International Journal of Statistics and Analysis.
ISSN 2248-9959 Volume 8, Number 1 (2018), pp. 1-8

http://www.ripublication.com

# A Class of Estimators of a General Parameter of a Finite Population with Auxiliary Information on Two Variables

**P. A. Patel[1] and F. H. Shah[2*]**

[1*]*Department of Statistics, Sardar Patel University, Vallabh Vidyanagar Gujarat, India.*

[2]*Department of Statistics, Sardar Patel University, Vallabh Vidyanagar Gujarat, India.*

**Corresponding author*

## ABTRACT

A class of estimators is considered for general parameter of a finite population using auxiliary variable at estimation stage. The lower bound for the asymptotic variance of the class is shown to be equal to the asymptotic variance of the linear regression estimator of the general parameter. As special cases, various classes of estimators of linear and non-linear parameters are discussed. A small scale Monte Carlo simulation is carried out to study the performance various estimators of linear and non-linear parameters.

**Keywords:** Auxiliary variable, Finite population, Asymptotic variance, Regression estimator, Monte Carlo Simulation

## 1. INTRODUCTION

Most of the emphasis has been placed on the problem of estimating a finite population mean or variance of the survey variable in presence of auxiliary information. The estimation of coefficient of variation (C.V.), correlation coefficient, regression coefficient, variance of a linear estimator, etc., are also important. Das and Tripathi [1, 2], Patel and Shah [3, 4] suggested various estimators of coefficient of variance. Patel and Chaudhari [5, 6], Patel and Patel [7] addressed the problem of estimation of variance of the Horvitz-Thompson estimator of the population total. Two well-known regression estimators have appeared in the literature, namely the generalized regression (GREG) estimator (see, Särndal et al., [8]) and the optimal (OPT) estimator (see, Montanari, [9]) for estimating finite population means of survey variables. There are a number of ways to construct a GREG estimator of the population mean, see, e.g., Fuller [10]. Shah [11] discussed various approaches for estimation of the population regression coefficient and consequently suggested many regression estimators of the population mean.

In this article we consider a problem of estimation of a general parameter that can be used to estimate various parameters of the finite population incorporating information on two auxiliary variables. A class of estimators of the general parameter is suggested and the lower bound of the variance of this estimator is obtained in sections 2 and 3. A special case is discussed and numerical study is presented in section 4. Section 5 states the conclusion for the study.

## 2. THE CLASS OF ESTIMATORS

Consider a finite population $U = \{1, \dots, i, \dots, N\}$ of $N$ elements and assume that $y$ to be a study variable and variables $x$ and $z$ to be known auxiliary variables. Let a population function $G_y$ of $y$ be a parameter of interest. For instance, $G_y$ may be a population linear function (mean or total), a quadratic function (variance of an estimator of a linear function or population variance), or a non-linear function (coefficient of variation, correlation coefficient, regression coefficient). We seek to estimate $G_y$, incorporating auxiliary information on $x$ and $z$, using a probability sample $s$ of fixed-size $n$ drawn from U with any sampling design $p$

Let $g_y, g_x \; and \; g_z$ denote any unbiased/consistent estimators of $G_y, G_x$ and $G_z$, respectively, for any sampling design and let $(g_y, g_x, g_z, G_x G_z)$ assume values in a closed convex subset $\mathbb{R}^5$, of the five dimensional real space containing the point $(G_y, G_x, G_z, G_x G_z)$. Here $G_x$ and $G_z$ are assumed to be known. For example, $\overline{y},\ \overline{x}$ and $\overline{z}$ are consistent estimators for $\overline{Y},\ \overline{X}$ and $\overline{Z}$, where $\overline{X}$ and $\overline{Z}$ are known. Let $h(g_y, g_x, g_z, G_x, G_z)$ be a continuous function of $g_y, g_x, g_z, G_x, G_z$ having continuous first and second derivatives which are bounded in $\mathbb{R}^5$ and such that $h(G_y, G_x, G_z, G_x, G_z) = G_y$ for all $G_y$. We consider the class of estimators $\hat{G}_y$ defined by

