}\right]_{(1\times l)}$$

$$\bar{D}_{x_d} = \left[\mathrm{e}_{x_{(1)1}}-\mathrm{e}_{x_{(2)1}} \quad \mathrm{e}_{x_{(1)2}}-\mathrm{e}_{x_{(2)2}} \quad . \quad . \quad \mathrm{e}_{x_{(1)l}}-\mathrm{e}_{x_{(2)l}}\right]_{(1\times l)}$$

$$\left.\begin{aligned} &E_1E_{2/1}\left(\Delta_y{}'\Delta_y\right) = \gamma_2 \Sigma_{y_{(1\times 1)}}, E_1E_{2/1}\left(\Delta_x{}'\Delta_y\right) = \left(\gamma_1-\gamma_2\right)\Sigma_{xy_{(l\times 1)}} \\ &E_1E_{2/1}\left(\Delta_y{}'\bar{D}_{x_d}\right) = \left(\gamma_1-\gamma_2\right)\Sigma_{yx_{d(1\times l)}}, E_1E_{2/1}\left(\bar{D}_{x_d}{}'\bar{D}_{x_d}\right) = \left(\gamma_2-\gamma_1\right)\Sigma_{x_{d\,(l\times l)}} \\ &E_1E_{2/1}\left(\Delta_x{}'\Delta_x\right) = \left(\gamma_2-\gamma_1\right)\Sigma_{x_{(l\times l)}} \end{aligned}\right| \tag{9}$$

3. The Proposed Estimator

Motivated by Abu-Dayyeh & Ahmend (2005) and Sunaullah et al. (2016) a regression-cum-exponential type estimator using several auxiliary variables is proposed. Let $s_{x(1)k}^2$ and $s_{x(2)k}^2$ be the sample variances of auxiliary variable *k* (unknown) when samples of size n_1 and n_2 are selected as the first phase and the second phase samples, respectively. Further, it is assumed that $s_{y(2)}^2$ denote the sample variance of the study variable of size n_2 selected at second-phase ($n_2 \subset n_1$). The proposed estimator along with the derivation of mean square error is expressed as:

$$t_s = \left(s_{y_{(2)}}^2 + \sum_{k=1}^{l}\eta_k (s_{x(1)k}^2 - s_{x(2)k}^2)\right)\exp\left(\sum_{k=1}^{l} b_k \left\{\frac{s_{x(1)k}^2 - s_{x(2)k}^2}{s_{x(1)k}^2 + s_{x(2)k}^2}\right\}\right), \ k=1,2,...,l. \tag{10}$$

Note that positive values of b_k produce different families of univariate exponential ratio estimators, and negative values for this constant will yield different families of univariate exponential product estimators. For this particular study, η_k is assumed to be unknown and we thus need to estimate its optimum value.

Up to first order of approximation, (10) can be written as in ε_k's,

$$t_s = \left(S^2_{y_{(2)}} \left(1+\varepsilon_{y_{(2)}}\right) + \sum_{k=1}^{l} \eta_k \left\{ \varepsilon_{x_{(1)k}} + \varepsilon_{x_{(2)k}} \right\} \right) \exp\left(\sum_{k=1}^{l} b_k \left\{ \frac{\left(\varepsilon_{x_{(1)k}} - \varepsilon_{x_{(2)k}}\right)}{2} \left(1 + \frac{\left(\varepsilon_{x_{(1)k}} + \varepsilon_{x_{(2)k}}\right)}{2} \right)^{-1} \right\} \right),$$

Now retaining the terms up to the order one and after some simplification we have,

$$t_s = S^2_{y_{(2)}} \left(1 + \varepsilon_{y_{(2)}} + \frac{1}{2} \sum_{k=1}^{l} \left(2\eta_k \frac{S^2_{x_k}}{S^2_{y_{(2)}}} + b_k \right) \left(\varepsilon_{x_{(1)k}} - \varepsilon_{x_{(2)k}} \right) \right), \tag{11}$$

Subtracting $S^2_{y_{(2)}}$ from both side of (10) and we have,

$$t_s - S^2_{y_{(2)}} = S^2_{y_{(2)}} \left(\varepsilon_{y_{(2)}} + \frac{1}{2} \sum_{k=1}^{l} \left(2\eta_k \phi + b_k \right) \left(\varepsilon_{x_{(1)k}} - \varepsilon_{x_{(2)k}} \right) \right), \tag{12}$$

where $\phi = \dfrac{S^2_{x_k}}{S^2_{y_{(2)}}}$.

