REVSTAT – Statistical Journal
Volume 22, Number 3, July 2024, 277–292
https://doi.org/10.57805/revstat.v22i3.508

Estimation of Population Variance for a Sensitive Variable in Stratified Sampling using Randomized Response Technique

Authors: Badr Aloraini
– Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, North Carolina, USA
boalorai@uncg.edu

Sadia Khalil
– Department of Statistics, Lahore College for Women University, Lahore, Pakistan
sadia_khalil@hotmail.com

Muhammad Nouman Qureshi
– Department of Statistics, National College of Business Administration and Economics, Lahore, Pakistan
nqureshi633@gmail.com

Sat Gupta ✉
– Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, North Carolina, USA
sngupta@uncg.edu

Received: May 2021 Revised: June 2022 Accepted: June 2022

Abstract:

- In this paper, Randomized Response technique (RRT) is used to propose some separate and combined variance estimators for a sensitive variable using stratified random sampling. The performances of the proposed estimators are examined using a unified measure of respondent privacy and estimator efficiency.

Keywords:

- *auxiliary information; mean squared error; stratified random sampling; respondent privacy; variance estimation.*

AMS Subject Classification:

- 62D05.

✉ Corresponding author

1. INTRODUCTION

Our main focus in this study is on variance estimation for sensitive variables in stratified sampling. Many researchers have dealt with the problem of mean and variance estimation under simple random sampling and stratified random sampling when the study variable is non-sensitive and is directly observable. Zahid and Shabbir (2018) [17] and many other authors have investigated the problem of mean estimation in stratified random sampling when the study variable is non-sensitive. Important contributions in the area of variance estimation in stratified random sampling for non-sensitive random variables have been made by Kadilar and Cingi (2006) [8], Sidelel *et al.* (2014) [12], Özel *et al.* (2014) [10], Clement (2018) [2], Sanaullah *et al.* (2017) [11], Younis and Shabbir (2019) [16], and Asghar *et al.* (2019) [1]. In all of these studies, the study variable is directly observed and an auxiliary variable is used to increase the efficiency of estimation.

In research involving sensitive survey questions, standard estimation techniques are unreliable. Warner(1965) [14] introduced the Randomized Response Technique (RRT) as a research method to reduce response Bias in estimation of a sensitive study variable and at the same time improve the respondent cooperation. Many authors, including Kalucha *et al.* (2017) [9] and Zhang *et al.* (2021) [18], have estimated the mean of a sensitive study variable under stratified sampling. However, not much work exists for variance estimation under RRT. Gupta *et al.* (2020) [5] introduced several variance estimators under RRT in simple random sampling. The primary goal of this study is to re-examine the Gupta *et al.* (2020) [5] study in the context of stratified random sampling.

Let us consider Y and X to be the observed and auxiliary variables defined on a finite population $U = \{U_1, U_2, ..., U_N\}$. We assume that Y is sensitive in nature and we observe a scrambled version of it given by $Z = TY + S$, where T, S, Y and X are mutually uncorrelated. Let the population be divided into L homogeneous strata with N_h unites $(h = 1, 2, ..., L)$ in the h^{th} stratum such that $\sum_{h=1}^{L} N_h = N$. From h^{th} stratum, a simple random sample of size n_h is drawn without replacement such that $\sum_{h=1}^{L} n_h = n$. Let (x_{hi}, y_{hi}, z_{hi}) be the observed values on the variables X, Y, and Z in the h^{th} stratum. Let $\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$, $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$, $\bar{z}_{st} = \sum_{h=1}^{L} W_h \bar{z}_h$ be the stratified sample means where $\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$, $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$, $\bar{z}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} z_{hi}$ are the stratum sample means and $\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$, $\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$, $\bar{Z}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} z_{hi}$ are corresponding population stratum means. Let $W_h = \frac{N_h}{N}$ $(h = 1, 2, ..., L)$ be the known stratum weights.

The population variance of the study variable in stratified sampling is given by Kadilar and Cingi (2006) [8] as

$$\sigma_{c0}^2 = \sum_{h=1}^{L} W_h \sigma_{yh}^2 + \sum_{h=1}^{L} W_h (\bar{Y}_h - \bar{Y})^2. \tag{1.1}$$

The combined ordinary and combined ratio estimators of population variance given by Kadilar and Cingi (2006) [8] in stratified sampling are given, respectively, by

$$t_{c0} = \sum_{h=1}^{L} W_h s_{yh}^2 + \sum_{h=1}^{L} W_h (\bar{y}_h - \bar{y}_{st})^2, \tag{1.2}$$

and

$$t_{c1} = t_{c0}\left(\frac{\sigma_x^2}{s_{xst}^2}\right), \text{ where } s_{xst}^2 = \sum_{h=1}^{L} W_h s_{xh}^2 + \sum_{h=1}^{L} W_h (\bar{x}_h - \bar{x}_{st})^2. \tag{1.3}$$

Some authors including Özel *et al.* (2014) [10] have suggested the separate ordinary and separate ratio estimators of population variance in stratified sampling which are given respectively by

$$t_{s0} = \sum_{h=1}^{L} W_h s_{yh}^2, \tag{1.4}$$

and

$$t_{s1} = \sum_{h=1}^{L} W_h \left(\frac{s_{yh}^2}{s_{xh}^2}\right) \sigma_{xh}^2. \tag{1.5}$$

In this paper, we have considered the problem of estimating population variance using auxiliary information by adapting Kadilar and Cingi (2006) [8], Özel *et al.* (2014) [10], and Gupta *et al.* (2020) [5] under RRT. We will discuss the proposed combined variance estimators in detail in Section 2. Separate variance estimators will be discussed in detail in Section 3. We also examine the effect of ignoring the term $\sum_{h=1}^{L} W_h (\bar{y}_h - \bar{y}_{st})^2$ in (1.2) on the estimates of the variance in stratified random sampling. Section 4 presents the results of a simulation study; Section 5 presents a real data example; and Section 6 provides some concluding remarks.

2. SOME COMBINED VARIANCE ESTIMATORS IN STRATIFIED RANDOM SAMPLING

In this study, the respondent is asked to provide a scrambled response for the sensitive study Y by using the generalized RRT model given by $Z = TY + S$, as in Diana and Perri (2011) [3], where S and T are uncorrelated scrambling variables such that $E(S) = 0$ and $E(T) = 1$. Gupta *et al.* (2020) [5] used this RRT model for estimating the population variance in simple random sampling. They proposed the following estimators:

$$t_0(R) = \frac{s_z^2 - \sigma_S^2 - \sigma_T^2 * \bar{z}^2}{\sigma_T^2 + 1}, \tag{2.1}$$

$$t_1(R) = t_0(R) * \left(\frac{\sigma_x^2}{s_x^2}\right), \tag{2.2}$$

and

$$t_p(R) = \left[t_0(R) + (\sigma_x^2 - s_x^2)\right] * \left(\frac{(\alpha\sigma_x^2 + \beta)}{\omega(\alpha s_x^2 + \beta) + (1-\omega)(\alpha\sigma_x^2 + \beta)}\right)^g, \tag{2.3}$$

where α and β are suitably chosen constants associated with the auxiliary variable X. With $g = 1$, one can obtain various ratio estimators, and with $g = -1$ one can obtain various product estimators. ω is an unknown whose optimal value will be used.

Motivated by Gupta *et al.* (2020) [5] and Kadilar and Cingi (2006) [8], we propose the following combined variance estimators in the stratified random sampling.

