Science World Journal Vol. 19(No 3) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i3.17

Full Length Research Article

EXPONENTIAL TYPE ESTIMATOR FOR ESTIMATING FINITE POPULATION MEAN WITH AUXILIARY VARIABLES UNDER SIMPLE RANDOM SAMPLING

*M.E. Kanwai, Y. Yakubu, A. Isah, A. Usman

Federal University of Technology, Minna, Niger State, Nigeria

*Corresponding Author Email Address: mekanwai3@gmail.com

ABSTRACT
In this study, a ratio-product-cum-exponential type estimators for estimating the population mean in single phase sampling were proposed. The biases and Mean Square Errors (MSEs) of these estimators were obtained up to the first order of approximation. Theoretical and empirical comparative approach using real datasets and simulation study were investigated. The results showed that the proposed estimators were more efficient than the sample mean, ratio, product, exponential ratio and product estimator. Furthermore, the efficiency of the proposed estimators were investigated at different correlation levels and it was found that as the correlation increases the efficiency also changes positively. This suggests that when the auxiliary and study variables are more strongly correlated, the estimators become more efficient, reducing estimation errors and increasing precision.

Keywords: Exponential Type Estimator, Auxiliary Variables, Single-phase samplings, Efficiency.

INTRODUCTION
In sample survey, it is well established that auxiliary information is often used to improve the precision of estimators of population parameters (Singh, 2022). The auxiliary information typically is easy to measure; whereas the variable of interest may be expensive or difficult to measure. These auxiliary variables can be utilized at any of the following stage: pre-selection stage or designing stage; selection stage and post-selection or estimation stage. The ratio, product, regression, and difference estimators take advantage of auxiliary information at the estimation stage. The ratio estimator is most effective when the relationship between the study variable (y) and the auxiliary variable (x) is linear through the origin and the variance of y is proportional to x . The ratio estimator can be applied in areas such as the production of wheat and the area cultivated, the number of bullocks on a holding and its area in acres, the total production and is the number of workers, fuel consumed and distance travelled. But when the auxiliary variable is negatively correlated with the study variate, the product estimator is more appropriate for the estimation of the population mean. Its practical applications may include price and demand of a commodity, number of employees and time to complete a task, sale of woolen cloth and temperature, travel time and speed of a vehicle (Solanki *et al*., 2013)
In recent years, several research papers have emerged focusing on ratio type, exponential ratio type, and regression-type estimators, incorporating different types of transformations. Perry (2007), asserted that, when two or more auxiliary variables are available, many estimators may be defined by linking together different estimators such as ratio, product or regression, each one exploiting a single variable in order to enhance the efficiency of the estimators. Considering this fact, an attempt is made to improve the performance of different estimators. Riaz *et al.,* (2014) developed an estimator by combining the concept of Bahl and Tuteja (1991) exponential type estimator and classical regression estimator for the estimation of the population mean. Among others, we highlight, various authors that proposed mixed exponential type estimators, Singh and Vishwakarma (2007), Shabbir and Gupta (2010) proposed a regression-ratio-type exponential estimator by combining Rao (1991) and Bedi (1996) estimators, Yadav and Kadılar (2013), Ozgul and Cingi (2014) ,Singh *et al.*(2019), Singh *et al.*(2020),Audu *et al.*(2020), Audu and Singh (2020), Abiodun *et al.*(2021).
Survey statisticians often prefer ratio and product estimators due to their clear conceptual advantages (Perry, 2007). This preference likely drives the extensive research focused on improving the performance of these estimators. However, most of the ratio-based estimators can only be applied when the correlation between the study and auxiliary variables is positively strong. and the product-based estimators, when the estimators are negatively correlated. It is based on this background, that an alternative ratio-product-cum-exponential type estimator that provides more efficient estimates than some of the existing conventional estimators is proposed. This study focuses on deriving the properties of the newly developed ratio-product-cum-exponential type estimator, including its bias and Mean Square Error (MSE). Additionally, it establishes the theoretical efficiency conditions of the proposed estimator in comparison with existing estimators.

MATERIALS AND METHODS
Research Design
Consider a finite population $U = (U_1, U_2, \dots, U_N)$ of size N units. Let y and x denote the study variable and auxiliary variable respectively. The purpose is to estimate the population mean of the study variable $\bar{Y}$, a sample of size $n\ (n < N)$ assuming a simple random sampling without replacement (SRSWOR) s of size $n\ (n < N)$ is drawn from the population $U\ (s \subset U)$.
where: $\bar{y} = \frac{1}{n}\sum y_i$; $\rho_{yx} = \frac{s_{yx}}{s_x s_y}$, denote the sample correlation between y and x , while $C_y = \frac{s_y}{\bar{y}}$ and $C_x = \frac{s_x}{\bar{x}}$, denote the coefficient of variation of y and x respectively.

