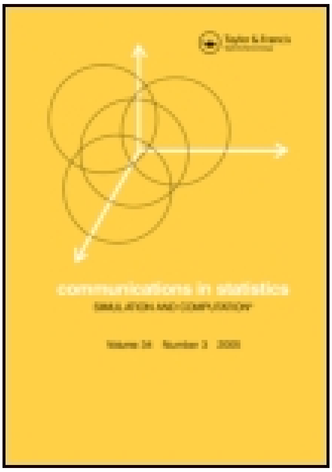

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Exponential-Type Estimators of the Mean of a Sensitive Variable in the Presence of Nonsensitive Auxiliary Information

Nursel Koyuncu [a], Sat Gupta [b] & Rita Sousa [c]

[a] Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey

[b] Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, North Carolina, USA

[c] New University of Lisbon & Statistics Portugal, Lisbon, Portugal

Accepted author version posted online: 15 Mar 2013.Published online: 14 Jan 2014.

Communications in Statistics—Simulation and Computation®, 43: 1583–1594, 2014

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2012.737492

Taylor & Francis
Taylor & Francis Group

Exponential-Type Estimators of the Mean of a Sensitive Variable in the Presence of Nonsensitive Auxiliary Information

NURSEL KOYUNCU,[1] SAT GUPTA,[2] AND RITA SOUSA[3]

[1]Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey
[2]Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, North Carolina, USA
[3]New University of Lisbon & Statistics Portugal, Lisbon, Portugal

Sousa et al. and Gupta et al. suggested ratio and regression-type estimators of the mean of a sensitive variable using nonsensitive auxiliary variable. This article proposes exponential-type estimators using one and two auxiliary variables to improve the efficiency of mean estimator based on a randomized response technique. The expressions for the mean squared errors (MSEs) and bias, up to first-order approximation, have been obtained. It is shown that the proposed exponential-type estimators are more efficient than the existing estimators. The gain in efficiency over the existing estimators has also been shown with a simulation study and by using real data.

Keywords Absolute relative bias; Exponential estimator; Mean squared error; Randomized response technique; Regression estimator.

Mathematics Subject Classification Primary 62D05.

1. Introduction

Randomized response technique (RRT) is used to estimate the proportion of people in a community bearing a stigmatizing characteristic such as habitual tax evasion, reckless driving, indiscriminate gambling, abortion, etc. In such situations, we cannot expect to get a truthful direct response to a sensitive question. Eichhron and Hayre (1983), Gupta and Shabbir (2004), Gupta et al. (2002, 2010), Wu et al. (2008), Saha (2008), Perri (2008), and many others have estimated the mean of a sensitive variable when the study variable is sensitive and there is no auxiliary variable. Sousa et al. (2010) and Gupta et al. (2012) suggested mean estimators based on RRT models using an auxiliary variable that can be directly observed. In sampling literature, Bahl and Tuteja (1991), Shabbir and Gupta (2007), Grover (2010), and Koyuncu (2012) have studied exponential-type estimators to get more efficient estimates. In this study, we have proposed exponential-type estimators of the mean of a sensitive variable using nonsensitive auxiliary information. We have discussed the cases when one or two nonsensitive auxiliary variables are available.

Received May 2, 2012; Accepted October 1, 2012
Address correspondence to Nursel Koyuncu, Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey; E-mail: nkoyuncu@hacettepe.edu.tr

Let Y be the study variable, a sensitive variable which cannot be observed directly. Let X_1 and X_2 be nonsensitive auxiliary variables that have a positive correlation with Y. Let S, be a scrambling variable, independent of Y,X_1, and X_2. The respondent is asked to report a scrambled response for Y given by $Z = Y + S$ but is asked to provide a true response for X_1 and X_2. Let a random sample of size n be drawn without replacement from a finite population $U = (U_1, U_2, \ldots, U_N)$. For the ith unit ($i = 1, 2, \ldots, N$), let y_i, x_{1i}, and x_{2i}, respectively, be the values of the study variable $Y$and auxiliary variables X_1 and X_2. Let $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, $\overline{x}_1 = \frac{\sum_{i=1}^{n} x_{1i}}{n}$, $\overline{x}_2 = \frac{\sum_{i=1}^{n} x_{1i}}{n}$, and $\overline{z} = \frac{\sum_{i=1}^{n} z_i}{n}$ be the sample means and $\overline{Y} = E(Y)$, $\overline{X}_1 = E(X_1)$, $\overline{X}_2 = E(X_2)$, and $\overline{Z} = E(Z)$ be the population means for Y, X_1, X_2, and Z, respectively. We assume that $\overline{X}_1$ and $\overline{X}_2$ are known and $\overline{S} = E(S) = 0$. Thus $E(Z) = E(Y)$ and $C_z^2 = C_y^2 + (S_s^2/\overline{Y}^2)$, where C_z and C_y are the coefficients of the variation of z and y, respectively.