2.1. The Combined Basic Variance Estimator

Based on the RRT model $Z = TY + S$, we have σ^2_{zh} as

$$\sigma^2_{zh} = \sigma^2_{Th}(\sigma^2_{yh} + \mu^2_{yh}) + \sigma^2_{yh} + \sigma^2_{Sh}.$$

Rearranging, we get

$$\sigma^2_{yh} = \frac{\sigma^2_{zh} - \sigma^2_{Sh} - (\sigma^2_{Th} * \bar{Z}^2_h)}{\sigma^2_{Th} + 1}.$$

The population variance of the study variable in stratified sampling is given by

$$\sigma^2_{c0}(R) = \sum_{h=1}^{L} W_h \left(\frac{\sigma^2_{zh} - \sigma^2_{Sh} - \sigma^2_{Th} * \bar{Z}^2_h}{\sigma^2_{Th} + 1} \right) + \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2. \tag{2.4}$$

Let

$$\sigma^2_{c0}(R) = A_1 + B_1,$$

where

$$A_1 = \sum_{h=1}^{L} W_h \left(\frac{\sigma^2_{zh} - \sigma^2_{Sh} - \sigma^2_{Th} * \bar{Z}^2_h}{\sigma^2_{Th} + 1} \right) \text{ and } B_1 = \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2.$$

We have our first proposed combined estimator given by

$$t_{c0}(R) = \sum_{h=1}^{L} W_h \left(\frac{s^2_{zh} - \sigma^2_{Sh} - \sigma^2_{Th} * \bar{z}^2_h}{\sigma^2_{Th} + 1} \right) + \sum_{h=1}^{L} W_h (\bar{z}_h - \bar{z}_{st})^2. \tag{2.5}$$

Let

$$t_{c0}(R) = \hat{A}_1 + \hat{B}_1,$$

where

$$\hat{A}_1 = \sum_{h=1}^{L} W_h \left(\frac{s^2_{zh} - \sigma^2_{Sh} - \sigma^2_{Th} * \bar{z}^2_h}{\sigma^2_{Th} + 1} \right) \text{ and } \hat{B}_1 = \sum_{h=1}^{L} W_h (\bar{z}_h - \bar{z}_{st})^2.$$

To obtain the Bias and MSE expressions for the proposed estimators in the stratified random sampling, we define the following error terms

$$\delta_{zh} = \frac{s^2_{zh} - \sigma^2_{zh}}{\sigma^2_{zh}}, \quad e_{zh} = \frac{\bar{z}_h - \bar{Z}_h}{\bar{Z}_h}, \quad e_{zst} = \frac{\bar{z_{st}} - \bar{Z}}{\bar{Z}}, \quad e_{xst} = \frac{\bar{x_{st}} - \bar{X}}{\bar{X}},$$

such that

$$E(\delta_{zh}) = E(e_{zh}) = E(e_{zst}) = E(e_{xst}) = 0,$$

$$E(\delta^2_{zh}) = \theta_h(\lambda_{40h} - 1), \quad E(\delta^2_{xh}) = \theta_h(\lambda_{04h} - 1), \quad E(\delta_{zh}\delta_{xh}) = \theta_h(\lambda_{22h} - 1),$$

$$E(\delta_{zh}e_{zh}) = \theta_h\lambda_{30h}C_{zh}, \quad E(\delta_{xh}e_{zh}) = \theta_h\lambda_{12h}C_{zh}, \quad E(e^2_{zh}) = \theta_h C^2_{zh},$$

$$E(e_{zst}e_{zh}) = \sum_{h=1}^{L} W_h\theta_h\sigma^2_{zh}, \quad E(e^2_{zst}) = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L} W^2_h\theta_h\sigma^2_{zh}, \quad E(e^2_{xst}) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W^2_h\theta_h\sigma^2_{xh},$$

$$E(e_{xst}e_{zh}) = \sum_{h=1}^{L} W_h\theta_h\sigma_{zxh}, \quad E(e_{xst}e_{zst}) = \frac{1}{\bar{Z}\bar{X}}\sum_{h=1}^{L} W^2_h\theta_h\sigma_{zxh},$$

where

$$\sigma_{zxh} = \rho_{zxh}\sigma_{zh}\sigma_{xh}, \quad \rho_{zxh} = \frac{\rho_{yxh}}{\sqrt{1 + \frac{\sigma_{Th}^2(\sigma_{yh}^2 + \mu_{yh}^2) + \sigma_{Sh}^2}{\sigma_{yh}^2}}}, \quad \lambda_{rsh} = \frac{\mu_{rsh}}{\mu_{20h}^{\frac{r}{2}}\mu_{02h}^{\frac{s}{2}}},$$

$$\mu_{rsh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(Z_{hi} - \bar{Z}_h)^r(X_{hi} - \bar{X}_h)^s \quad \text{and} \quad C_{zh}^2 = C_{yh}^2\sigma_{Th}^2 + \left(\frac{\sigma_{Sh}^2}{\bar{Y}_h^{\ 2}}\right).$$

Consider the first term

$$\hat{A}_1 = \sum_{h=1}^{L} W_h\left(\frac{s_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1}\right). \tag{2.6}$$

Rewriting (2.6), we have

$$\hat{A}_1 = \sum_{h=1}^{L} W_h\left(\frac{\sigma_{zh}^2(1 + \delta_{zh}) - \sigma_{Sh}^2 - \sigma_{Th}^2[\bar{Z}_h(1 + e_{zh})]^2}{\sigma_{Th}^2 + 1}\right).$$

Subtracting A_1 on both sides, we obtain

$$(\hat{A}_1 - A_1) = \sum_{h=1}^{L} W_h\left(\frac{\sigma_{zh}^2\delta_{zh} - 2\sigma_{Th}^2\bar{Z}_h^2 e_{zh} - \sigma_{Th}^2\bar{Z}_h^2 e_{zh}^2}{\sigma_{Th}^2 + 1}\right). \tag{2.7}$$

Taking the expectation on both sides of (2.7), the Bias of $\hat{A}_1$ is obtained as

$$\text{Bias}(\hat{A}_1) \approx -\sum_{h=1}^{L} \theta_h W_h\left(\frac{\sigma_{Th}^2\bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1}\right). \tag{2.8}$$

By squaring both sides of (2.7) and using the first order approximation, the MSE is obtained as

(2.9)

$$\text{MSE}(\hat{A}_1) \approx \sum_{h=1}^{L} \theta_h W_h^2\left(\frac{1}{(\sigma_{Th}^2 + 1)^2}\right)\left(\sigma_{zh}^4(\lambda_{40h} - 1) + 4\sigma_{Th}^4\bar{Z}_h^4 C_{zh}^2 - 4\sigma_{zh}^2\sigma_{Th}^2\bar{Z}_h^2\lambda_{30h}C_{zh}\right).$$

Consider the second term

$$\hat{B}_1 = \sum_{h=1}^{L} W_h(\bar{z}_h - \bar{z}_{st})^2. \tag{2.10}$$

Rewriting (2.10), we have

$$\hat{B}_1 = \sum_{h=1}^{L} W_h[\bar{Z}_h(1 + e_{zh}) - \bar{Z}(1 + e_{zst})]^2. \tag{2.11}$$