$$\hat{G}_y = h\big(g_y, g_x, g_z, G_x, G_z\big) \tag{1}$$

For instance, $\hat{G}_y = \overline{y}\big(\overline{X}/\overline{x}\big)\big(\overline{Z}/\overline{z}\big)$ is an estimator of $G_y = \overline{Y}\big(\overline{X}/\overline{X}\big)\big(\overline{Z}/\overline{X}\big) = \overline{Y}$. Denote the first partial derivatives of $h\big(g_y, g_x, g_z, G_x, G_z\big)$ with respect to $g_y,\ g_x$ and $g_z$ at the point $(G_y, G_x, G_z, G_x, G_z)$ by $h'_y\big(G_y, G_x, G_z, G_x, G_z\big)$, $h'_x(G_y, G_x, G_z, G_x, G_z)$ and $h'_z(G_y, G_x, G_z, G_x, G_z)$, respectively. Expanding $h\big(g_y, g_x, g_z, G_x, G_z\big)$ about the point $(G_y, G_x, G_z, G_x, G_z)$ in a second order Taylor's series and noting that $h'_y\big(G_y, G_x, G_z, G_x, G_z\big) = 1$ we obtain

$$\begin{aligned}\hat{G}_y = G_y + h'_y\big(G_y, G_x, G_z, G_x, G_z\big)\big(g_y - G_y\big) + h'_x(G_y, G_x, G_z, G_x, G_z)(g_x - G_x) \\ + h'_z(G_y, G_x, G_z, G_x, G_z)(g_z - G_z) + O(n^{-1})\end{aligned} \tag{2}$$

From (2) it is found that the bias $E(\hat{G}_y - G_y)$ of $\hat{G}_y$ is of the order $n^{-1}$ and its asymptotic variance as

$$V(\hat{G}_y) = V_{yy} + h_x'^2(G_y, G_x, G_z, G_x, G_z)V_{xx} + h_z'^2(G_y, G_x, G_z, G_x, G_z)V_{zz}$$
$$+2h_x'\big(G_y, G_x, G_z, G_x, G_z\big)V_{yx} + 2h_z'\big(G_y, G_x, G_z, G_x, G_z\big)V_{yz}$$
$$+2h_x'(G_y, G_x, G_z, G_x, G_z)h_z'(G_y, G_x, G_z, G_x, G_z)V_{xz} \qquad (3)$$

where $V_{uv}$ denotes the covariance between $u\ and\ v;\ u, v = x, y, z.$

The variance of $\hat{G}_y$, given at (3), is minimized for

$$h_x'\big(G_y, G_x, G_z, G_x, G_z\big) = \frac{V_{yz}V_{xz} - V_{zz}V_{yx}}{V_{xx}V_{zz} - {V_{xz}}^2} = -\,\gamma_x, \qquad \text{say,}$$

$$h_z'\big(G_y, G_x, G_z, G_x, G_z\big) = \frac{V_{yx}V_{xz} - V_{xx}V_{yz}}{V_{xx}V_{zz} - {V_{xz}}^2} = -\,\gamma_z, \qquad \text{say}$$

Here, $\gamma_x$ and $\gamma_z$ can further be simplified as

$$\gamma_x = \sqrt{\frac{V_{yy}}{V_{xx}}}\ \rho_x^* \quad \text{and} \quad \gamma_z = \sqrt{\frac{V_{yy}}{V_{zz}}}\rho_z^*$$

where

$$\rho_x^* = \frac{\big(\rho_{yx}^* - \rho_{yz}^*\rho_{xz}^*\big)}{1 - \rho_{xz}^{*2}} \quad \text{and} \quad \rho_z^* = \frac{\big(\rho_{yz}^* - \rho_{yx}^*\,\rho_{xz}^*\big)}{1 - \rho_{xz}^{*2}}$$

