

ALTERNATIVE TO FACTOR-TYPE ESTIMATOR UNDER SINGLE-PHASE SAMPLING

Audu Ahmed^{1*} and Adewara Adedayo Amos²

¹Department of Mathematics, Usmanu Danfodiyo University Sokoto, Nigeria

²Department of Statistics, University of Ilorin, Kwara State, Nigeria

*Corresponding author E-mail: ahmed.audu@udusok.edu.ng, H/P: +2348065350476

ABSTRACT

In this paper alternative estimator to factor-type estimator for estimating population mean has been proposed using information on $(N - n)$ population units yet to be drawn from both study and auxiliary variables. The expressions for bias and MSE for the proposed estimator in the form of population parameters using the concept of large sample approximation have been derived and the conditions for its efficiency over the conventional estimators have been established. Also empirical study to demonstrate the efficiencies of the proposed estimator over traditional estimators have been performed and the results show that the proposed estimator is less biased and performed better.

Keywords: Auxiliary variable, Factor-Type estimator, Precision, Bias, Mean Squared Error

INTRODUCTION

The utilization of auxiliary information either at planning stage or at design stage to improve the estimate of population parameters has been a common phenomenon in the sample surveys. Auxiliary information can improved the efficiencies of estimators if they are judiciously utilized especially during estimation stage. The use of auxiliary information for formulating factor estimator for estimating population mean $\bar{Y}$ was introduced by Singh and Shukla (1987) and (1993) under single phase sampling. The use of factor estimators fully depends on the information provided by the sample of size n and n_1 drawn from the population of size N . In many situations the information provided by the sample drawn may be insufficient, bias or not representative of population of interest. In such situations information on $(N - n)$ population units yet to be drawn from both study and auxiliary variables is used for the estimation as it contains the adequate and representative information required.

In this paper alternative estimator for factor-type estimator which uses information on $(N - n)$ population units yet to be drawn from both study and auxiliary variables to transform both $\bar{x}$ and $\bar{y}$ has been proposed and its efficiency over conventional estimators has been theoretically and empirically demonstrated.

Let $\Omega = (1, 2, 3 \dots N)$ be a population of size N and Y, X be two real valued functions having values $(Y_i, X_i) \in \mathbb{R}^+ > 0$ on the i^{th} unit of $U(1 \leq i \leq N)$. We assume positive correlation $\rho > 0$ between the study variable Y and auxiliary variable X . Let $\bar{Y}$ and $\bar{X}$ be the population means of Y and X respectively with C_y and C_x as coefficients of variation Y and X .

Let a pair (y_i, x_i) of simple random sample of size n be drawn without replacement from the population $\Omega = (1, 2, 3 \dots N)$ and $\bar{y}, \bar{x}$ be sample means based on the sample drawn. The usual sample mean $\bar{y}$, traditional ratio estimator $\bar{y}_r$ (Cochran, 1977), and Dual to ratio estimator $\bar{y}_{st}$ (Srivenkataramana, 1980) are defined as;

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}} \bar{X} \quad (1)$$

$$MSE(\bar{y}_r) = \bar{Y} \frac{1-f}{n} (C_y^2 + C_x^2 - 2\rho C_y C_x) \quad (2)$$

$$\bar{y}_{st} = \bar{y} \frac{N\bar{x} - n\bar{x}}{(N-n)\bar{x}} \quad (3)$$

$$MSE(\bar{y}_{st}) = \bar{Y} \frac{1-f}{n} \left(C_y^2 + \left(\frac{n}{N-n} \right)^2 C_x^2 - 2 \frac{n}{N-n} \rho C_y C_x \right) \quad (4)$$