Expanding (2.11), and restricting to terms up to order 2, we have

$$\hat{B}_1 = \sum_{h=1}^{L} W_h[(\bar{Z}_h - \bar{Z})^2 + (\bar{Z}_h e_{zh} - \bar{Z}e_{zst})^2 + 2(\bar{Z}_h^2 e_{zh} - \bar{Z}_h\bar{Z}e_{zst} - \bar{Z}_h\bar{Z}e_{zh} + \bar{Z}_h^2 e_{zst})]. \tag{2.12}$$

Subtracting B_1 on both sides, we obtain

(2.13)

$$(\hat{B}_1 - B_1) = \bar{Z}^2 e_{zst}^2 + \sum_{h=1}^{L} W_h[\bar{Z}_h^2 e_{zh}^2 - 2\bar{Z}_h\bar{Z}e_{zh}e_{zst} + 2(\bar{Z}_h^2 e_{zh} - \bar{Z}_h\bar{Z}e_{zst} - \bar{Z}_h\bar{Z}e_{zh} + \bar{Z}_h^2 e_{zst})].$$

Taking the expectation on both sides of (2.13), the Bias of $\hat{B}_1$ is obtained as

$$\text{Bias}(\hat{B}_1) \approx \bar{Z}^2 \sum_{h=1}^{L} W_h^2 \theta_h C_{zh}^2 + \sum_{h=1}^{L} W_h \theta_h [\bar{Z}_h^2 C_{zh}^2 - 2\bar{Z}_h \bar{Z} \sigma_{zh}^2]. \tag{2.14}$$

By squaring both sides of (2.13), using the first order approximation and simplifying, the MSE is obtained as

$$\begin{aligned}\text{MSE}(\hat{B}_1) \approx\ & 4\bar{Z}^4 \sum_{h=1}^{L} W_h^2 \theta_h C_{zh}^2 \\ & + \sum_{h=1}^{L} W_h^2 \theta_h \left[4\bar{Z}_h^2 C_{zh}^2 - 8\bar{Z}_h \bar{Z} \sum_{h=1}^{L} W_h \sigma_{zh}^2 (\bar{Z}_h - \bar{Z})^2 + 4\bar{Z}_h \bar{Z}^2 \sum_{h=1}^{L} W_h^2 C_{zh}^2 (\bar{Z}_h - 2\bar{Z}) \right].\end{aligned} \tag{2.15}$$

The expressions for Bias and MSE of $t_{c0}(R)$ are given by

$$\text{Bias}(t_{c0}(R)) = \text{Bias}(\hat{A}_1) + \text{Bias}(\hat{B}_1), \tag{2.16}$$

and

$$\text{MSE}(t_{c0}(R)) \approx \text{MSE}(\hat{A}_1) + \text{MSE}(\hat{B}_1). \tag{2.17}$$

In (2.17), we assume that $\hat{A}_1$ and $\hat{B}_1$ are uncorrelated. This is not an unreasonable assumption since the sample mean and the sample variance are uncorrelated for normal data. This is also confirmed by large number of simulated values of $\hat{A}_1$ and $\hat{B}_1$ that we generated.

2.2. The Combined Ratio Variance Estimator

$$t_{c1}(R) = \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) * \left(\frac{\sigma_{xh}^2}{s_{xh}^2} \right) \right] + \sum_{h=1}^{L} W_h \left(\bar{z}_h - \frac{\bar{z}_{st}}{\bar{x}_{st}} \bar{X} \right)^2, \tag{2.18}$$

$$t_{c1}(R) = \hat{A}_2 + \hat{B}_2.$$

Consider the first term:

$$\hat{A}_2 = \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) * \left(\frac{\sigma_{xh}^2}{s_{xh}^2} \right) \right]. \tag{2.19}$$

Rewriting (2.19), we have

$$\hat{A}_2 = \sum_{h=1}^{L} W_h \left[\frac{\sigma_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 \bar{Z}_h^2}{\sigma_{Th}^2 + 1} + \frac{2\sigma_{Th}^2 \bar{Z}_h^2 e_{zh} \delta_{xh} - \sigma_{zh}^2 \delta_{zh} \delta_{xh} - \sigma_{Th}^2 \bar{Z}_h^2 e_{zh}^2}{\sigma_{Th}^2 + 1} \right].$$

Subtracting A_1 and taking the expectation on both sides, the Bias of $\hat{A}_2$ is obtained as

$$\text{Bias}(\hat{A}_2) \approx \sum_{h=1}^{L} \theta_h W_h \left[\left(\frac{2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh} - \sigma_{zh}^2 (\lambda_{22h} - 1) - \sigma_{Th}^2 \bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1} \right) \right]. \tag{2.20}$$

For MSE, we have

$$\hat{A}_2 = \sum_{h=1}^{L} W_h \Bigg[\frac{\sigma_{zh}^2 + \sigma_{zh}^2\delta_{zh} - \sigma_{Sh}^2 - \sigma_{Th}^2\bar{Z}_h^2 - 2\sigma_{Th}^2\bar{Z}_h^2 e_{zh} - \sigma_{Th}^2\bar{Z}_h^2 e_{zh}^2}{\sigma_{Th}^2+1}$$
$$\frac{-\sigma_{zh}^2\delta_{xh} - \sigma_{zh}^2\delta_{zh}\delta_{xh} + \sigma_{Sh}^2\delta_{xh} + \sigma_{Th}^2\bar{Z}_h^2\delta_{xh} + 2\sigma_{Th}^2\bar{Z}_h^2 e_{zh}\delta_{xh} + \sigma_{Th}^2\bar{Z}_h^2 e_{zh}^2\delta_{xh}}{\sigma_{Th}^2+1} \Bigg].$$

Simplifying and ignoring second and higher order terms,

$$\hat{A}_2 = \sum_{h=1}^{L} W_h \left[\frac{\sigma_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2\bar{Z}_h^2}{\sigma_{Th}^2+1} + \frac{\sigma_{zh}^2\delta_{zh} - 2\sigma_{Th}^2\bar{Z}_h^2 e_{zh} - \sigma_{zh}^2\delta_{xh} + \sigma_{Sh}^2\delta_{xh} + \sigma_{Th}^2\bar{Z}_h^2\delta_{xh}}{\sigma_{Th}^2+1} \right].$$

Squaring and taking the expectation on both sides, we have

$$\hat{A}_2 = \sum_{h=1}^{L} W_h^2 E\left(\frac{\sigma_{zh}^2\delta_{zh}}{\sigma_{Th}^2+1} - \frac{2\sigma_{Th}^2\bar{Z}_h^2 e_{zh}}{\sigma_{Th}^2+1} - \sigma_{yh}^2\delta_{xh} \right)^2.$$

After some simplifications, the MSE of $\hat{A}_2$ is obtained as

$$\text{MSE}(\hat{A}_2) \approx \sum_{h=1}^{L} \frac{W_h^2\theta_h}{(\sigma_{Th}^2+1)^2} \Bigg[\sigma_{zh}^4(\lambda_{40h}-1) - 2\sigma_{zh}^2\sigma_{yh}^2(\lambda_{22h}-1)(\sigma_{Th}^2+1) + \sigma_{yh}^4(\lambda_{04h}-1)(\sigma_{Th}^2+1)^2$$
$$+ 4C_{zh}\Big(\sigma_{Th}^4\bar{Z}_h^4 C_{zh} - \sigma_{zh}^2\sigma_{Th}^2\bar{Z}_h^2\lambda_{30h} + \sigma_{Th}^2\sigma_{yh}^2\bar{Z}_h^2\lambda_{12h}(\sigma_{Th}^2+1) \Big) \Bigg]. \tag{2.21}$$