with

$$\rho_{uv}^* = \frac{V_{uv}}{(V_{uu}V_{vv})^{1/2}}$$

Inserting $h_x'\big(G_y, G_x, G_z, G_x, G_z\big) = -\rho_x^*\sqrt{V_{yy}/V_{xx}}$ and $h_z'\big(G_y, G_x, G_z, G_x, G_z\big) = -\rho_z^*\sqrt{V_{yy}/V_{xx}}$ in (2.3) we obtain the minimum asymptotic variance (i.e., lower bound of the variance) as

$$minV\big(\hat{G}_y\big) = V_{yy}\big[1 + \rho_x^{*2} + \rho_z^{*2} - 2\rho_{yx}^*\rho_x^* - 2\rho_{yz}^*\rho_z^* + 2\rho_x^*\rho_y^*\rho_{xz}^*\big] \qquad (4)$$

**Remark 1.** The minimum variance of $\hat{G}_y$, given in (4), is identical to the variance of the linear regression estimator

$$\hat{G}_{yreg} = g_y + \beta_x(G_x - g_x) + \beta_z(G_z - g_z)$$

where $\beta_x$ and $\beta_z$ are known regression coefficients.

In next section we illustrate above theory of estimation for the population mean under simple random sampling without replacement (SRSWOR).

## 3. ESTIMATION OF THE POPULATION MEAN

Consider the estimating function as the population mean

$$G_y = \bar{Y} \qquad (5)$$

The class of estimators given in (1) reduces to $\hat{G}_y = h(\bar{y},\bar{x},\bar{z},\bar{X},\bar{Z})$.

Many estimators of $\bar{Y}$ are available in the literature from which some of them are listed below.

Mohanty [12] has made ratio-adjustment to the regression estimator and suggested the estimator

$$t_M = [\bar{y} + \beta_{yx}(\bar{X} - \bar{x})]\frac{\bar{Z}}{\bar{z}} \tag{6}$$

where the optimum value of $\beta_{yx}$ is $opt\beta_{yx} = \bar{Y}C_y\rho_{yx}/\bar{X}C_x$. The minimum variance of $t_M$ is

$$V(t_M) \approx \theta\bar{Y}^2\left[C_y^2(1 - \rho_{yx}^2 - \rho_{yz}^2) + (C_z - C_y\rho_{yz})^2 + 2C_yC_z\rho_{yx}\rho_{xz}\right] \tag{7}$$

where $\theta = n^{-1} - N^{-1}$.

Samiuddin and Hanif [13] have suggested three estimators of $\bar{Y}$:

(1) A chain ratio estimator

$$t_{SH(1)} = \bar{y}\frac{\bar{X}}{\bar{x}}\frac{\bar{Z}}{\bar{z}} \tag{8}$$

with

$$V(t_{SH(1)}) \approx \theta\bar{Y}^2[C_y^2 + C_x^2 + C_z^2 - 2\rho_{yx}C_yC_x - 2\rho_{yz}C_yC_z - 2\rho_{xz}C_xC_z] \tag{9}$$

(2) A composite estimator constructed using two ratio estimators

$$t_{SH(2)} = a\left(\bar{y}\frac{\bar{X}}{\bar{x}}\right) + (1 - a)\left(\bar{y}\frac{\bar{Z}}{\bar{z}}\right) \tag{10}$$

The minimum variance of $t_{SH(2)}$ is

$$V(t_{SH(2)}) \approx \theta\bar{Y}^2\left[(C_y^2 + C_z^2 - 2\rho_{yz}C_yC_z) - \frac{(\rho_{yx}C_yC_x - \rho_{yz}C_yC_z - \rho_{xz}C_xC_z + C_z^2)^2}{C_x^2 + C_z^2 - 2\rho_{xz}C_xC_z}\right] \tag{11}$$