To obtain the bias and mean squared error (MSE) expressions, let us define

$$e_0 = (\overline{z} - \overline{Z})/\overline{Z},\ e_1 = (\overline{x}_1 - \overline{X}_1)/\overline{X}_1,\ e_2 = (\overline{x}_2 - \overline{X}_2)/\overline{X}_2,$$
$$e_3 = \left(s_{x1}^2 - S_{x1}^2\right)/S_{x1}^2,\ e_4 = \left(s_{zx1}^2 - S_{zx1}^2\right)/S_{zx1}^2.$$

Using these notations,

$$E(e_i) = 0 \qquad i = 0, 1, 2, 3, 4.$$
$$E\left(e_0^2\right) = \lambda C_z^2,\ E\left(e_1^2\right) = \lambda C_{x1}^2,\ E\left(e_2^2\right) = \lambda C_{x2}^2,\ E(e_0 e_1) = \lambda C_{zx1},$$
$$E(e_0 e_2) = \lambda C_{zx2},\ E(e_1 e_2) = \lambda C_{x1x2},\ E(e_1 e_3) = \lambda \frac{1}{\overline{X}_1}\frac{\mu_{03}}{\mu_{02}},\ E(e_1 e_4) = \lambda \frac{1}{\overline{X}_1}\frac{\mu_{12}}{\mu_{11}},$$

where $\lambda = \frac{1-f}{n}$ and $\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \overline{Z})^r(x_{1i} - \overline{X}_1)^s$.

We describe below some existing mean estimators and their bias and MSE formulas.

(i) ***Ordinary sample mean ($\overline{z}$) of scrambled responses*:**

$$\hat{\mu}_y = \overline{z} \tag{1.1}$$
$$\text{MSE}(\hat{\mu}_y) = \lambda\left(S_y^2 + S_s^2\right) \tag{1.2}$$

where

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2, \quad S_{x1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{1i} - \overline{X}_1)^2,$$
$$S_s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(s_i - \overline{S})^2.$$

(ii) ***Sousa et al.'s (2010) ratio-type estimator*:**

$$\hat{\mu}_R = \overline{z}\frac{\overline{X}_1}{\overline{x}_1} \tag{1.3}$$
$$\text{Bias}(\hat{\mu}_R) \cong \overline{Y}\lambda\left(C_{x1}^2 - C_{x1z}\right), \tag{1.4}$$

where $C_z^2 = (C_y^2 + \frac{S_s^2}{\overline{Y}^2})$, $\rho_{zx1} = \frac{\rho_{yx1}}{\sqrt{1+\frac{S_s^2}{S_y^2}}}$, $\overline{Z} = \overline{Y}$.

$$\text{MSE}\left(\hat{\mu}_R\right) \cong \lambda \overline{Y}^2 \left(C_z^2 + C_{x1}^2 - 2C_{x1z}\right). \tag{1.5}$$

(iii) ***Gupta et al.'s (2012) regression estimator*:**

$$\hat{\mu}_{\text{Reg}} = \overline{z} + \hat{\beta}_{zx1}(\overline{X}_1 - \overline{x}_1), \tag{1.6}$$

where $\hat{\beta}_{zx1} = \frac{s_{zx1}}{s_{x1}^2} = \frac{s_{yx1}}{s_{x1}^2}$, sample regression coefficient between Z and X_1.

$$\text{Bias}\left(\hat{\mu}_{\text{Reg}}\right) \cong -\beta_{zx1}\lambda\left(\frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{03}}{\mu_{02}}\right), \tag{1.7}$$

where $\beta_{zx1} = \frac{S_{zx1}}{S_{x1}^2} = \frac{S_{yx1}}{S_{x1}^2} = \rho_{yx1}\frac{S_y}{S_{x1}} = \beta_{yx1}$ population regression coefficient between Z and X_1. Recognizing $\overline{Z} = \overline{Y}$,

$$\text{MSE}\left(\hat{\mu}_{\text{Reg}}\right) \cong \lambda \overline{Y}^2 C_z^2 \left[1 - \frac{S_{zx1}^2}{S_{x1}^2 S_z^2}\right] = \lambda \overline{Y}^2 C_z^2 \left[1 - \rho_{zx1}^2\right]$$

$$\text{MSE}\left(\hat{\mu}_{\text{Reg}}\right) \cong \lambda S_y^2 \left[\left(1 + \frac{S_s^2}{S_y^2}\right) - \rho_{yx1}^2\right]. \tag{1.8}$$

(iv) ***Gupta et al.'s (2012) generalized regression-cum-ratio estimator*:**

$$\hat{\mu}_{\text{GRR}} = \left[k_1\overline{z} + k_2(\overline{X} - \overline{x})\right]\left(\frac{\overline{X}}{\overline{x}}\right), \tag{1.9}$$

where k_1 and k_2 are constants.

$$\text{Bias}\left(\hat{\mu}_{\text{GRR}}\right) \cong (k_1 - 1)\overline{Y} + k_1\overline{Y}\lambda\left\{C_x^2 - \rho_{zx}C_zC_x\right\} + k_2\overline{X}\lambda C_x^2 \tag{1.10}$$

$$\begin{aligned}\text{MSE}\left(\hat{\mu}_{\text{GRR}}\right) \cong{}& (k_1 - 1)\overline{Y}^2 + k_1^2\overline{Y}^2\lambda\left\{C_z^2 + 3C_x^2 - 4\rho_{zx}C_zC_x\right\} \\ &+ k_2^2\overline{X}^2\lambda C_x^2 - 2k_1\overline{Y}^2\lambda\left\{C_x^2 - \rho_{zx}C_zC_x\right\} \\ &- 2k_2\overline{YX}\lambda C_x^2 - 2k_1k_2\overline{YX}\lambda\left\{\rho_{zx}C_zC_x - 2C_x^2\right\}. \end{aligned} \tag{1.11}$$