Consider the second term:

$$\hat{B}_2 = \sum_{h=1}^{L} W_h \left(\bar{z}_h - \frac{\bar{z}_{st}}{\bar{x}_{st}} \bar{X} \right)^2. \tag{2.22}$$

Repeating the procedure outlined in steps (2.10)–(2.15) for the estimator (2.22), yields definitions of Bias and MSE for $\hat{B}_2$ as

$$\text{Bias}(\hat{B}_2) \approx \bar{Z}^2 \sum_{h=1}^{L} W_h^2\theta_h(C_{zh}^2 + C_{xh}^2) + \sum_{h=1}^{L} W_h\theta_h \Bigg[\bar{Z}_h^2 C_{zh}^2 - 2\bar{Z}_h\bar{Z}\sum_{h=1}^{L} W_h\sigma_{zh}^2$$
$$+ 2\left(\frac{\bar{Z}_h}{\bar{X}} \sum_{h=1}^{L} W_h^2\sigma_{zxh} + \frac{\bar{Z}_h}{\bar{X}} \sum_{h=1}^{L} W_h\sigma_{zxh} - 2\left(\frac{\bar{Z}}{\bar{X}} \sum_{h=1}^{L} W_h^2\sigma_{zxh} \right) \right) \Bigg], \tag{2.23}$$

$$\text{MSE}(\hat{B}_2) \approx 4\bar{Z}^4 \sum_{h=1}^{L} W_h^2\theta_h(C_{zh}^2 + C_{xh}^2) + \sum_{h=1}^{L} W_h^2\theta_h \Bigg[4\bar{Z}_h^2 C_{zh}^2(\bar{Z}_h - \bar{Z})^2$$
$$+ 4\bar{Z}_h^2\bar{Z}^2 \sum_{h=1}^{L} W_h^2(C_{zh}^2 + C_{xh}^2) + 8\bar{Z}_h^3\bar{Z}\sum_{h=1}^{L} W_h\left(\frac{\sigma_z xh}{\bar{Z}\bar{X}} - \sigma_{zh}^2 \right)$$
$$- 8\bar{Z}_h^2\bar{Z}^2\left(\frac{2}{\bar{Z}\bar{X}} \sum_{h=1}^{L} W_h\sigma_{zxh} - \sum_{h=1}^{L} 2W_h\sigma_{zh}^2 + \frac{1}{\bar{Z}\bar{X}} \sum_{h=1}^{L} W_h^2\sigma_{zxh} \right)$$
$$+ 8\bar{Z}_h\bar{Z}^3\left(\frac{1}{\bar{Z}\bar{X}} \sum_{h=1}^{L} W_h^2\sigma_{zxh} - \sum_{h=1}^{L} W_h^2 C_{zh}^2 + \frac{1}{\bar{Z}\bar{X}} \sum_{h=1}^{L} W_h\sigma_{zxh} - \sum_{h=1}^{L} W_h\sigma_{zh}^2 \right)$$
$$- 8\bar{Z}^2 \frac{1}{\bar{Z}\bar{X}} \sum_{h=1}^{L} W_h^2\sigma_{zxh} \Bigg]. \tag{2.24}$$

The expressions for Bias and MSE of $t_{c1}(R)$ are given by

$$\text{Bias}(t_{c1}(R)) = \text{Bias}(\hat{A}_2) + \text{Bias}(\hat{B}_2), \tag{2.25}$$

and

$$\text{MSE}(t_{c1}(R)) \approx \text{MSE}(\hat{A}_2) + \text{MSE}(\hat{B}_2). \tag{2.26}$$

2.3. A Combined Generalized Variance Estimator

We now propose the following class of generalized population variance estimators:

$$\begin{aligned} t_{cp}(R) &= \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) + (\sigma_{xh}^2 - s_{xh}^2) \right] \\ &\quad * \left(\frac{(\alpha\sigma_{xh}^2 + \beta)}{\omega(\alpha s_{xh}^2 + \beta) + (1-\omega)(\alpha\sigma_{xh}^2 + \beta)} \right)^g \\ &\quad + \sum_{h=1}^{L} W_h \left[\left(\bar{z}_h - \left[\bar{z}_{st} + (\bar{X} - \bar{x}_{st}) \right] \right) * \left(\frac{(\alpha\bar{X} + \beta)}{\lambda(\alpha\bar{x}_{st} + \beta) + (1-\lambda)(\alpha\bar{X} + \beta)} \right)^g \right]^2, \end{aligned} \tag{2.27}$$

$$t_{cp}(R) = \hat{A}_3 + \hat{B}_3.$$

Consider the first term:

$$\hat{A}_3 = \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{Sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) + (\sigma_{xh}^2 - s_{xh}^2) \right] * \left(\frac{(\alpha\sigma_{xh}^2 + \beta)}{\omega(\alpha s_{xh}^2 + \beta) + (1-\omega)(\alpha\sigma_{xh}^2 + \beta)} \right)^g.$$

Using Taylor series approximation, we obtain the Bias in $\hat{A}_3$ as

$$\text{Bias}(\hat{A}_3) = \sum_{h=1}^{L} -W_h\theta_h \left[\frac{\sigma_{Th}^2 \bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1} - (g\omega\psi_h) \left(\frac{\sigma_{zh}^2(\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh}}{\sigma_{Th}^2 + 1} - \sigma_{xh}^2(\lambda_{04h} - 1) \right) \right], \tag{2.28}$$

where $\psi_h = \sum_{h=1}^{L} \frac{\alpha\sigma_{xh}^2}{\alpha\sigma_{xh}^2 + \beta}$.

The mean square error is given by

$$\begin{aligned} \text{MSE}(\hat{A}_3) &= \sum_{h=1}^{L} W_h^2 \theta_h \left[\left(\frac{\sigma_{zh}^4(\lambda_{40h} - 1) + 4\sigma_{Th}^4 \bar{Z}_h^4 C_{zh}^2 - 4\sigma_{zh}^2 \sigma_{Th}^2 \bar{Z}_h^2 \lambda_{30h} C_{zh}}{(\sigma_{Th}^2 + 1)^2} \right) \right. \\ &\quad \left. + \left((\sigma_{xh}^2 + Q_h\sigma_{yh}^2)^2 (\lambda_{04h} - 1) \right) - 2 \left(\frac{\sigma_{zh}^2(\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh}}{\sigma_{Th}^2 + 1} \right) (\sigma_{xh}^2 + Q_h\sigma_{yh}^2) \right], \end{aligned} \tag{2.29}$$

where $Q_h = g\omega\psi_h$.