(3) A composite estimator constructed using two regression estimators

$$t_{SH(3)} = a\left([\bar{y} + \beta_{yx}(\bar{X} - \bar{x})]\frac{\bar{Z}}{\bar{z}}\right) + (1 - a)\left([\bar{y} + \beta_{yz}(\bar{Z} - \bar{z})]\frac{\bar{X}}{\bar{x}}\right) \tag{12}$$

The minimum variance of $t_{SH(2)}$ is

$$V\left(t_{SH(3)}\right) \approx \theta \overline{Y}^2 C_y^2 \left[1 - \frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{1 - \rho_{xz}^2}\right] \tag{13}$$

A regression estimator

$$t_{reg} = \bar{y} + b_x(\bar{X} - \bar{x}) + b_z(\bar{Z} - \bar{z}) \tag{14}$$

where $b_x = s_{yx}/s_x^2$ and $b_z = s_{yz}/s_z^2$. Its approximate variance can be obtained as

$$V\left(t_{reg}\right) \approx \theta \overline{Y}^2 C_y^2 \left[1 - \rho_{yx}^2 - \rho_{yz}^2 + 2\rho_{yx}\rho_{yz}\rho_{xz}\right] \tag{15}$$

Motivated from Abu Dayyeh et al. [14], Kadilar and Cingi [15] proposed the estimator

$$t_{KC} = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha 1}\left(\frac{\bar{Y}}{\bar{y}}\right)^{\alpha 2} + b_x(\bar{X} - \bar{x}) + b_z(\bar{Z} - \bar{z}) \tag{16}$$

where $\alpha_1$ and $\alpha_2$ are constants to be determined such that $t_{KC}$ has minimum variance. They derived the minimum variance as

$$minMSE(t_{KC}) \approx \theta \overline{Y}^2 C_y^2 \left\{1 + C_1^2 + C_2^2 + 2\rho_{xz}C_1C_2 - 2\rho_{yx}C_1 - 2\rho_{yz}C_2\right\} \tag{17}$$

where

$$C_1 = \rho_{yx} + \frac{\rho_{xz}\left(\rho_{yx}\rho_{xz} - \rho_{yz}\right)}{1 - \rho_{xz}^2} \quad \text{and} \quad C_2 = \rho_{yz} + \frac{\rho_{xz}\left(\rho_{yz}\rho_{xz} - \rho_{yx}\right)}{1 - \rho_{xz}^2}$$

Upon simplification (17) reduces to (8).

**Remark 1.** The estimators given in this section are member of the class $\hat{G}_y = h\left(\bar{y}, \bar{x}, \bar{z}, \bar{X}, \bar{Z}\right)$. Under SRSWOR the minimum variance of this class, given in (4), is simplified as

$$V\left(\hat{G}_y\right) \approx \theta \overline{Y}^2 C_y^2 \left[1 - \frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{1 - \rho_{xz}^2}\right] \tag{18}$$

**Remark 2.** The estimators $t_{SH(3)}$ and $t_{KC}$ have attained the lower bound of the variance (18).

**Remark 3.** The theory discussed in Section 2 can be applied to estimate the population variance, coefficient of variation, coefficient of correlation, coefficient of regression, variance of the Horvitz-Thompson estimator etc., on the line of Section 3.

## 4. NUMERICAL COMPARISON

To check whether the variance of each of the estimators given in (6), (8) – (10), (12), (13) and (14) attain the lower bound or not we calculate their variances using the population data on *y* = apple production amount in 1999, *x* = number of apple trees in

1999 and $z$ = apple production amount in 1998 of 204 villages in Black Sea Region of turkey (source: Institute of statistics, Republic of Turkey).