From Eq. (1.11), the optimum values of k_1 and k_2 are given by

$$k_{1(\text{opt})} = \frac{1 - \lambda C_x^2}{1 - \lambda\left\{C_x^2 - C_z^2\left(1 - \rho_{zx}^2\right)\right\}}, \tag{1.12}$$

$$k_{2(\text{opt})} = \frac{\overline{Y}}{\overline{X}}\left\{1 + k_{1(\text{opt})}\left(\frac{\rho_{zx}C_z}{C_x} - 2\right)\right\}; \tag{1.13}$$

the minimum MSE of $\hat{\mu}_{\text{GRR}}$ can be written as follows:

$$\text{MSE}\left(\hat{\mu}_{\text{GRR}}\right)_{\min} \cong \frac{\overline{Y}^2 C_z^2\left(1 - \rho_{zx}^2\right)\lambda\left\{1 - \lambda C_x^2\right\}}{C_z^2\left(1 - \rho_{zx}^2\right)\lambda + \left\{1 - \lambda C_x^2\right\}}. \tag{1.14}$$

2. Proposed Exponential-Type Estimator

Our first proposed estimator, which we call "generalized regression-cum-exponential estimator" follows Grover (2010) and Shabbir and Gupta (2007) and is given by

$$\hat{\mu}_{\exp 1} = [w_1 \bar{z} + w_2(\overline{X}_1 - \bar{x}_1)] \exp\left(\frac{\overline{X}_1 - \bar{x}_1}{\overline{X}_1 + \bar{x}_1}\right), \tag{2.1}$$

where w_1 and w_2 are suitable weights. Expressing (2.1) in terms of e's and retaining terms in e's up to second order, we have

$$\hat{\mu}_{\exp 1} - \overline{Z} \cong \left[w_1\overline{Z} + w_1\overline{Z}e_0 - w_2\overline{X}_1 e_1 - \frac{1}{2}w_1\overline{Z}e_1 - \frac{1}{2}w_1\overline{Z}e_0e_1 + \frac{1}{2}w_2\overline{X}_1e_1^2 + \frac{3}{8}w_1\overline{Z}e_1^2 - \overline{Z}\right]. \tag{2.2}$$

The bias and MSE of $\hat{\mu}_{\exp 1}$, to the first order of approximation, are given by

$$\text{Bias}\,(\hat{\mu}_{\exp 1}) \cong (w_1 - 1)\,\overline{Y} + \lambda\left\{\frac{1}{2}w_1\overline{Y}\left(\frac{3}{4}C_{x1}^2 - C_{zx1}\right) + \frac{1}{2}w_2\overline{X}_1C_{x1}^2\right\} \tag{2.3}$$

$$\begin{aligned} MSE\,(\hat{\mu}_{\exp 1}) \cong \big\{&\overline{Y}^2 + w_1^2\overline{Y}^2\left(1 + \lambda\left(C_z^2 + C_{x1}^2 - 2C_{zx1}\right)\right) \\ &+ w_2^2\overline{X}^2\lambda C_{x1}^2 + w_1\overline{Y}^2\left(\lambda\left(C_{zx1} - \frac{3}{4}C_{x1}^2\right) - 2\right) \\ &- w_2\overline{YX}\lambda C_{x1}^2 + 2w_1w_2\overline{YX}\lambda\left(C_{x1}^2 - C_{zx1}\right)\big\} \end{aligned} \tag{2.4}$$

and optimum values of w_1 and w_2, respectively, are found as

$$w_1^* = \frac{1 - \frac{1}{8}\lambda C_{x1}^2}{1 + \lambda C_z^2\left(1 - \rho_{zx1}^2\right)} \tag{2.5}$$

$$w_2^* = \frac{\overline{Y}}{2\overline{X}_1}\frac{C_{x1}^2 - 2C_{x1}^2 + 2C_{zx1} + \lambda C_{x1}^2\left(C_z^2\left(1 - \rho_{zx1}^2\right) + \frac{1}{4}\left(C_{x1}^2 - C_{zx1}\right)\right)}{C_{x1}^2\left[1 + \lambda C_z^2\left(1 - \rho_{zx1}^2\right)\right]}. \tag{2.6}$$