Differentiate (2.29) w.r.t Q_h:

$$2\sigma_{yh}^2(\sigma_{xh}^2 + Q_h\sigma_{yh}^2)(\lambda_{04h} - 1) = 2\sigma_{yh}^2 \left(\frac{\sigma_{zh}^2(\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh}}{\sigma_{Th}^2 + 1} \right),$$

$$Q_{h\text{opt}} = \sum_{h=1}^{L} \frac{1}{\sigma_{yh}^2} \left[\left(\frac{\sigma_{zh}^2(\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh}}{\sigma_{Th}^2 + 1} \right) \left(\frac{1}{(\lambda_{04h} - 1)} \right) - \sigma_{xh}^2 \right].$$

The MSE at this optimum value is given by

$$\begin{aligned}\mathrm{MSE}(\hat{A}_3)_{\mathrm{opt}} = \sum_{h=1}^{L}\frac{W_h^2\theta_h}{(\sigma_{Th}^2+1)^2}\Bigg[&\Big(\sigma_{zh}^4(\lambda_{40h}-1)+4\sigma_{Th}^4\bar{Z}_h^4C_{zh}^2-4\sigma_{zh}^2\sigma_{Th}^2\bar{Z}_h^2\lambda_{30h}C_{zh}\Big)\\ &-\frac{1}{(\lambda_{04h}-1)}\Big(\sigma_{zh}^2(\lambda_{22h}-1)-2\sigma_{Th}^2\bar{Z}_h^2\lambda_{12h}C_{zh}\Big)^2\Bigg].\end{aligned} \tag{2.30}$$

Consider the second term:

$$\hat{B}_3 = \sum_{h=1}^{L} W_h\left[\Big(\bar{z}_h-\big[\bar{z}_{st}+(\bar{X}-\bar{x}_{st})\big]\Big) * \left(\frac{(\alpha\bar{X}+\beta)}{\lambda(\alpha\bar{x}_{st}+\beta)+(1-\lambda)(\alpha\bar{X}+\beta)}\right)^g\right]^2. \tag{2.31}$$

Repeating the procedure outlined in steps (2.10)–(2.15) for the estimator (2.31), yields definitions of Bias and MSE for $\hat{B}_3$ as

$$\begin{aligned}\mathrm{Bias}(\hat{B}_3) \approx\ & \bar{Z}^2\sum_{h=1}^{L}W_h^2\theta_h\Big(C_{zh}^2+D^2C_{xh}^2\Big)+\bar{X}^2\sum_{h=1}^{L}W_h^2C_{xh}^2+\sum_{h=1}^{L}W_h\theta_h\Bigg[\bar{Z}_h^2C_{zh}^2\\ &+\frac{2\bar{Z}_h}{\bar{Z}}\sum_{h=1}^{L}W_h\sigma_{zxh}+2D\Bigg(\frac{\bar{Z}_h}{\bar{X}}\sum_{h=1}^{L}W_h^2\sigma_{zxh}-\bar{Z}_h\bar{X}\sum_{h=1}^{L}W_h^2C_{xh}^2+\frac{\bar{Z}_h}{\bar{X}}\sum_{h=1}^{L}W_h\sigma_{zxh}-\frac{2\bar{Z}}{\bar{X}}\sum_{h=1}^{L}W_h^2\sigma_{zxh}\\ &+\bar{Z}\bar{X}\sum_{h=1}^{L}W_h^2C_{xh}^2+\sum_{h=1}^{L}W_h^2\sigma_{zxh}\Bigg)-2\bar{Z}_h\bar{Z}\sum_{h=1}^{L}W_h\sigma_{zh}^2-2\sum_{h=1}^{L}W_h^2\sigma_{zxh}\Bigg],\end{aligned} \tag{2.32}$$

where $D=(g\lambda\phi)$ and $\phi=\frac{\alpha\bar{X}}{\alpha\bar{X}+\beta}$;

$$\begin{aligned}\mathrm{MSE}(\hat{B}_3)_{\mathrm{opt}} \approx\ & \theta\Bigg\{\bar{Z}^2\bar{X}^2\sum_{h=1}^{L}W_h^2C_{xh}^2+\bar{Z}^4\sum_{h=1}^{L}W_h^2C_{zh}^2-\bar{Z}^2\sum_{h=1}^{L}W_h^2\sigma_{zxh}\\ &+D_{\mathrm{opt}}\Bigg[D_{\mathrm{opt}}\bar{Z}^4\sum_{h=1}^{L}W_h^2C_{xh}^2+\bar{Z}^3\bar{X}\sum_{h=1}^{L}W_h^2C_{xh}^2-\frac{\bar{Z}^3}{\bar{X}}\sum_{h=1}^{L}W_h^2\sigma_{zxh}\Bigg]\\ &+\sum_{h=1}^{L}W_h^2\Bigg[4\Bigg(\big(\bar{Z}_h^4-\bar{Z}_h^3\bar{Z}\big)C_{zh}^2+\bar{Z}_h^2\bar{Z}^2\Big(\sum_{h=1}^{L}W_h^2C_{zh}^2+C_{zh}^2+\sum_{h=1}^{L}W_h\sigma_{zh}^2\Big)\Bigg)\\ &+\big(\bar{Z}_h^2\bar{X}^2-\bar{Z}_h\bar{Z}\bar{X}^2\big)\sum_{h=1}^{L}W_h^2C_{xh}^2-\big(\bar{Z}_h^3\bar{Z}^2+\bar{Z}_h\bar{Z}^3\big)\sum_{h=1}^{L}W_h\sigma_{zh}^2\\ &+\frac{\big(\bar{Z}_h^3\bar{X}+\bar{Z}_h\bar{Z}^2\bar{X}\big)}{\bar{Z}\bar{X}}\sum_{h=1}^{L}W_h\sigma_{zxh}-2\bar{Z}_h^2\sum_{h=1}^{L}W_h\sigma_{zxh}+\bar{Z}_h\bar{Z}\sum_{h=1}^{L}W_h^2\sigma_{zxh}\\ &+D_{\mathrm{opt}}\Bigg(D_{\mathrm{opt}}\bar{Z}_h^2\bar{Z}^2\sum_{h=1}^{L}W_h^2C_{xh}^2+\frac{\big(2\bar{Z}_h^3-2\bar{Z}_h^2\bar{Z}\big)}{\bar{X}}\sum_{h=1}^{L}W_h\sigma_{zxh}\\ &+\big(\bar{Z}_h^2\bar{X}\bar{Z}-2\bar{Z}_h\bar{X}\bar{Z}^2-\bar{Z}_h\bar{Z}^3\big)\sum_{h=1}^{L}W_h^2C_{xh}^2+\frac{\bar{Z}_h\bar{Z}^2}{\bar{X}}\sum_{h=1}^{L}W_h^2\sigma_{zxh}\Bigg)\Bigg]\Bigg\}\end{aligned} \tag{2.33}$$

where

$$D_{\text{opt}} = \frac{-\sum_{h=1}^{L}\left[\frac{\left(2\bar{Z}_h^3-2\bar{Z}_h^2\bar{Z}\right)}{\bar{X}}\sum_{h=1}^{L} W_h\sigma_{zxh} + \bar{Z}^2\left(\bar{Z}_h\bar{X} - 2\bar{X}\bar{Z} - \bar{Z}^2\right)\sum_{h=1}^{L} W_h^2 C_{xh}^2 + \frac{\bar{Z}_h\bar{Z}^2}{\bar{X}}\sum_{h=1}^{L} W_h^2\sigma_{zxh}\right]}{2\left\{\bar{Z}_h^4\sum_{h=1}^{L} W_h^2 C_{xh}^2 + \sum_{h=1}^{L} W_h^2\left[\bar{Z}_h^2\bar{Z}^2\sum_{h=1}^{L} W_h^2 C_{xh}^2\right] + \left[\bar{Z}^3\bar{X}\sum_{h=1}^{L} W_h^2 C_{xh}^2 - \frac{\bar{Z}^4}{\bar{X}}\sum_{h=1}^{L} W_h^2\sigma_{zxh}\right]\right\}}.$$

The expressions for Bias and MSE of $t_{cp}(R)$ are given by

$$\text{Bias}(t_{cp}(R)) = \text{Bias}(\hat{A}_3) + \text{Bias}(\hat{B}_3), \tag{2.34}$$

and

$$\text{MSE}(t_{cp}(R))_{\text{opt}} \approx \text{MSE}(\hat{A}_3)_{\text{opt}} + \text{MSE}(\hat{B}_3)_{\text{opt}}. \tag{2.35}$$

3. SOME SEPARATE VARIANCE ESTIMATORS IN STRATIFIED RANDOM SAMPLING

Some authors, including Özel *et al.* (2014) [10], Clement (2018) [2] and Younis and Shabbir (2019) [16], have presented separate variance estimators. In doing so, they have ignored the B_1 term introduced in (1.2). We examine the following separate variance estimators in stratified random sampling mainly to show that ignoring the B_1 term can give misleadingly low MSE values.