Existing Estimators
Some of the existing estimators with their Bias and MSE (Mean Square Error) up to the first order of approximation are given below.

Estimator without auxiliary variable
The sample variance estimator of the finite population variance,

www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

$$\bar{y} = s_y^2 \tag{1}$$

which is an unbiased estimator of finite population variance (S_y^2) and its variance is

$$\text{Var.}(\bar{y}) = \theta \bar{Y}^2 C_y^2 \tag{2}$$

$$\text{Bias}(\bar{y}) = 0 \tag{3}$$

Estimators with auxiliary variable

The usual ratio [Cochran (1940)] and product [Robson (1957) and Murthy (1964)] as cited in Singh *et al.,* (*2020*) estimators of population mean have been defined as:

Ratio estimator:

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}} \tag{4}$$

$$\text{MSE}(\bar{y}_R) = \theta \bar{Y}^2 (C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x) \tag{5}$$

$$\text{Bias}(\bar{y}_R) = \theta \bar{Y} (C_x^2 - \rho_{yx} C_y C_x) \tag{6}$$

Product estimator:

$$t_P = \bar{y}\frac{\bar{x}}{\bar{X}} \tag{7}$$

$$\text{MSE}(\bar{y}_p) = \theta \bar{Y}^2 (C_y^2 + C_x^2 + 2\rho_{yx} C_y C_x) \tag{8}$$

$$\text{Bias}(\bar{y}_P) = \theta \bar{Y} \rho_{yx} C_y C_x \tag{9}$$

Bahl and Tuteja (1991, as cited in Abiodun *et al.*, 2021), exponential ratio and product type estimator

Exponential Ratio:

$$\bar{y}_{ER} = \bar{y}\, exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \tag{10}$$

$$\text{MSE}(\bar{y}_{ER}) = \bar{Y}^2 \theta \left[C_y^2 + \frac{1}{4} C_x^2 - \rho_{yx} C_y C_x\right] \tag{11}$$

$$\text{Bias}\ (\bar{y}_{ER}) = \theta \bar{Y}\left[\frac{3C_x^2}{8} - \frac{\rho_{yx} C_y C_x}{2}\right] \tag{12}$$

Exponential Product:

$$\bar{y}_{EP} = \bar{y}\, exp\left[\frac{\bar{x} - \bar{X}}{\bar{x} + \bar{X}}\right] \tag{13}$$

$$\text{MSE}\ (\bar{y}_{EP}) = \theta \bar{Y}^2 \left[C_y^2 + \frac{1}{4} C_x^2 + \rho_{yx} C_y C_x\right] \tag{14}$$

$$\text{Bias}\ (\bar{y}_{EP}) = \theta \bar{Y}\left[\frac{\rho_{yx} C_y C_x}{2} - \frac{C_y^2}{8}\right] \tag{15}$$

Proposed Estimators

Using Cochran (1940), Murthy (1964), and Bahl and Tuteja (1991) estimators, the following are the proposed estimators: ratio-product-cum-exponential type of estimators for optimum and non-optimum scenario respectively for the estimation of population mean with one auxiliary variable in simple random sampling with partial information case.

$$T_{p1} = 2^{-1}\bar{y}\left\{\left(\frac{\bar{X}}{\bar{x}}\right)^{\varepsilon} + \left(\frac{\bar{x}}{\bar{X}}\right)^{1-\varepsilon}\right\}\left\{exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right)\right\} \tag{16}$$

$$T_{p2} = 2^{-1}\bar{y}\left\{\left(\frac{\bar{X}}{\bar{x}}\right) + \left(\frac{\bar{x}}{\bar{X}}\right)\right\}\left\{exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right)\right\} \tag{17}$$

To derive the biases and mean squares errors (MSEs) of the proposed estimators we define the following:

$$E(\bar{e}_0^2)=\theta C^2{}_y,\ E(\bar{e}_1^2)=\theta C^2{}_x,\ E(\bar{e}_0\bar{e}_1) = \theta C_y C_{x,}\rho_{yx} \tag{18}$$

$$\bar{y} = \bar{Y}(1 + \bar{e}_0),\quad \bar{x} = \bar{X}(1 + \bar{e}_1)\ \text{, s}\ \theta = \frac{1}{n} - \frac{1}{N} \tag{19}$$