The MSE of can be obtained by squaring (12) and taking expectations,

$$MSE(t_s) = E\left(t_s - S_y^4\right)^2 = S_y^4 E_1 E_{2/1} \left(\Delta_{y_{(1\times1)}} + \frac{1}{2} \Delta_{x_{(1\times l)}} \Psi_{(l\times1)} \right)' \left(\Delta_{y_{(1\times1)}} + \frac{1}{2} \Delta_{x_{(1\times l)}} \Psi_{(l\times1)} \right), \tag{13}$$

where,

$$\Psi_{(l\times1)} = \left(2\eta_k \phi + b_k \right)_{(l\times1)} \tag{14}$$

Using the results in Section 2, the MSE expression is obtained as:

$$MSE(t_s) = S_y^4 \begin{pmatrix} \gamma_2 \Sigma_{y_{(1\times1)}} + \frac{1}{2}(\gamma_1 - \gamma_2) \Sigma_{yx_{(1\times l)}} \Psi_{(l\times1)} + \frac{1}{2}(\gamma_1 - \gamma_2) \\ \Psi'_{(l\times1)} \Sigma_{yx_{(1\times l)}} + \frac{1}{4}(\gamma_2 - \gamma_1) \Psi'_{(1\times l)} \Sigma_{x_{(l\times l)}} \Psi_{(l\times1)} \end{pmatrix}. \tag{15}$$

By differentiating (15) with respect to Ψ,

$$\frac{\partial MSE(t_s)}{\partial \Psi} = 0,$$

We get the optimum value $\Psi_{opt_{(l\times1)}} = 2 \Sigma^{-1}_{x_{(l\times l)}} \Sigma_{yx_{(l\times1)}}$.

Now substituting $\Psi_{opt_{(l\times1)}}$ in (15), we get the minimum mean square error as:

$$_{\min} MSE(t_s) = S_y^4 \left(\gamma_2 \Sigma_{y_{(1\times1)}} - (\gamma_2 - \gamma_1) \Sigma_{yx_{(1\times l)}} \Sigma^{-1}_{x_{(l\times l)}} \Sigma_{yx_{(l\times1)}} \right). \tag{16}$$

3.1. Remark

It is observed that for b_k we may assume (-1, 0, 1) in (10) and we may obtain a univariate regression-cum-exponential product type, usual regression estimator and regression-cum-exponential ratio type estimator respectively. We may obtain different univariate estimators by assigning different number of auxiliary variables in (10). Furthermore it is also observed that by setting b_k = 0, we may get the regression estimator as in (3) and we get Singh et al. (2009) exponential type ratio and product estimator by setting η_k =0 & b_k =1, b_k =-1 using single auxiliary variable.

4. Simulation Study

We now describe a simulation study to demonstrate the efficiency of our proposed estimator for the univariate case. We consider a study variable (Y) along with four auxiliary variables (X_1, X_2, X_3, X_4). We calculate the minimum mean squares errors of our proposed and modified existing estimators. The percent relative efficiencies (PREs) have also been computed of our proposed estimator t_s,

$$PREs(t_{\bullet}, t_0) = \frac{Var(t_0)}{MSE(t_{\bullet})} \times 100$$

As our proposed estimator is regression-cum-exponential type so here we need to think about the probability distribution for estimating our auxiliary variables. We chose the gamma distribution is the suitable one for the exponential part. It is a continuous distribution and has two parameters; a shape and a scale parameter.

The gamma distribution is characterized by two parameters: a shape parameter (k) and a scale parameter (θ). These parameters govern the shape and scale of the distribution, allowing for a diverse range of distributions.