Adewara (2015) proposed alternative to Kadilar (2014) ratio-estimator using the means of both the study and the auxiliary variables using information on $(N - n)$ population units yet to

be drawn under single-phase sampling. The estimator proposed and the MSE are given below

$$\bar{y}_{aa} = \bar{y}^* \frac{\bar{x}^*}{\bar{x}^*} \quad (7)$$

Where $\bar{x}^* = \bar{X} \left(1 - \left(\frac{n}{N-n} \right) \Delta_{\bar{x}} \right)$, $\bar{y}^* = \bar{Y} \left(1 - \left(\frac{n}{N-n} \right) \Delta_{\bar{y}} \right)$, $\Delta_{\bar{x}} = \frac{\bar{x} - \bar{X}}{\bar{X}}$, $\Delta_{\bar{y}} = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$ such that $|\Delta_{\bar{x}}| < 1$, $|\Delta_{\bar{y}}| < 1$

$$MSE(\bar{y}_{aa}) = \bar{Y}^2 \left(\frac{n}{N-n} \right)^2 \frac{1-f}{n} (C_y^2 + C_x^2 - 2\rho C_x C_y) \quad (8)$$

Adewara (2015) obtained in his work that alternative to Kadilar and Cingi (2004) ratio-estimator $\bar{y}_{aa}$ was more efficient than classical ratio estimator $\bar{y}_r$ if $0 < n < \frac{N}{2}$.

Factor-Type Estimator

Singh and Shukla (1987, 1993) suggested a factor-type estimator for population mean as

$$\bar{y}_{FT} = \bar{y} \frac{(A+C)\bar{X} + fB\bar{x}}{(A+fB)\bar{X} + C\bar{x}} \quad (5)$$

$$MSE(\bar{y}_{FT}) = \bar{Y}^2 \frac{1-f}{n} (C_y^2 + P^2 C_x^2 + 2P\rho C_x C_y) \quad (6)$$

Where $A = (d-1)(d-2)$, $B = (d-1)(d-4)$, and $C = (d-2)(d-3)(d-4)$,

$$f = \frac{n}{N}, \theta_1 = \frac{fB}{A+fB+C}, \theta_2 = \frac{C}{A+fB+C}, P = \theta_1 - \theta_2, \quad 0 < d < \infty$$

In their work, it was observed that factor-type estimator $\bar{y}_{FT}$ was more efficient than classical ratio estimator $\bar{y}_r$ if $P > \frac{2\rho C_y}{C_x} - 1$.

Proposed Estimator

Let $\bar{x}^*$ and $\bar{y}^*$ be the means of X and Y yet to be drawn (Srivenkataramana, 1980) i.e. the means corresponding to $(N - n)$ population units. Motivated by Adewara (2005, 2006 and 2015), the new factor estimator is given by

$$\bar{y}_{FT}^* = \bar{y}^* \frac{(A+C)\bar{X} + fB\bar{x}^*}{(A+fB)\bar{X} + C\bar{x}^*} \quad (7)$$

Where $\bar{X} = f\bar{x} + (1-f)\bar{x}^*$, $\bar{Y} = f\bar{y} + (1-f)\bar{y}^*$, $\bar{x} = \bar{X}(1 + \Delta_{\bar{x}})$, $\bar{y} = \bar{Y}(1 + \Delta_{\bar{y}})$,

$\bar{x}^* = \bar{X}\left(1 - \left(\frac{n}{N-n}\right)\Delta_{\bar{x}}\right)$, $\bar{y}^* = \bar{Y}\left(1 - \left(\frac{n}{N-n}\right)\Delta_{\bar{y}}\right)$, $\Delta_x = \frac{\bar{x}-\bar{X}}{\bar{X}}$, $\Delta_y = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$ such that

$$|\Delta_{\bar{x}}| < 1, |\Delta_{\bar{y}}| < 1.$$

Bias and MSE of the Proposed Estimator $\bar{y}_{FT}^*$

The proposed $\bar{y}_{FT}^*$ estimator can be expressed in terms of $\Delta_{\bar{x}}$ and $\Delta_{\bar{y}}$ as

$$\bar{y}_{FT}^* = \bar{Y}\left(1 - \frac{n}{N-n}\Delta_{\bar{y}}\right) \frac{(A+C)\bar{X} + fB\left(1 - \frac{n}{N-n}\Delta_{\bar{x}}\right)\bar{X}}{(A+fB)\bar{X} + C\left(1 - \frac{n}{N-n}\Delta_{\bar{x}}\right)\bar{X}}$$