To obtain the $MSE\ (T_{p1})$ to the first degree of approximation, express equation (16), in terms of $e's$, we have:

$$T_{p1} = 2^{-1}\bar{Y}(1 + \bar{e}_0)\left[\left(\frac{\bar{X}}{\bar{X}(1+\bar{e}_1)}\right)^{\varepsilon} + \left(\frac{\bar{X}(1+\bar{e}_1)}{\bar{X}}\right)^{1-\varepsilon}\right]\left[exp\left\{\frac{\bar{X} - \bar{X}(1+\bar{e}_1)}{\bar{X} + \bar{X}(1+\bar{e}_1)}\right\}\right] \tag{20}$$

$$T_{p1} = 2^{-1}\ \bar{Y}(1 + e_0)\left[(1 + e_1)^{-\varepsilon} + (1 + e_1)^{1-\varepsilon}\right]\left[exp\left(\frac{-\bar{X}e_1}{2\bar{X} + \bar{X}e_1}\right)\right] \tag{21}$$

$$T_{p1} = 2^{-1}\bar{Y}(1+e_0)\left[(1+e_1)^{-\varepsilon} + (1+e_1)^{1-\varepsilon}\right]\left[exp\left\{-\frac{\bar{e}_1}{2}\left(1+\frac{\bar{e}_1}{2}\right)^{-1}\right\}\right] \quad (22)$$

$$T_{p1} = 2^{-1}\bar{Y}(1+e_0)\left[\left(2+\bar{e}_1 - 2\varepsilon\bar{e}_1 + \varepsilon^2\bar{e}_1{}^2\right)\right]\left[exp\left(-\frac{\bar{e}_1}{2}+\frac{\bar{e}_1^2}{4}\right)\right] \quad (23)$$

$$T_{p1} = 2^{-1}\bar{Y}(1+e_0)\left[\left(2+\bar{e}_1 - 2\varepsilon\bar{e}_1 + \varepsilon^2\bar{e}_1{}^2\right)\right]\left[1-\frac{\bar{e}_1}{2}+\frac{3\bar{e}_1^2}{8}\right] \quad (24)$$

$$T_{p1} = \left[\bar{Y} + \bar{Y}\bar{e}_0 - \varepsilon\bar{Y}\bar{e}_1 - \varepsilon\bar{Y}\ \bar{e}_0\,\bar{e}_1 + \frac{1}{2}\left(\frac{1}{4}+\varepsilon+\varepsilon^2\right)\bar{Y}\bar{e}_1^2\right] \quad (25)$$

Subtracting $\bar{Y}$ from both sides of (25) and taking expectations, we get the bias of the estimator T_{p1}

$$Bias(T_{p1}) = E\left(T_{p1} - \bar{Y}\right) = \bar{Y}E\left[\bar{e}_0 - \varepsilon\bar{e}_1 - \varepsilon\bar{e}_0\,\bar{e}_1 + \frac{1}{2}\left(\frac{1}{4}+\varepsilon+\varepsilon^2\right)\bar{e}_1^2\right] \quad (26)$$

$$Bias(T_{p1}) = \bar{Y}\theta\left[\frac{1}{2}\left(\frac{1}{4}+\varepsilon+\varepsilon^2\right)C_x^2 - \varepsilon C_y C_x \rho_{yx}\right] \quad (27)$$

To obtain the Mean Squared Error (MSE) of the proposed estimator up to the first order of approximation, we subtract $\bar{Y}$ from both sides of (25), squaring both sides and then taking expectations of both sides.

$$MSE(t_{p1}) = E\left(T_{p1} - \bar{Y}\right)^2 = \bar{Y}^2E\left[\bar{e}_0^2 + \varepsilon^2\bar{e}_1^2 - 2\varepsilon\bar{e}_0\,\bar{e}_1\right] \quad (28)$$

Substituting equation (18) and (19) in equation (28) will gives:

$$MSE(T_{p1}) = \bar{Y}^2\theta\left[C_y^2 + \varepsilon^2C_x^2 - 2\varepsilon C_yC_x\rho_{yx}\right] \quad (29)$$

Special cases of proposed generalized estimators:

From (29) to investigate three common scenarios for the value of ε

Case (I): $\varepsilon = -1$

$$MSE(T_{p1}) = \bar{Y}^2\theta\left[C_y^2 + C_x^2 + 2C_yC_x\rho_{yx}\right] \quad (30)$$

This is equal to the MSE of the ratio estimator of Cochran (1940):

$$MSE(T_{p1}) = \text{MSE}(\bar{y}_P) \quad (31)$$

Case (II): $\varepsilon = 0$

$$MSE(T_{p1}) = \bar{Y}^2\theta C_y^2 \quad (32)$$

It led to the sample variance of the sample mean:

$$MSE(T_{p1}) = \text{Var.}(\bar{y}) \quad (33)$$

Case (III): $\varepsilon = 1$

$$MSE(T_{p1}) = \bar{Y}^2\theta\left[C_y^2 + C_x^2 - 2C_yC_x\rho_{yx}\right] \quad (34)$$