Substituting these optimum values in (2.4), the minimum MSE of $\hat{\mu}_{\exp 1}$ can be written as follows:

$$\text{MSE}_{\min}(\hat{\mu}_{\exp 1}) \cong \overline{Y}^2\left[1 - \frac{\lambda^2C_{x1}^2\left(\frac{1}{16}C_{x1}^2 + C_z^2\left(1 - \rho_{zx1}^2\right)\right) + 4}{4\left[1 + \lambda C_z^2\left(1 - \rho_{zx1}^2\right)\right]}\right] \tag{2.7}$$

$$\text{MSE}_{\min}(\hat{\mu}_{\exp 1}) \cong \left[\frac{\text{MSE}(\hat{\mu}_{\text{Reg}})}{\left[1 + \frac{\text{MSE}(\hat{\mu}_{\text{Reg}})}{\overline{Y}^2}\right]} - \frac{\lambda C_{x1}^2\left(\text{MSE}(\hat{\mu}_{\text{Reg}}) + \lambda\frac{1}{16}C_{x1}^2\overline{Y}^2\right)}{4\left[1 + \frac{\text{MSE}(\hat{\mu}_{\text{Reg}})}{\overline{Y}^2}\right]}\right]. \tag{2.8}$$

Note that the optimum choice of the constants w_1 and w_2 involves unknown parameters. These quantities can be guessed through a pilot sample survey or through experience gathered in due course of time, as mentioned in Upadhyaya and Singh (2006) and Koyuncu and Kadilar (2009).

The estimator defined in (2.1) can be generalized to the case of multiple auxiliary variables. We consider below the case of two auxiliary nonsensitive variables. This estimator is given by

$$\hat{\mu}_{\exp 2} = [d_1\bar{z} + d_2(\overline{X}_1 - \bar{x}_1) + d_3(\overline{X}_2 - \bar{x}_2)] \exp\left(\frac{(\overline{X}_1 - \bar{x}_1) + (\overline{X}_2 - \bar{x}_2)}{(\overline{X}_1 + \bar{x}_1) + (\overline{X}_2 + \bar{x}_2)}\right). \tag{2.9}$$

Expressing (2.9) in terms of e's and retaining up to second-order terms in e's, we have

$$\hat{\mu}_{\exp 2} = \left\{d_1\overline{Z}(1+e_0) - d_2\overline{X}_1 e_1 - d_3\overline{X}_2 e_2\right\}\left\{1 - \frac{\overline{X}_1}{2(\overline{X}_1+\overline{X}_2)}e_1 - \frac{\overline{X}_2}{2(\overline{X}_1+\overline{X}_2)}e_2 + \frac{3\overline{X}_1^2}{8(\overline{X}_1+\overline{X}_2)^2}e_1^2 + \frac{6\overline{X}_1\overline{X}_2}{8(\overline{X}_1+\overline{X}_2)^2}e_1e_2 + \frac{3\overline{X}_2^2}{8(\overline{X}_1+\overline{X}_2)^2}e_2^2\right\}. \tag{2.10}$$

The bias and MSE of $\hat{\mu}_{\exp 2}$, to the first order of approximation, are given by

$$\begin{aligned}\text{Bias}(\hat{\mu}_{\exp 2}) = \Bigg\{&(d_1-1)\overline{Z} + \frac{d_1\lambda\overline{Z}}{2(\overline{X}_1+\overline{X}_2)}\Bigg(-\overline{X}_1C_{zx1} - \overline{X}_2C_{zx2} + \frac{3\overline{X}_1^2}{4(\overline{X}_1+\overline{X}_2)}C_{x1}^2 \\ &+ \frac{3\overline{X}_2^2}{4(\overline{X}_1+\overline{X}_2)}C_{x2}^2 + \frac{3\overline{X}_1\overline{X}_2}{2(\overline{X}_1+\overline{X}_2)}C_{x1x2}\Bigg) \\ &+ \frac{d_2\lambda\overline{X}_1}{2(\overline{X}_1+\overline{X}_2)}\left(\overline{X}_1C_{x1}^2 + \overline{X}_2C_{x1x2}\right) + \frac{d_3\overline{X}_2}{2(\overline{X}_1+\overline{X}_2)}\lambda\left(\overline{X}_1C_{x1x2} + \overline{X}_2C_{x2}^2\right)\Bigg\}\end{aligned} \tag{2.11}$$

$$\begin{aligned}\text{MSE}(\hat{\mu}_{\exp 2}) = \big\{&\overline{Z}^2 + d_1A - d_2B - d_3C + d_1^2D + d_2^2\overline{X}_1^2\lambda C_{x1}^2 + d_3^2\overline{X}_2^2\lambda C_{x2}^2 \\ &+ 2d_1d_2F + 2d_1d_3G + 2d_2d_3\overline{X}_1\overline{X}_2\lambda C_{x1x2}\big\},\end{aligned} \tag{2.12}$$