Using power series expansion, the simplification up to first order approximation $O(n^{-1})$ is given by

$$\bar{y}_{FT}^* = \bar{Y}\left(1 - \frac{n}{N-n}\Delta_{\bar{y}}\right) \frac{1 - \frac{fB\left(\frac{n}{N-n}\Delta_{\bar{x}}\right)}{A+C+fB}}{1 - \frac{C\left(\frac{n}{N-n}\Delta_{\bar{x}}\right)}{A+C+fB}}$$

$$\bar{y}_{FT}^* = \bar{Y}\left(1 - \frac{n}{N-n}\Delta_{\bar{y}}\right)\left(1 - \theta_1 \frac{n}{N-n}\Delta_{\bar{x}}\right)\left(1 - \theta_2 \frac{n}{N-n}\Delta_{\bar{x}}\right)^{-1}$$

$$\bar{y}_{FT}^* - \bar{Y} = \bar{Y}\left((\theta_1 - \theta_2)\left(\frac{n}{N-n}\right)^2 \Delta_{\bar{y}}\Delta_{\bar{x}} - \frac{n}{N-n}\Delta_{\bar{y}} - (\theta_1 - \theta_2)\frac{n}{N-n}\Delta_{\bar{x}} - (\theta_1 - \theta_2)\left(\frac{n}{N-n}\right)^2 \Delta_{\bar{x}}^2\right)$$

Let $E(\Delta_{\bar{y}}) = E(\Delta_{\bar{x}}) = 0$, $E(\Delta_{\bar{y}})^2 = \frac{1-f}{n}C_y^2$, $E(\Delta_{\bar{x}})^2 = \frac{1-f}{n}C_x^2$, $E(\Delta_{\bar{y}}\Delta_{\bar{x}}) = \frac{1-f}{n}\rho C_y C_x$

Taking expectation of $\bar{y}_{FT}^* - \bar{Y}$, the bias of the proposed estimator is obtained as

$$Bias(\bar{y}_{FT}^*) = \bar{Y} \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 P(\rho C_y C_x - \theta_2 C_x^2)$$

Squaring both sides of $\bar{y}_{FT}^* - \bar{Y}$, then taking expectation, we obtain the MSE of the proposed estimator to terms of order n^{-1} as

$$MSE(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x)$$

$$MSE(\bar{y}_{FT}^*) = \left(\frac{n}{N-n}\right)^2 MSE(\bar{y}_{FT})$$

The bias and MSE of the proposed estimator to terms of order n^{-1} are given as;

$$Bias(\bar{y}_{FT}^*) = \bar{Y} \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 P(\rho C_y C_x - \theta_2 C_x^2) \quad (8)$$

$$MSE(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) \quad (9)$$

$$MSE(\bar{y}_{FT}^*) = \left(\frac{n}{N-n}\right)^2 MSE(\bar{y}_{FT}) \quad (10)$$

The MSEs of (7) and (9) is at minimum when $P = \theta_1 - \theta_2 = -\rho \frac{C_y}{C_x} = -v$ (say).

Simplifying relation $P = -v$, we obtained the cubic equation in the form of d as

$$(v-1)d^3 + (fv + f - 8v + 9)d^2 - (5fv + 5f - 23v + 26)d + (4fv + 4f - 22v + 24) = 0 \quad (11)$$

By solving (11), at most 3 zeros d_1, d_2 and d_3 of the polynomials for which (5) and (7) is optimal will be obtained.

Efficiency Comparisons

In this section, efficiency of the proposed estimator is compared with that of classical ratio estimator, factor-type estimator and conditions are obtained under which the proposed estimator is more efficient

$$(i) \text{MSE}(\bar{y}) - \text{MSE}(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} C_y^2 - \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) > 0$$

$$C_y^2 > \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(1 - \left(\frac{n}{N-n}\right)^2\right) C_y^2 > \left(\frac{n}{N-n}\right)^2 (P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} > P(P+2v)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} - 2v < P < \left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2}$$

And for $n \rightarrow \frac{N}{2}$, we have $-2v < P < 0$

This implies that $\text{MSE}(\bar{y}) - \text{MSE}(\bar{y}_{FT}^*)$ is positive if $0 < n < \frac{N}{2}$ and $-2v < P < 0$.