3.1. The Separate Basic Variance Estimator

Following the authors listed above, the separate population variance of the study variable in stratified sampling is given by

$$\sigma_{s0}^2(R) = \sum_{h=1}^{L} W_h\left(\frac{\sigma_{zh}^2 - \sigma_{sh}^2 - \sigma_{Th}^2 * \bar{Z}_h^2}{\sigma_{Th}^2 + 1}\right). \tag{3.1}$$

This leads to the following estimator:

$$t_{s0}(R) = \sum_{h=1}^{L} W_h\left(\frac{s_{zh}^2 - \sigma_{sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1}\right). \tag{3.2}$$

The Bias and MSE of $t_{s0}(R)$ are given respectively as

$$\text{Bias}(t_{s0}(R)) \approx -\sum_{h=1}^{L}\theta_h W_h\left(\frac{\sigma_{Th}^2\bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1}\right), \tag{3.3}$$

and

$$\text{MSE}(t_{s0}(R)) \approx \sum_{h=1}^{L}\theta_h W_h^2\left(\frac{1}{(\sigma_{Th}^2+1)^2}\right)\left(\sigma_{zh}^4(\lambda_{40h}-1) + 4\sigma_{Th}^4\bar{Z}_h^4 C_{zh}^2 - 4\sigma_{zh}^2\sigma_{Th}^2\bar{Z}_h^2\lambda_{30h}C_{zh}\right). \tag{3.4}$$

3.2. The Separate Ratio Variance Estimator

$$t_{s1}(R) = \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) * \left(\frac{\sigma_{xh}^2}{s_{xh}^2} \right) \right]. \tag{3.5}$$

The Bias and MSE of $t_{s1}(R)$ are given respectively as

$$\text{Bias}(t_{s1}(R)) \approx \sum_{h=1}^{L} \theta_h W_h \left[\left(\frac{2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh} - \sigma_{zh}^2(\lambda_{22h} - 1) - \sigma_{Th}^2 \bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1} \right) \right], \tag{3.6}$$

and

$$\begin{aligned} \text{MSE}(t_{s1}(R)) \approx \sum_{h=1}^{L} \frac{W_h^2 \theta_h}{(\sigma_{Th}^2 + 1)^2} \Big[& \sigma_{zh}^4(\lambda_{40h} - 1) - 2\sigma_{zh}^2 \sigma_{yh}^2 (\lambda_{22h} - 1)(\sigma_{Th}^2 + 1) \\ & + \sigma_{yh}^4 (\lambda_{04h} - 1)(\sigma_{Th}^2 + 1)^2 \\ & + 4C_{zh} \left(\sigma_{Th}^4 \bar{Z}_h^4 C_{zh} - \sigma_{zh}^2 \sigma_{Th}^2 \bar{Z}_h^2 \lambda_{30h} + \sigma_{Th}^2 \sigma_{yh}^2 \bar{Z}_h^2 \lambda_{12h} (\sigma_{Th}^2 + 1) \right) \Big] \end{aligned} \tag{3.7}$$

3.3. A Separate Generalized Variance Estimator

The generalized separate population variance estimators can be written as

$$\begin{aligned} t_{sp}(R) = & \sum_{h=1}^{L} W_h \left[\left(\frac{s_{zh}^2 - \sigma_{sh}^2 - \sigma_{Th}^2 * \bar{z}_h^2}{\sigma_{Th}^2 + 1} \right) + (\sigma_{xh}^2 - s_{xh}^2) \right] \\ & * \left(\frac{(\alpha \sigma_{xh}^2 + \beta)}{\omega(\alpha s_{xh}^2 + \beta) + (1 - \omega)(\alpha \sigma_{xh}^2 + \beta)} \right)^g. \end{aligned} \tag{3.8}$$

The Bias and MSE of $t_{sp}(R)$ are given respectively as

$$\begin{aligned} \text{Bias}(t_{sp}(R)) = & \sum_{h=1}^{L} - W_h \theta_h \left[\frac{\sigma_{Th}^2 \bar{Z}_h^2 C_{zh}^2}{\sigma_{Th}^2 + 1} - (g\omega\psi_h) \right. \\ & \left. * \left(\frac{\sigma_{zh}^2(\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh}}{\sigma_{Th}^2 + 1} - \sigma_{xh}^2(\lambda_{04h} - 1) \right) \right], \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} \text{MSE}(t_{sp}(R))_{\text{opt}} = & \sum_{h=1}^{L} \frac{W_h^2 \theta_h}{(\sigma_{Th}^2 + 1)^2} \left[\left(\sigma_{zh}^4(\lambda_{40h} - 1) + 4\sigma_{Th}^4 \bar{Z}_h^4 C_{zh}^2 - 4\sigma_{zh}^2 \sigma_{Th}^2 \bar{Z}_h^2 \lambda_{30h} C_{zh} \right) \right. \\ & \left. - \frac{1}{(\lambda_{04h} - 1)} \left(\sigma_{zh}^2 (\lambda_{22h} - 1) - 2\sigma_{Th}^2 \bar{Z}_h^2 \lambda_{12h} C_{zh} \right)^2 \right]. \end{aligned} \tag{3.10}$$

4. SIMULATION STUDY

We consider a sample of size N = 2000 from two bivariate normal populations for $\begin{bmatrix} X \\ Y \end{bmatrix}$ determined by the following means and covariance matrices with $N_1 = 1200$ and $N_2 = 800$:

$$\text{Stratum 1:} \quad \mu = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 2.7 \\ 2.7 & 6 \end{bmatrix}, \quad \rho_{yx} = 0.80,$$

$$\text{Stratum 2:} \quad \mu = \begin{bmatrix} 6 \\ 4 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 2.2 \\ 2.2 & 5 \end{bmatrix}, \quad \rho_{yx} = 0.70.$$

These 2000 observations are treated as our finite populations. For the 2000 values generated from these distributions, the means, variances, and correlations are given by:

$$\text{Stratum 1:} \quad \mu_{x1} = 4.021, \ \mu_{y1} = 2.010, \ \sigma^2_{x1} = 1.975, \ \sigma^2_{y1} = 5.987, \ \rho_{yx1} = 0.797,$$

$$\text{Stratum 2:} \quad \mu_{x2} = 6.070, \ \mu_{y2} = 4.006, \ \sigma^2_{x2} = 1.982, \ \sigma^2_{y2} = 4.977, \ \rho_{yx2} = 0.702.$$

Overall parameter values are given by

$$\mu_x = 4.8413, \quad \mu_y = 2.8791, \quad \sigma^2_x = 2.9671, \quad \sigma^2_y = 6.4644, \quad \rho_{yx} = 0.7596.$$

We consider a sample of size $n = 600$, where $n_1 = 360$ and $n_2 = 240$. The stratum sample size n_h $(h = 1, 2)$ is based on the proportional allocation, that is, $n_h = W_h \times n$. The scrambling variable S and T are assumed to have normal distributions with $E(S) = 0$, $E(T) = 1$, $\text{Var}(S) = 0.5$ and different values for $\text{Var}(T)$. In the combined and separate generalized variance estimators, we choose $\alpha = 1$, $\beta = 0$ and $g = 1$. Other choices of α and β in our simulations had minimal impact.