If we have specific information or data for estimating these parameters (k and θ), you can proceed with generating the auxiliary variables using the gamma distribution. The gamma distribution is often denoted as Gamma, Gamma(k,θ), where k>0 is the shape parameter and θ>0 is the scale parameter.

By using the following information, we generated $X's$ as:

$X_1 \sim \Gamma(0.015, 8)$, $X_2 \sim \Gamma(0.016, 6)$, $X_3 \sim \Gamma(0.017, 7)$, $X_4 \sim \Gamma(0.018, 4)$

And the study variable is simulated as:

$$Y = \sum_{k}^{l} X_k * \varepsilon \qquad \varepsilon \sim \Gamma(2, 2)$$

The study variable Y with four auxiliary variables under our simulation model can be written as:

$$Y_1 = 0.3X_1 + 0.4X_2 + 0.8X_3 + 0.4X_4$$
$$Y = Y_1 * \varepsilon$$

We have selected n_1 units in the first phase sampling using Simple Random Sampling Without Replacement (SRSWOR), followed by selecting a small sample from the previously chosen units in the second phase (n_2), again using SRSWOR. This process has been repeated 1000 times to calculate the values of our proposed and existing estimators.

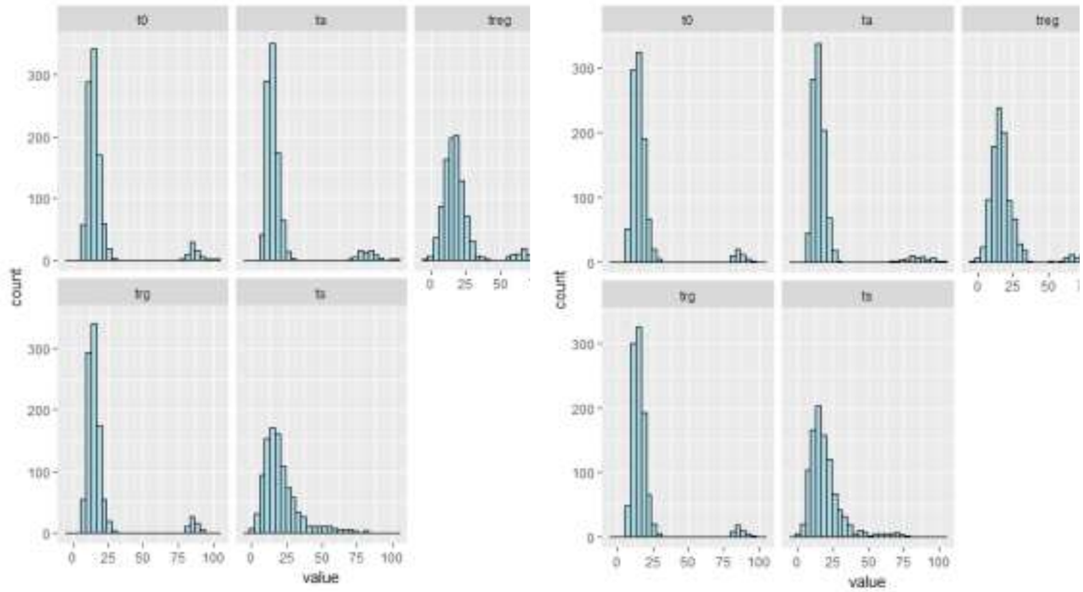
$$\text{var}(t_i) = \frac{1}{q}\sum_{i=1}^{q}(t_i - T)^2, \quad T = \frac{1}{q}\sum_{i=1}^{q} t_i$$

The code for the simulation was carried out using the R statistical package (2016) and the broom (2016) and tidyverse packages (2016).