where

$$\begin{aligned}A &= \overline{Z}^2\Bigg(-2 + \lambda\Bigg\{\frac{\overline{X}_1C_{zx1}}{(\overline{X}_1+\overline{X}_2)} + \frac{\overline{X}_2C_{zx2}}{(\overline{X}_1+\overline{X}_2)} - \frac{3\overline{X}_1^2C_{x1}^2}{4(\overline{X}_1+\overline{X}_2)^2} \\ &\quad - \frac{6\overline{X}_1\overline{X}_2C_{x1x2}}{4(\overline{X}_1+\overline{X}_2)^2} - \frac{3\overline{X}_2^2C_{x2}^2}{4(\overline{X}_1+\overline{X}_2)^2}\Bigg\}\Bigg), \\ B &= \lambda\frac{\overline{Z}}{(\overline{X}_1+\overline{X}_2)}\left(\overline{X}_1^2C_{x1}^2 + \overline{X}_1\overline{X}_2C_{x1x2}\right), \\ C &= \lambda\frac{\overline{Z}}{(\overline{X}_1+\overline{X}_2)}\left(\overline{X}_2^2C_{x2}^2 + \overline{X}_1\overline{X}_2C_{x1x2}\right), \\ D &= \Bigg(\overline{Z}^2 + \lambda\Bigg(\overline{Z}^2C_z^2 + \frac{\overline{X}_1^2\overline{Z}^2C_{x1}^2}{(\overline{X}_1+\overline{X}_2)^2} + \frac{\overline{X}_2^2\overline{Z}^2C_{x2}^2}{(\overline{X}_1+\overline{X}_2)^2} - 2\frac{\overline{X}_1\overline{Z}^2C_{zx1}}{(\overline{X}_1+\overline{X}_2)} \\ &\quad - 2\frac{\overline{X}_2\overline{Z}^2C_{zx2}}{(\overline{X}_1+\overline{X}_2)} + 2\frac{\overline{X}_1\overline{X}_2\overline{Z}^2C_{x1x2}}{(\overline{X}_1+\overline{X}_2)^2}\Bigg)\Bigg),\end{aligned}$$

$$F = \lambda\left(\frac{\overline{ZX}_1^2}{(\overline{X}_1+\overline{X}_2)}C_{x1}^2 + \frac{\overline{ZX}_1\overline{X}_2}{(\overline{X}_1+\overline{X}_2)}C_{x1x2} - \overline{ZX}_1 C_{zx1}\right),$$

$$G = \lambda\left(\frac{\overline{ZX}_2^2}{(\overline{X}_1+\overline{X}_2)}C_{x2}^2 + \frac{\overline{X}_1\overline{X}_2\overline{Z}}{(\overline{X}_1+\overline{X}_2)}C_{x1x2} - \overline{ZX}_2 C_{zx2}\right)$$

and optimum values of d_1, d_2, and d_3 are, respectively, found as

$$d_1^* = \frac{1}{2D}\frac{\begin{array}{l} A\,(D\lambda S_{x1x2}-FG)^2 + (BDG+CDF+2AFG)\,(D\lambda S_{x1x2}-FG) \\ -G\,(CD+AG)\left(D\lambda S_{x1}^2-F^2\right) - F\,(AF+BD)\left(D\lambda S_{x2}^2-G^2\right) \\ -A\left(D\lambda S_{x1}^2-F^2\right)\left(D\lambda S_{x2}^2-G^2\right)\end{array}}{\left(D\lambda S_{x1}^2-F^2\right)\left(D\lambda S_{x2}^2-G^2\right)-(D\lambda S_{x1x2}-FG)^2}, \tag{2.13}$$

$$d_2^* = \frac{1}{2}\frac{(AF+BD)\left(D\lambda S_{x2}^2-G^2\right)-(AG+CD)\,(D\lambda S_{x1x2}-FG)}{\left(D\lambda S_{x1}^2-F^2\right)\left(D\lambda S_{x2}^2-G^2\right)-(D\lambda S_{x1x2}-FG)^2}, \tag{2.14}$$

$$d_3^* = \frac{1}{2}\frac{(AG+CD)\left(D\lambda S_{x1}^2-F^2\right)-(AF+BD)\,(D\lambda S_{x1x2}-FG)}{\left(D\lambda S_{x1}^2-F^2\right)\left(D\lambda S_{x2}^2-G^2\right)-(D\lambda S_{x1x2}-FG)^2}. \tag{2.15}$$

Substituting these optimum values in (2.12), the minimum MSE of $\hat{\mu}_{\exp 2}$ can be written as follows:

$$\begin{aligned}&\mathrm{MSE}_{\min}(\hat{\mu}_{\exp 2}) \\ &= \overline{Z}^2 - \frac{A^2}{4D} - \frac{1}{4D}\frac{\begin{array}{l}(AG+CD)^2\left(D\lambda S_{x1}^2-F^2\right)+(AF+BD)^2\left(D\lambda S_{x2}^2-G^2\right) \\ -2\,(AG+CD)\,(AF+BD)\,(D\lambda S_{x1x2}-FG)\end{array}}{\left(D\lambda S_{x1}^2-F^2\right)\left(D\lambda S_{x2}^2-G^2\right)-(D\lambda S_{x1x2}-GF)^2}.\end{aligned} \tag{2.16}$$