The Percent Relative Efficiency (PRE) with respect to the stratified sampling is defined as

$$\text{PRE} = \frac{\text{MSE}(t_{c0}(R))}{\text{MSE}(t_{ci}(R))} \times 100, \ \text{where } i = 0, 1 \text{ and } p.$$

Since we are developing the proposed estimators based on randomized data, it is important to consider the privacy level as well. Gupta *et al.* (2018) [6] introduced a unified measure of estimator quality (δ) given by

$$\delta = \frac{\text{Theoretical MSE}}{\Delta_{DP}}, \quad \text{where } \Delta_{DP} = \sum_{h=1}^{L} W_h \Delta_{DPh}$$

is the privacy level for the model $Z{=}TY{+}S$ as given by Yan *et al.* (2009) [15].

Theoretical and empirical MSEs and PREs for both the separate variance estimators and combined variance estimators are reported in Table 1. For either separate or combined estimators, the generalized estimator is clearly more efficient than the basic estimator and the ratio estimator. One can note that the MSEs increase as the variances of T increase, which is on expected lines due to extra noise in the data. However, this loss in efficiency is off-set by the gain in privacy as shown by the δ-column. For example, the MSEs of the combined generalized variance estimator $t_{cp}(R)$ increases from 0.2842 to 0.5024 when $\text{Var}(T)$ increases from 0.5 to 1, but δ value decreases from 0.0364 to 0.0323. In general, the proposed variance estimators under the additive model ($Z = Y + S$) where $\mathrm{Var}(T) = 0$ are more efficient compared to the generalized model ($Z = TY + S$) where $\mathrm{Var}(T) > 0$ by providing smaller MSEs. However, the proposed variance estimators under the generalized model ($Z = TY + S$) are better by providing smaller δ values if we consider the efficiency and the privacy simultaneously.

Table 1: Theoretical (in **bold**) and empirical MSEs and PREs of the variance estimators with $\sigma_y^2 = 6.4644$.

Var(S)	Var(T)	Estimator	$\hat{\sigma_y^2}$	MSE	PRE	δ	Estimator	$\hat{\sigma_y^2}$	MSE	PRE	δ
0.5	0	$t_{s0}(R)$	5.4368	**0.0857** 0.0858	**100** 100	0.1664	$t_{c0}(R)$	6.4141	**0.1119** 0.1131	**100** 100	0.2144
		$t_{s1}(R)$	5.4542	**0.0752** 0.0765	**113.9627** 112.1568	0.1441	$t_{c1}(R)$	6.4322	**0.0985** 0.1014	**113.6040** 111.5384	0.1887
		$t_{sp}(R)$	5.4402	**0.0635** 0.0647	**134.9606** 132.6120	0.1368	$t_{cp}(R)$	6.4182	**0.0858** 0.0863	**130.4195** 131.0544	0.1644
	0.3	$t_{s0}(R)$	5.5456	**0.2142** 0.2151	**100** 100	0.0442	$t_{c0}(R)$	6.4434	**0.2588** 0.2606	**100** 100	0.0530
		$t_{s1}(R)$	5.5547	**0.1951** 0.1964	**109.8923** 109.5213	0.0400	$t_{c1}(R)$	6.4534	**0.2376** 0.2399	**108.9225** 108.6285	0.0486
		$t_{sp}(R)$	5.5443	**0.1757** 0.1769	**122.0261** 121.5941	0.0374	$t_{cp}(R)$	6.4429	**0.2169** 0.2174	**119.3176** 119.8712	0.0444
	0.5	$t_{s0}(R)$	5.5357	**0.2788** 0.2781	**100** 100	0.0357	$t_{c0}(R)$	6.4141	**0.3361** 0.3370	**100** 100	0.0431
		$t_{s1}(R)$	5.5446	**0.2579** 0.2596	**108.1039** 107.1263	0.0330	$t_{c1}(R)$	6.4239	**0.3147** 0.3158	**106.8001** 106.7131	0.0404
		$t_{sp}(R)$	5.5344	**0.2351** 0.2491	**118.5878** 111.6419	0.0306	$t_{cp}(R)$	6.4137	**0.2842** 0.2922	**118.2617** 114.9897	0.0364
	0.8	$t_{s0}(R)$	5.5128	**0.3801** 0.3766	**100** 100	0.0310	$t_{c0}(R)$	6.3931	**0.4560** 0.4528	**100** 100	0.0375
		$t_{s1}(R)$	5.5216	**0.3621** 0.3589	**104.9710** 104.9317	0.0294	$t_{c1}(R)$	6.4041	**0.4355** 0.4331	**104.7072** 104.5486	0.0358
		$t_{sp}(R)$	5.5115	**0.3406** 0.3399	**111.5971** 110.7972	0.0279	$t_{cp}(R)$	6.3981	**0.4047** 0.4141	**112.6760** 109.3455	0.0333
	1	$t_{s0}(R)$	5.4970	**0.4550** 0.4512	**100** 100	0.0299	$t_{c0}(R)$	6.3897	**0.5431** 0.5382	**100** 100	0.0360
		$t_{s1}(R)$	5.5057	**0.4399** 0.4372	**103.4325** 103.2021	0.0287	$t_{c1}(R)$	6.3929	**0.5235** 0.5202	**103.7440** 103.4602	0.0347
		$t_{sp}(R)$	5.4957	**0.4207** 0.4169	**108.1530** 108.2273	0.0262	$t_{cp}(R)$	6.3928	**0.5024** 0.4950	**108.1011** 108.7272	0.0323

Comparing the proposed separate variance estimators to the proposed combined variance estimators, it may appear that the separate estimators are better since they have smaller MSE. However, one should note that these estimators degrade accuracy in comparison to the combined estimators. For example, as the true variance of Y is 6.4644, the estimated variance of Y is 6.4137 for the combined variance estimator when $\mathrm{Var}(T) = 0.5$. However, the estimated variance of Y is 5.5344 for the separate variance estimator when $\mathrm{Var}(T) = 0.5$. The same is true for the other cases. This indicates that the proposed combined variance estimators are more accurate than the separate variance estimators.

5. APPLICATION

In this section a real data set is used to compare the performances of the combined variance estimators. The data is obtained from Eurostat (2008) [4], and the sampling details are provided in Sousa *et al.* (2014) [13], a paper that was co-authored by one of the co-authors of the current paper. There are 1698 records in the population. The volume of purchase orders reported by the Information and Communication Technologies for 2010 is taken as the study variable Y. Turnover for the individual enterprises is the auxiliary variable X. The study variable Y is scrambled using the additive scrambling variable S assumed to be a normally distributed random variable with mean 0 and variance 0.5, and the multiplicative scrambling variable T assumed to be a normally distributed random variable with mean 1 and four different choices for its variance (0, 0.3, 0.5, and 1). Data summary is provided in Table 2.

Table 2: Population Characteristics and Sampling Information.