Table 1: PREs of our proposed estimator with respect to Sample variance (t_0)

Estimator	**PREs** $N = 10,000$		
	$n_1 = 2000$, $n_2 = 600$	$n_1 = 1800$, $n_2 = 600$	$n_1 = 1500$, $n_2 = 600$
$t_0 = s_y^2$	100	100	100
t_s	156.0718	133.2068	121.0597
t_{reg}	100.1531	100.3522	100.0922
t_{rg}	100.016	100.0291	100.0016
t_a	111.0891	104.4206	102.4855

Table 1 summarizes the results of the simulation study. In all cases, our proposed estimator is the most efficient estimator. Because the regression-cum-exponential estimator has been implemented for the no information case, when $n_1 = 1500$, then the gain in efficiency of the estimator is lower as compared to $n_1 = 1800$ and 2000 because there is less auxiliary information from the first phase ($n_1 = 1500$). Therefore, for small phase 1 samples, we observe smaller gains in efficiency.



(a) $n_1 = 2000$ (b) $n_1 = 1500$

Figure 1: Distribution of five estimators of the population variance for simulated data for two phase n_1 sample sizes: the unbiased estimator (t_0); exponential estimator with mean auxiliary information (t_a); regression estimator using variance auxiliary information (t_{reg}); regression estimator with mean auxiliary information (t_{rg}); and our proposed estimator (t_s).

By increasing the phase-I sample size, we see more dramatic gains in efficiency. Figure 1 shows that the gains in the efficiency result from stabilizing the estimator when it takes on large values. In each subfigure (a) and (b), we see that the upper tail of the distribution for our proposed estimator is much lighter than the tails of the other estimators (in particular, there is either a much smaller or completely eliminated secondary mode).

5. Conclusion and Discussion

We use a simulation model to check the efficiency of our proposed estimator and we find that our regression-cum-exponential estimator has less mean square error than the unbiased estimator, Isaki's (1983) traditional modified regression estimator and the modified Sanaullah et al. (2016) estimator as shown in Table 1. We consider three scenarios by taking different sizes of n_1 with the same size of n_2. When we are taking a small size of n_1 from the population, it is found that we have less opportunity to exploit the auxiliary information for no information case. By increasing the size of phase one sample, our auxiliary variables provide much more information and lead to the best results at n_1 =2000. We use the gamma distribution and it is well known that exponential distribution is the special case of continuous gamma distribution. The usual regression estimator behaves well when the simulation model generates symmetric error distributions and linear associations between the auxiliary variables and the outcome. However, in many practical situations, we observe non-symmetric and skewed situations and then the regression-cum-exponential type estimators should provide better results than the simple regression estimator. In this paper, we have clearly identified the importance of exponential type estimator. In Fig. 1 (a), it is shown that our estimator is positively skewed (due to the gamma distribution properties and the multiplicative error) and one can observe that without using the exponential part as in t_{reg} and t_{rg}, a bimodal sampling distribution occurs, with many very poor estimates of the target population variance. Similar results have been shown in Fig.1 (b) by changing the size of phase one sample i.e. n_1.

In future research, we plan to expand our work by considering the extension of our proposed methodology to bivariate and multivariate scenarios, utilizing multi-auxiliary variables. This extension would enable us to tackle more complex real-life phenomena and improve our understanding of the relationships between variables. To provide a clearer explanation of our work, we intend to illustrate its application to a real-life phenomenon through simulated results. This approach will help bridge the gap between theoretical concepts and practical implications, offering a more tangible understanding of our proposed estimator. Furthermore, our results have demonstrated the superior efficiency of our proposed estimator compared to existing modified estimators that utilize multi-auxiliary variables. This finding emphasizes the practical utility and effectiveness of our approach in accurately estimating population variances.

References

Arcos, A., Rueda, M., Martinez, M. D., Gonzalez, S. and Roman, Y. (2005). Incorporating the auxiliary information available in variance estimation. *Applied Mathematics and Computation*, 160 (2), 387–399.

Abu-Dayyeh, W. and Ahmed, M. (2005). Ratio and Regression Estimators for the Variance under Two-Phase sampling. *International Journal of Statistical Sciences,* 4, 49-56.

Abid, M., Sherwani, K.,A., R., Tahir, M., Nazir, Z., H., and Riaz, M. (2020). An Improved and Robust Class of Variance Estimator. *Scientia Arania*:

Asghar, A., Sanaullah A. and Hanif, M. (2014). Generalized exponential type estimator for population variance in survey sampling. *Revista Colombiana de Estadistica*, 37(1), 211–222.