3. Comparison with Gupta et al.'s (2012) Estimators

First, we compare the proposed generalized regression-cum-exponential estimator with the Gupta et al.'s (2012) regression estimator. Note that

$$\mathrm{MSE}_{\min}\left(\hat{\mu}_{\exp 1}\right) < \mathrm{MSE}\left(\hat{\mu}_{\mathrm{Reg}}\right) \text{ iff}$$

$$\mathrm{MSE}\left(\hat{\mu}_{\mathrm{Reg}}\right) - \frac{\mathrm{MSE}\left(\hat{\mu}_{\mathrm{Reg}}\right)}{\left[1+\frac{\mathrm{MSE}(\hat{\mu}_{\mathrm{Reg}})}{\overline{Y}^2}\right]} + \frac{\lambda C_x^2\left(\mathrm{MSE}\left(\hat{\mu}_{\mathrm{Reg}}\right)+\lambda\frac{1}{16}C_x^2\overline{Z}^2\right)}{4\left[1+\frac{\mathrm{MSE}(\hat{\mu}_{\mathrm{Reg}})}{\overline{Y}^2}\right]} > 0 \text{ iff}$$

$$\frac{\frac{(\mathrm{MSE}(\hat{\mu}_{\mathrm{Reg}}))^2}{\overline{Y}^2}}{\left[1+\frac{\mathrm{MSE}(\hat{\mu}_{\mathrm{Reg}})}{\overline{Y}^2}\right]} + \frac{\lambda C_x^2\left(\mathrm{MSE}\left(\hat{\mu}_{\mathrm{Reg}}\right)+\lambda\frac{1}{16}C_x^2\overline{Z}^2\right)}{4\left[1+\frac{\mathrm{MSE}(\hat{\mu}_{\mathrm{Reg}})}{\overline{Y}^2}\right]} > 0. \tag{3.1}$$

From (3.1), we can see easily that the proposed generalized regression-cum-exponential estimator is always more efficient than the regression estimator of Gupta et al. (2012).

Second, we compare the proposed generalized regression-cum-exponential estimator with the Gupta et al.'s (2012) generalized regression-cum-ratio estimator

$$\mathrm{MSE}_{\min}\left(\hat{\mu}_{\exp 1}\right) < \mathrm{MSE}\left(\hat{\mu}_{\mathrm{GRR}}\right)_{\min} \text{ iff}$$

$$\frac{\left(\frac{1-f}{n}\right)^2 C_x^2\left(\frac{1}{16}C_x^2 + C_z^2\left(1-\rho_{zx}^2\right)\right)+4}{4\left[1+\frac{1-f}{n}C_z^2\left(1-\rho_{zx}^2\right)\right]} + \frac{C_z^2\left(1-\rho_{zx}^2\right)\frac{1-f}{n}\left\{1-\frac{1-f}{n}C_x^2\right\}}{C_z^2\left(1-\rho_{zx}^2\right)\frac{1-f}{n}+\left\{1-\frac{1-f}{n}C_x^2\right\}} > 1. \quad (3.2)$$

When the condition (3.2) is satisfied, we can infer that the suggested estimator is more efficient than Gupta et al.'s (2012) generalized regression-cum-ratio estimator.

4. Simulation Studies

In this section, we investigate the efficiency of the proposed exponential estimators to existing estimators. The simulation study is carried out to compare the MSEs of the estimators both empirically and theoretically. In the simulation study, we consider two finite populations of size $N = 1000$ generated from a multivariate normal distribution with the same theoretical mean of $[Y, X_1, X_2]$ as $\mu = \begin{bmatrix} 5 & 5 & 5 \end{bmatrix}$ and different covariance matrices as given below.

Population 1:

$$\sigma^2 = \begin{bmatrix} 10 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix} \quad \rho_{x1y} = 0.6817, \quad \rho_{x2y} = 0.6705,$$

Population 2:

$$\sigma^2 = \begin{bmatrix} 6 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix} \quad \rho_{x1y} = 0.8706, \quad \rho_{x2y} = 0.8706.$$

The scrambling variable S is taken to be a normal distribution with mean equal to zero and standard deviation equal to 10% of the standard deviation of X_1. The reported response is given by $Z = Y + S$. For each population, we considered four sample sizes: $n = 50$, 100, 200, and 300. The percent relative efficiency (PRE) is calculated from the following equations:

$$\mathrm{PRE} = \frac{\mathrm{MSE}\left(\hat{\mu}_Y\right)}{\mathrm{MSE}\left(\hat{\mu}_i\right)} \times 100$$

where $i = R$, Reg, GRR, exp 1, exp 2.

The empirical MSE, theoretical MSE, and PRE values for all estimators are given in Tables 1 and 2.

From Tables 1 and 2, we can confirm that the suggested generalized regression-cum-exponential estimator is always more efficient than Gupta et al.'s (2012) regression estimator. Generalized regression-cum-exponential estimator is the most efficient estimator for using one auxiliary variable. The suggested exponential estimator with two auxiliary variables performs better than the estimator with one auxiliary variable, as expected.