| size | Mean | Variance/covariance | Correlations |
|---|---|---|---|
| N = 250 | $\bar{Y} = 966$ | $\bar{X} = 26441$ | $\bar{Z} = 1014$ |
| n = 50 | $S_y = 2389.76$ | $S_x = 45402.7$ | $S_z = 2521.40$ |
| | $S_{xy} = 77372777$ | $S_{yz} = 5684276$ | $S_{xz} = 94636084$ |
| | $\rho^*_{yx} = 0.7162$ | $\rho^*_{yz} = 0.9427$ | $\rho^*_{xz} = 0.8283$ |
| | $\rho^*_1 = -0.9223$ | $\rho^*_2 = 0.170$ | |

**Table 1** The approximate variances

| Estimator | Variance | RE |
|---|---|---|
| $\bar{y}$ | 91214.61 | 100.00 |
| $t_M$ | 159025.10 | 57.36 |
| $t_{SH(1)}$ | 415380.32 | 21.96 |
| $t_{SH(2)}$ | 10885.79 | 837.92 |
| $t_{SH(3)} = t_{KC}$ | 9172.48 | 994.44 |
| $t_{reg}$ | 65691.31 | 138.85 |

## 5. CONCLUSION

Table 1 shows that the minimum variance of the suggested class of estimators is 10885.79. The estimators suggested by Samiuddin and Hanif (13) and Kalidar and Cingi (15) have attained the lower bound of variance.

## REFERENCES

[1] Das, A. K. and Tripathi, T.P. 1981a. Sampling strategies for coefficient of variation using knowledge of the mean using an auxiliary character. *Tech.Rep.Stat.Math*.5/81.ISI,Calcutta.

[2] Das, A. K. and Tripathi, T.P. 1981b. A class of estimators for coefficient of variation using knowledge on coefficient of variation of an auxiliary character. Paper presented at $35^{th}$ annual conference of Ind. Soc. Agricultural Statistics. Held at New Delhi, India.

[3] Patel, P. A. & Shah, R. 2009. A Monte Carlo comparison of some suggested estimators of coefficient of variation in finite population, Journal of Statistics Sciences, 1 (2), 137-148.

[4] Patel, P. A. & Shah, R. 2012. Estimation of the finite population coefficient of variation in presence of two auxiliary variables, "*Journal of Statistics sciences",* 4 (1), 55 –63.

[5] Patel, P. A & Chaudhari, R.D. 2006. Design-based Horvitz-Thompson variance estimation: π-weighted ratio type estimator, *Statistics in Transition*, 7(6), 1277-1293.

[6] Patel, P. A & Chaudhari, R.D. 2008. Model -based Variance estimation under unequal probability sampling, *Metrika* 67, 171-187.

[7] Patel, P. A. & Patel, J. S. 2010. A Monte Carlo comparison of some variance estimators of Horvitz- Thompson estimator, J. Statistical Computing and Simulation, 80, 5, 489-502.

[8] Sarandal, C. E., Swensson, B., and Wretman, J. H. (1992), Model Assisted Survey Sampling, Springer Verlag, New York.

[9] Montanari, G. E. (1998), On regression estimation of finite population mean. Survey Methodology, 24:69-77

[10] Fuller, W. A. 2002. Regression estimation for sample surveys. Survey Methodology, 28, 5-25.

[11] Shah Rina 2012. On estimating non-linear functions of a finite populations using auxiliary information at the estimation stage, Unpublished Ph. D. Thesis submitted to Sardar Patel University, Vallabh Vidhyanagar.

[12] Mohanty S. 1967. Combination of Regression and Ratio Estimate. Journal of Indian Statistical Association 5, 16-19.

[13] Samiuddin M. and Hanif M. 2006. Estimation in two phase sampling with complete and incomplete information. *Proc. 8th Islamic Countries Conference on Statistical Sciences*. 13, 479-495.

[14] W.A. Abu-Dayyeh, M.S. Ahmed, R.A. Ahmed, H.A. Muttlak 2003. Some estimators of a finite population mean using auxiliary information, Applied Mathematics and Computation 139, 287–298.

[15] Kadilar, C. and Cingi, H. 2005. A new ratio estimator using two auxiliary variables. *Applied Mathematics and Computation*, 162,901-908.