$$(ii) \text{MSE}(\bar{y}_r) - \text{MSE}(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} (C_y^2 + C_x^2 - 2\rho C_y C_x) - \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) > 0$$

$$(C_y^2 + C_x^2 - 2\rho C_y C_x) > \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(1 - \left(\frac{n}{N-n}\right)^2\right) C_y^2 + (C_x^2 - 2\rho C_x C_y) > \left(\frac{n}{N-n}\right)^2 (P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(\frac{N-n}{n}\right)^2 (1-2v) > P(P+2v)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(\frac{N-n}{n}\right)^2 (1-2v) > P(P+2v)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(\frac{N-n}{n}\right)^2 \left(1-2v \left[1 + \left(\frac{n}{N-n}\right)^2\right]\right) < P < \left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(\frac{N-n}{n}\right)^2 (1-2v)$$

And for $n \rightarrow \frac{N}{2}$, we have $(1-4v) < P < 0$

This implies that $\text{MSE}(\bar{y}_r) - \text{MSE}(\bar{y}_{FT}^*)$ is positive if $0 < n < \frac{N}{2}$ and $(1-4v) < P < (1-2v)$

$$(iii) \text{MSE}(\bar{y}_{st}) - \text{MSE}(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} \left(C_y^2 + \left(\frac{n}{N-n}\right)^2 C_x^2 - 2\frac{n}{N-n} \rho C_y C_x\right) - \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) > 0$$

$$\left(C_y^2 + \left(\frac{n}{N-n}\right)^2 C_x^2 - 2\frac{n}{N-n} \rho C_y C_x\right) > \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(1 - \left(\frac{n}{N-n}\right)^2\right) C_y^2 + \left(\frac{n}{N-n}\right)^2 \left(C_x^2 - 2\left(\frac{N-n}{n}\right) \rho C_x C_y\right) > \left(\frac{n}{N-n}\right)^2 (P^2 C_x^2 + 2P\rho C_y C_x)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(1 - 2\left(\frac{N-n}{n}\right)v\right) > P(P+2v)$$

$$\left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(1 - 2v \left[1 + \left(\frac{N-n}{n}\right)\right]\right) < P < \left(\frac{N(N-2n)}{n^2}\right) \frac{C_y^2}{C_x^2} + \left(1 - 2\left(\frac{N-n}{n}\right)v\right)$$

And for $n \rightarrow \frac{N}{2}$, we have $(1 - 4v) < P < (1 - 2v)$

This implies that $MSE(\bar{y}_{st}) - MSE(\bar{y}_{FT}^*)$ is positive if $0 < n < \frac{N}{2}$ and $(1 - 4v) < P < (1 - 2v)$

$$(iv) \quad MSE(\bar{y}_{FT}) - MSE(\bar{y}_{FT}^*) = \bar{Y}^2 \frac{1-f}{n} (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) - \bar{Y}^2 \frac{1-f}{n} \left(\frac{n}{N-n}\right)^2 (C_y^2 + P^2 C_x^2 + 2P\rho C_y C_x) > 0$$

$$1 > \left(\frac{n}{N-n}\right)^2$$

This implies that $MSE(\bar{y}_{FT}) - MSE(\bar{y}_{FT}^*)$ is positive if $0 < n < \frac{N}{2}$

EMPIRICAL STUDY

In order to demonstrate numerically the efficiency of the proposed estimator, we consider the data in the table below;

Table 1: Data Sets for Empirical Study

Population	Parameters								
	S_x^2	S_y^2	S_{xy}	$\bar{Y}$	$\bar{X}$	R	ρ	N	n
1	86858.54	97720.88	85207.57	454.6	457.3	0.9941	0.9735	288	20
2	791376.06	151031.29	268108.86	733.35	1372.6	0.0807	0.7755	912	30
3	242453.03	4127.63	25940.47	15.37	243.76	0.0631	0.82	106	20

Sources: Population1 (Adeniyi, 2014), Population (Okafor, 2002), Population 3 (Kadilar and Cingi, 2006)