Stratum	N	ρ_{yx}	μ_y	σ_y	μ_y	σ_x	Population
1	979	0.7802	2.15	2.46	3.12	2.68	$N = 1698,\ \rho_y x = 0.9368$
2	362	0.7952	16.67	6.86	20.31	6.02	$\mu_y = 14.44,\ \sigma_y^2 = 501.31$
3	357	0.8408	45.88	30.21	56.33	30.18	$\mu_x = 17.97,\ \sigma_x^2 = 640.59$

Table 3: Theoretical (in **bold**) and empirical MSEs and PREs of the variance estimators.

Var(S)	Var(T)	Estimator	$\hat{\sigma_y^2}$	MSE	PRE	δ
0.5	0	$t_{c0}(R)$	502.1499	**1230.678** 1228.833	**100** 100	2485.8154
		$t_{c1}(R)$	502.6576	**1079.764** 1086.812	**113.9765** 113.0676	2180.9880
		$t_{cp}(R)$	501.8341	**955.227** 947.223	**128.8361** 129.7300	1929.4389
	0.3	$t_{c0}(R)$	500.7685	**3615.078** 3656.872	**100** 100	15.8680
		$t_{c1}(R)$	501.557	**3483.887** 3457.51	**103.7656** 105.7660	15.2921
		$t_{cp}(R)$	503.1153	**3353.681** 3337.229	**111.1073** 112.9630	14.2817
	0.5	$t_{c0}(R)$	503.8318	**5389.539** 5353.592	**100** 100	14.2064
		$t_{c1}(R)$	501.3672	**5092.119** 5004.276	**105.8407** 106.9803	13.4224
		$t_{cp}(R)$	501.8327	**4878.576** 4841.631	**110.4736** 110.5741	12.8596
	1	$t_{c0}(R)$	501.3862	**9792.979** 9719.734	**100** 100	12.9152
		$t_{c1}(R)$	503.1442	**9540.721** 9574.217	**102.6440** 101.5198	12.5825
		$t_{cp}(R)$	500.9520	**9240.276** 9209.5	**105.9814** 105.5403	12.1863

Theoretical and empirical MSEs and PREs are provided in Table 3 for each of the proposed combined estimators. We used only the combined estimators in this numerical application because of the inherent drawback in the separate estimators as pointed out at the beginning of Section 3. The combined generalized variance estimator is clearly more efficient than both the combined basic variance estimator and the combined ratio variance estimator. Furthermore, the MSE increases as the variance of T is increased, meanwhile the unified measure (δ) value decreases. For example, for the combined generalized variance estimator, theoretical MSE is 3353.681 for $\sigma_T^2 = 0.3$ but increases to 9240.276 for $\sigma_T^2 = 1$. In contrast, the (δ) value decreases from 14.2817 to 12.1863 indicating that using the multiplicative noise T lowers the efficiency but the added privacy because of this more than compensates this loss.

6. CONCLUSION

Separate and combined variance estimators are considered under RRT in stratified random sampling. The simulation study shows that the generalized variance estimator is more efficient than the other estimators. Also, the proposed combined variance estimators are more accurate than the separate variance estimators. Furthermore, if one considers efficiency and privacy simultaneously, the linear combination model $Z = TY + S$, where $\mathrm{Var}(T) > 0$, produces better variance estimators compared to the additive model $Z = Y + S$ where $\mathrm{Var}(T) = 0$. This can be attributed to the fact that proposed variance estimators under $Z = TY + S$ have higher privacy level and hence smaller δ values. The real data application in Section 5 shows the same improvement with the generalized estimator as was seen in the simulation results of Section 4. We would like to mention that this work can be extended in several directions in new studies. For example, one can work with the case when the mean of the auxiliary variable is unknown. Also, the generalized estimator we suggest is not the only option. One can use other forms of generalizations.

ACKNOWLEDGMENTS

The authors appreciate the helpful comments by the two reviewers which helped improve the presentation of the paper.

REFERENCES

[1] Asghar, A.; Sanaullah, A. and Hanif, M. (2019). Generalized class of variance estimators under two-phase sampling for partial information case, *Electronic Journal of Applied Statistical Analysis*, **12**(1), 44–54.

[2] Clement, E.P. (2018). Improved Family of Ratio Estimators of Finite Population Variance in Stratified Random Sampling, *Biostat. Biometrics Open Acc. J.*, **5**(2), 555–659.

[3] Diana, G. and Perri, P.F. (2011). A class of estimators for quantitative sensitive data, *Statistical Papers*, **52**(3), 633–650.

[4] Eurostat (2008). NACE Rev.2 — statistical classification of economic activities in the european community, *Official Publications of the European Communities*, 112–285 and 306–311.

[5] Gupta, S.; Aloraini, B.; Qureshi, M. and Khalil, S. (2020). Variance Estimation using Randomized Response Technique, *REVSTAT*, **18**(2), 165–176.

[6] Gupta, S.; Samridhi, M.; Shabbir, J. and Khalil, S. (2018). A class of estimators for quantitative sensitive data, *Statistical Papers*, **52**(3), 633–650.

[7] Isaki, C.T. (1983). Variance estimation using auxiliary information, *Journal of the American Statistical Association*, **78**(381), 117–123.

[8] Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling, *Applied Mathematics and Computation*, **173**(2), 1047–1059.

[9] Kalucha, G.; Sousa, R.; Gupta, S. and Shabbir, J. (2017). Improved Ratio and Regression Estimators of the Mean of a Sensitive Variable in Stratified Sampling, *Statistics and Applications*, **15**(1-2), 63–78.

[10] Özel, G.; Çingi, H. and Oğuz, M. (2014). Separate Ratio Estimators for the Population Variance in Stratified Random Sampling, *Communications in Statistics – Theory and Methods*, **43**(22), 4766–4779. DOI: 10.1080/03610926.2012.729642

[11] Sanaullah, A.; Asghar, A. and Hanif, M. (2017). General class of exponential estimator for estimating finite population variance, *Journal of Reliability and Statistical Studies*, 1–16.

[12] Sidelel, E.B.; Orwa, G.O. and Otieno, R.O. (2014). Variance estimation in stratified random sampling in the presence of two auxiliary random variables, *International Journal of Science and Research*, **3**(9), 2453–2459.

[13] Sousa, R.; Gupta, S.; Shabbir, J. and Corte-Real, P. (2014). Improved mean estimation of a sensitive variable using auxiliary information in stratified sampling, *Journal of Statistics and Management Systems*, **17**(5-6), 503–518.

[14] Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, **60**(309), 63–69.

[15] Yan, Z.; Wang, J. and Lai, J. (2009). An efficiency and protection degree-based comparison among the quantitative randomized response strategies, *Communications in Statistics – Theory and Methods*, **38**, 404–408.

[16] Younis, F. and Shabbir, J. (2019). Estimation of General Parameter Under Stratified Adaptive Cluster Sampling Based on Dual Use of Auxiliary Information, *Scientia Iranica*. DOI: 10.24200/sci.2019.52515.2753

[17] Zahid, E. and Shabbir, J. (2018). Estimation of population mean in the presence of measurement error and non response under stratified random sampling, *PloS one*, **13**(2), e0191572.

[18] Zhang, Q.; Khalil, S. and Gupta, S. (2021). Estimation in the Simultaneous Presence of Measurement Errors and Non-response using Optional RRT Models under Stratified Sampling, *Journal of Statistical Computation and Simulation* (to appear).