Table 1
Mean squared error (MSE) and the percent relative efficiency (PRE) of all estimators

Population			MSE estimation		
N	ρ_{X1Y} ρ_{X2Y}	n	Empirical	Theoretical	PRE
			0.1993[a]	0.1953[a]	100.00[a]
		50	0.1193[b]	0.1145[b]	170.64[b]
			0.1083[c]	0.1047[c]	186.50[c]
			0.1094[d]	0.1043[d]	187.26[d]
			0.1089[e]	0.1042[e]	187.34[e]
			0.0857[f]	0.0827[f]	236.13[f]
			0.0900[a]	0.0925[a]	100.00[a]
		100	0.0544[b]	0.0542[b]	170.64[b]
			0.0499[c]	0.0496[c]	186.50[c]
			0.0503[d]	0.0495[d]	186.86[d]
			0.0501[e]	0.0495[e]	186.90[e]
1,000	**0.6817**		0.0390[f]	0.0393[f]	235.69[f]
	0.6705		0.0404[a]	0.0411[a]	100.00[a]
		200	0.0240[b]	0.0241[b]	170.64[b]
			0.0220[c]	0.0220[c]	186.50[c]
			0.0221[d]	0.0220[d]	186.66[d]
			0.0220[e]	0.0220[e]	186.67[e]
			0.0172[f]	0.0175[f]	235.47[f]
			0.0236[a]	0.0240[a]	100.00[a]
		300	0.0141[b]	0.0141[b]	170.64[b]
			0.0129[c]	0.0129[c]	186.50[c]
			0.0130[d]	0.0129[d]	186.59[d]
			0.0130[e]	0.0129[e]	186.60[e]
			0.0103[f]	0.0102[f]	235.40[f]

[a]Value for the ordinary mean.
[b]Value for the ratio estimator.
[c]Value for the regression estimator.
[d]Value for the generalized regression-cum-ratio estimator.
[e]Value for the generalized regression-cum-exponential estimator.
[f]Value for the exponential estimator with two auxiliary variables.

5. Numerical Example

We consider the real population used in Sousa et al. (2010) and in Gupta et al. (2012). It is based on the survey on Information and Communication Technologies (ICT) usage in enterprises in 2009 with seat in Portugal (Smilhily and Storm, 2010). This survey intends to promote the development of the national statistical system in the information society and to contribute to a deeper knowledge about the usage of the ICT by enterprises. The target population covers all industries with one and more persons employed in the sections of economic activity C to N and S, from NACE (Statistics of Economic Activities in the European Community) Rev. 2 (Eurostat, 2008). The data are essentially collected using Electronic Data Interchange, applying direct connection between information systems

Table 2
Mean squared error (MSE) and the percent relative efficiency (PRE) of all estimators

Population		MSE estimation			
N	ρ_{X1Y} ρ_{X2Y}	n	Empirical	Theoretical	PRE
			0.1198[a]	0.1181[a]	100.00[a]
		50	0.0400[b]	0.0395[b]	299.42[b]
			0.0287[c]	0.0289[c]	409.40[c]
			0.0291[d]	0.0288[d]	409.86[d]
			0.0289[e]	0.0288[e]	410.03[e]
			0.0078[f]	0.0073[f]	1626.30[f]
			0.0547[a]	0.0560[a]	100.00[a]
		100	0.0188[b]	0.0187[b]	299.42[b]
			0.0138[c]	0.0137[c]	409.40[c]
			0.0140[d]	0.0137[d]	409.62[d]
			0.0139[e]	0.0137[e]	409.70[e]
1,000	**0.8706**		0.0037[f]	0.0034[f]	1625.73[f]
	0.8428		0.0246[a]	0.0249[a]	100.00[a]
		200	0.0086[b]	0.0083[b]	299.42[b]
			0.0063[c]	0.0061[c]	409.40[c]
			0.0063[d]	0.0061[d]	409.49[d]
			0.0063[e]	0.0061[e]	409.53[e]
			0.0017[f]	0.0015[f]	1625.44[f]
			0.0143[a]	0.0145[a]	100.00[a]
		300	0.0049[b]	0.0048[b]	299.42[b]
			0.0037[c]	0.0035[c]	409.40[c]
			0.0037[d]	0.0035[d]	409.45[d]
			0.0037[e]	0.0035[e]	409.47[e]
			0.0011[f]	0.0009[f]	1625.35[f]

[a]Value for the ordinary mean.
[b]Value for the ratio estimator.
[c]Value for the regression estimator.
[d]Value for the generalized regression-cum-ratio estimator.
[e]Value for the generalized regression-cum-exponential estimator.
[f]Value for the exponential estimator with two auxiliary variables.

at the respondent and the National Statistics Institute. For some enterprises, the article questionnaire is still used. The questions in the structural business surveys mainly deal with characteristics that can be found in the organizations' annual reports and financial statements, such as employment, turnover, and investment.