Table 2: Bias, MSE and PRE of Proposed Estimator and Conventional Estimators

Population	Estimators	d	Bias	MSE	PRE
Population 1	$\bar{y}$	-----	0.0000	4546.724	100
	$\bar{y}_r$	-----	-0.3401	243.7679	1865.19
	$\bar{y}_{st}$	-----	0.6810	3949.808	115.11
	$\bar{y}_{FT}$	$d_1 = 2.2316$	-3.0775	237.7837	1912.13
		$d_2 = 0.8092$	-0.0479	237.7837	1912.13
		d_3^*	-----	-----	-----
	$\bar{y}_{FT}^*$	$d_1 = 2.2316$	-0.0171	1.3243	343330.3
		$d_2 = 0.8092$	-0.0003	1.3243	343330.4
		d_3^*	-----	-----	-----
Population 2	$\bar{y}$	-----	0.0000	4868.76	100
	$\bar{y}_r$	-----	3.6335	2915.677	166.99
	$\bar{y}_{st}$	-----	0.2142	4563.056	106.70
	$\bar{y}_{FT}$	$d_1 = 6.7807$	0.0426	1940.687	250.88
		$d_2 = 1.7971$	0.4784	1940.687	250.88
		$d_3 = 2.3023$	-0.6727	1940.687	250.88
	$\bar{y}_{FT}^*$	$d_1 = 6.7807$	0.00005	2.2452	216851.95
		$d_2 = 1.7971$	0.0006	2.2452	216851.95
		$d_3 = 2.3023$	-0.0008	2.2452	216851.95
Population 3	$\bar{y}$	-----	0.0000	167.435	100
	$\bar{y}_r$	-----	-1.7729	73.8426	226.75
	$\bar{y}_{st}$	-----	1.0040	138.6969	120.72
	$\bar{y}_{FT}$	$d_1 = 2.3740$	-3.8489	54.8545	305.235
		d_2^*	-----	-----	-----
		d_3^*	-----	-----	-----
	$\bar{y}_{FT}^*$	$d_1 = 2.3740$	-0.2082	2.9667	5643.812991
		d_2^*	-----	-----	-----
		d_3^*	-----	-----	-----

 d_i^* = Not Applicable

DISCUSSION AND CONCLUSION

From Table 2, it observed that the proposed alternative estimator to factor estimator is less biased compare to ratio estimator, Strivenkataramana estimator and factor-type estimator for all the population considered in the study. It is also observed that the proposed estimator is approximately unbiased for optimum choice of d . Table 2 also depicts the

efficiencies of the proposed estimator and traditional estimators and it is observed that the percent relative efficiency of the proposed estimator with respect to that of sample mean is higher and superior over the percent relative efficiencies of ratio estimator, Strivenkataramana estimator and factor-type estimator

REFERENCES

- Adeniyi, A. A. (2014): Application of simple random sampling and systematic sampling to students' enrolment in public secondary schools in Ondo State (Department of Statistics, Federal University of Technology, Akure).
- Adewara A. A. (2015): Alternative to Kadilar ratio type estimator, *Journal of the Mathematical Association of Nigeria*, Vol. 42, No. 2, pp. 55-62.
- Adewara A. A. (2006): Effect of improving both the auxiliary and variable of interest in ratio and product estimators, *Proc. Pakistan Academy Science*, Vol. 42, No. 4, pp. 275-278.
- Adewara A. A. (2005): Alternative to Ratio and Product estimators, Ph. D. thesis, University of Ilorin (Unpublished).
- Cochran, W. G. (1977). Sampling Techniques, Third U. S. Edition, Wiley Eastern Limited. New York
- Kadilar C. and Cingi, H. (2006). Improvement in estimating the population mean in simple random sampling, *Elsevier*, Vol. 19, pp. 75-79.
- Okafor, F. C. (2002): *Sample Survey Theory and Applications*. First Edition, Afro-Orbit Publications Ltd. University of Nigeria, Nsukka, Nigeria.
- Singh, V. K. and Shukla, D. (1987): One parameter family of factor type ratio estimator, *Metron*, Vol. 45, No. 1-2, pp. 273-283.
- Singh, V. K. and Shukla, D. (1993): An efficient one parameter family of factor - type estimator in sample survey, *Metron*, Vol.51, No. 1-2, pp. 139-159.
- Srivenkataramana, T. (1980): A dual of ratio estimator in sample surveys, *Biometrika*, Vol. 67, No. 1, pp. 199-204.