Adichwal, N. K., Kumar, J., and Singh, R. (2018). An improved generalized class of estimators for population variance using auxiliary variables. *Cogent Mathematics & Statistics*, 5(1), 1454579.

Ahmad S, Hussain S, Ullah K, Zahid E, Aamir M, Shabbir J, Ahmad Z, Alshanbari HM, Alajlan W (2022). A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling. *PLoS One*, 17(11), e0276540.

Cekim, O. H.,Kadilar, C. (2020) ln-Type Variance Estimators in Simple Random Sampling. *Pakistan Journal of Statistics and Operation Research*, pages 689-696.

Dubey, V. and Sharama, H.K. (2008). On estimation population variance using auxiliary information. *Statistics in Transition new series*, 9(1), 7-18.

Gupta, S. and Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. *Hacettepe Journal of Mathematics and Statistics,* 37(1), 57–67.

Isaki, C. (1983). Variance estimation using auxiliary information. *Journal of American Statistical Association*, 78, 117-123.

John Kung'u, J. , Chumba, G. and Odongo, L. (2014) Mixture Regression Estimators Using Multi-Auxiliary Variables and Attributes in Two-Phase Sampling. *Open Journal of Statistics*, 4, 355-366.

Kadilar, C. and Cingi, H. (2006). Improvement in variance using auxiliary Information. *Hecettepe Journal of Mathematics and Statistics*, 35 (1), 111–115.

Neyman, J. (1938). Contributions to the theory of sampling human populations. *Journal of the American Statistical Association,* 33, 101-116.

Masood, S. and Shabbir, J. (2016). Generalized Multi-Phase Regression-type Estimators under the Effect of Measurement Error to Estimate the Population Variance. *Hacettepe Journal of Mathematics and Statistics*, 45(4), 1297-1306.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika,* 45,154-165.

Prasad, B. and Singh, H. P. (1990). Some improved ratio type estimators of finite population variance in sample surveys. *Communication in Statistics: Theory and Methods,* 19, 1127-1139.

Sanaullah, A., Hanif, M. and Asghar, A. (2016). Generalized Exponential Estimators for Population Variance under Two-Phase. *International Journal of Applied Computational Mathematics,* 2(1), 75-84.

Sen, A.R. (1972). Successive Sampling with p(p≥ 1) auxiliary variables, T*he Annals of Mathematical Statistics,* 43(6), 2031-2034.

Singh, H. P., Chandra, P. and Singh, S. (2003). Variance estimation using multi-auxiliary information for random non-response in survey sampling. *STATISTICA*, 23-40.

Singh, R., Chauhan, P., Sawanand, N., Smarandache F. (2011). Improved exponential estimator for population variance using two auxiliary variables. *Italian Journal of Pure and Applied Mathematic*s, 28, 103-110.

Singh H.P. and Pal S. K. (2017) Improvement over variance estimation using auxiliary information in sample surveys, *Communications in Statistics - Theory and Methods*,15, 7732-7750.

Shabbir, J., Gupta, S. (2015). A Note on Generalized Exponential Type Estimator for Population Variance in Survey Sampling. *Revista Colombiana de Estadística,* 38(2), 385-397.

Singh, H. P., Srivastava, M. K., Srivastava, N., Tarray, T. A., et al. (2015). Chain regression-type estimator using multiple auxiliary information in successive sampling. *Hacettepe Journal of Mathematics and Statistics*, 44(5), 1247-1256.

Niaz, I., Sanaullah, A., Saleem, I., and Shabbir, J. (2021). An improved efficient class of estimators for the population variance. *Concurrency and Computation: Practice and Experience*,

Upadhyaya, L.N. and Singh, H.P. (1983). Use of auxiliary information in the estimation of population variance. *Mathematical Forum*, VI (2), 33-36.

Zaman, T., and Kadilar, C. (2021). New class of exponential estimators for finite population mean in two-phase sampling". *Communications in Statistics-Theory and Methods*, 50(4), 874-889.