In our application, the study variable Y is the purchase orders in 2010, collected by the ICT survey in that year. This is typically a confidential variable for enterprises, only known from business surveys. The auxiliary variable X is the turnover of each enterprise. This information can be easily obtained from enterprise records available in the public domain, as administrative information. In 2010, the population survey contained approximately 278,000 enterprises and we know the value of X for all these enterprises. The purchase orders information was collected in the ICT survey and we have the values of Y for 5,336

Table 3
Mean squared error (MSE) and the percent relative efficiency (PRE) for all estimators

Population			MSE estimation		
N	ρ_{XY}	n	Empirical	Theoretical	
		100	196.8002[a]	191.5088[a]	100.00[a]
			11.3683[b]	16.4393[b]	1164.94[b]
			16.7963[c]	16.3768[c]	1169.39[c]
			11.3909[d]	15.6644[d]	1222.58[d]
			12.3339[e]	13.1849[e]	1452.49[e]
		500	34.6507[a]	35.3757[a]	100.00[a]
			2.7259[b]	3.0367[b]	1164.94[b]
			3.0170[c]	3.0252[c]	1169.39[c]
			2.7509[d]	3.0069[d]	1176.50[d]
5336	0.9636		2.8631[e]	2.9173[e]	1212.61[e]
		1000	15.7543[a]	15.8591[a]	100.00[a]
			1.3092[b]	1.3614[b]	1164.94[b]
			1.3592[c]	1.3562[c]	1169.39[c]
			1.3175[d]	1.3526[d]	1172.47[d]
			1.3381[e]	1.3345[e]	1188.38[e]
		2000	6.2451[a]	6.1008[a]	100.00[a]
			0.5691[b]	0.5237[b]	1164.94[b]
			0.5573[c]	0.5217[c]	1169.39[c]
			0.5718[d]	0.5212[d]	1170.55[d]
			0.5673[e]	0.5185[e]	1176.62[e]

[a]Value for the ordinary mean.
[b]Value for the ratio estimator.
[c]Value for the regression estimator.
[d]Value for the generalized regression-cum-ratio estimator.
[e]Value for the generalized regression-cum-exponential estimator.

enterprises (which answered this question in the ICT survey in 2010). For this study, these 5,336 enterprises are considered as our population. The scrambling variable S is taken to be a normal random variable with mean equal to zero and standard deviation equal to 10% of the standard deviation of X, that is, $\sigma_S = 0.1\sigma_x$. The reported response is given by $Z = Y+S$ (the purchase order value plus a random quantity). The variables Y and X are strongly correlated, so we can take advantage of this correlation by using the ratio and regression estimators.

Population Characteristics:

$$N = 5336,\ \rho_{xy} = 0.9632$$
$$\mu_x = 22.99,\ \mu_Y = 30.19,\ \sigma_x = 172.09,\ \sigma_y = 138.65$$

and $\beta_{YX} = 0.7763$. We use the following sample sizes in our simulation study: n = 100, 500, 1000, and 2000.

The empirical, theoretical MSE and PRE values for all estimators are given in Table 3. From Table 3, we can say that the generalized regression-cum-exponential estimator has the largest PRE.

6. Conclusion

This article proposed exponential-type estimators using one or two nonsensitive auxiliary variables to improve the efficiency of RRT estimators of mean. The expression for bias and MSE are derived. We found that the proposed exponential-type estimators are more efficient than the existing estimators in literature. This results are also supported with a simulation study and using a real data.

Acknowledgments

The authors would like to thank the two anonymous referees for their constructive comments and suggestions.

References

Bahl, S., Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Information and Optimization Sciences* 12(1):159–163.

Eichhron, B. H., Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference* 7:307–316.

Eurostat. (2008). *Nace Rev. 2—Statistical classification of economic activities in European Community*. Official Publications of the European Communities. Working Paper. Luxembourg City, Luxembourg, pp. 112–285 and 306–311.

Grover, L. K. (2010). A correction note on improvement in variance estimation using auxiliary information. *Communications in Statistics- Theory and Methods* 39:753–764.

Gupta, S., Shabbir, J. (2004). Sensitivity estimation for personal interview survey questions. *Statistica* 64:643–653.

Gupta, S., Shabbir, J., Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference* 140(10):2870–2874.

Gupta, S., Shabbir, J., Sousa, R., Real, P. C. (2012). Estimation of the mean of a sensitive variable in the presence of auxiliary information. *Communications in Statistics—Theory and Methods* 41:1–12.

Gupta, S. N., Gupta, B. C., Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference* 100:239–247.

Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes. *Applied Mathematics and Computation* 218:10900–10905.

Koyuncu, N., Kadilar, C. (2009). Family of estimators of population mean using two auxiliary variables in stratified random sampling. *Communications in Statistics: Theory and Methods* 38(14):2398–2417.

Perri, P. F. (2008). Modified randomized devices for Simmons' model. *Model Assisted Statistics and Applications* 3:233–239.

Saha, A. (2008). A randomized response technique for quantitative data under unequal probability sampling. *Journal of Statistical Theory and Practice* 2(4):589–596.

Shabbir, J., Gupta, S. (2007). On improvement in variance estimation using auxiliary information. *Communications in Statistics—Theory and Methods* 36(12):2177–2185.

Smilhily, M., Storm, H. (2010). *ITC usage in enterprises 1—2009*. Luxembourg City, Luxembourg: Eurostat Publications, Issue 1.

Sousa, R., Shabbir, J., Real, P. C., Gupta, S. (2010). Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistical Theory and Practice* 4(3):495–507.

Upadhyaya, L. N., Singh, H. P. (2006). Almost unbiased ratio and product-type estimators of finite population variance in sample surveys. *Statistics in Transition* 7(5):1087–1096.

Wu, J.-W., Tian, G.-L., Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: Design and analysis. *Metrika* 67:251–263.