This yielded the MSE of the ratio estimator of Murthy (1964, as cited in Singh *et al., 2020*):

$$MSE(T_{p1}) = \text{MSE}(\bar{y}_R) \quad (35)$$

Differentiating (29) partially with respect to ε and equate to zero to obtain the optimum value of ε as

$$\varepsilon_{opt} = \frac{C_y\rho_{yx}}{C_x} \quad (36)$$

Substituting (36) in (29)and simplifying, the Mean Square Error of (17) is given as:

$$\text{MSE}\ (T_{p1})_{min} = \bar{Y}^2\theta\ C_y^2\left(1-\rho_{yx}^2\right) \quad (37)$$

The expression in (37) is equal to the MSE of linear regression estimator. Thus, the proposed estimator in (16) can be used as an alternative to the usual regression estimator in practice.

Now, to obtain the optimum bias, we substitute (36) in (27) and simplify

$$\text{Bias } (T_{p1})_{min} = \frac{1}{2}\bar{Y}\theta\left[\frac{1}{4}C_x^2 + C_yC_x\rho_{yx} - C_y^2\rho_{yx}^2\right] \tag{38}$$

Similarly, to obtain the MSE of T_2

$$T_{p2} = 2^{-1}\bar{Y}(1+\bar{e}_0)[(1-\bar{e}_1+\bar{e}_1^2)+(1+\bar{e}_1)]\left[1-\frac{\bar{e}_1}{2}+\frac{3\bar{e}_1^2}{8}\right] \tag{39}$$

$$T_{p2} = 2^{-1}\bar{Y}(1+\bar{e}_0)\left[\left(2-\bar{e}_1+\frac{3\bar{e}_1^2}{4}+\bar{e}_1^2\right)\right] \tag{40}$$

$$T_{p2} = \left[\bar{Y}+\bar{Y}\bar{e}_0-\frac{1}{2}\bar{Y}\bar{e}_1-\frac{1}{2}\bar{Y}\bar{e}_0\,\bar{e}_1+\frac{7}{8}\bar{Y}\,\bar{e}_1^2\right] \tag{41}$$

$$\text{Bias}(T_{p2}) = E(t_{p2}-\bar{Y}) = \bar{Y}\theta\frac{1}{2}\left[\frac{7}{4}C_x^2 - C_yC_x\rho_{yx}\right] \tag{42}$$

$$MSE(T_{p2}) = E(T_{p2}-\bar{Y})^2 \tag{43}$$

$$MSE(T_{p2}) = \bar{Y}^2E\left[\bar{e}_0-\frac{1}{2}\bar{e}_1\right]^2 \tag{44}$$

$$MSE(T_{p2}) = \bar{Y}^2\theta\left[C_y^2+\frac{1}{4}C_x^2-C_yC_x\rho_{yx}\right] \tag{45}$$

It can be observed from (45) that MSE of the non-optimum estimator (T_{p2}) is equal to that of the exponential ratio estimator of (12).

RESULTS AND DISCUSSION

Comparison of estimators

In this sub-section, theoretical comparison of the proposed estimators over other conventional estimators in single phase sampling is carried out.

The Mean Square Error (MSE) of an estimator is a key measure that combines both the variance and the bias of the estimator. It is defined as: MSE $(\hat{\theta})$ =Var. $(\hat{\theta})$ +Bias$[(\hat{\theta})]^2$, where $\hat{\theta}$ is the estimator.

Therefore, in comparing two estimators, say $\bar{y}_*$ and T_{p1} their MSEs provide a basis for determining which estimator is preferable. In practice, the estimator with the smaller MSE is typically preferred, as it minimizes the overall estimation error.

$$[Var(\bar{y}) - MSE(T_{p1})] = \theta\bar{Y}^2C_y^2\rho_{yx}^2 > 0 \tag{46}$$

$$[\text{MSE}(\bar{y}_R) - MSE(T_{p1})] = \theta\bar{Y}^2\left(C_x - \rho_{yx}^2C_y\right)^2 > 0 \tag{47}$$

$$[\text{MSE}(\bar{y}_P) - MSET_{p1})] = \theta\bar{Y}^2\left(C_x + \rho_{yx}^2C_y\right)^2 > 0 \tag{48}$$

$$[\text{MSE}(\bar{y}_{exp.R}) - MSE(T_{p1})] = \theta\bar{Y}^2\left(\frac{C_x}{2} - \rho_{yx}^2C_y\right)^2 > 0 \tag{49}$$

$$[\text{MSE}(\bar{y}_{\text{exp}.p}) - MSE(T_{p1})] = \bar{Y}^2\left(\frac{C_x}{2} + \rho_{yx}^2C_y\right)^2 > 0 \tag{50}$$

It is obvious from the above comparison that in terms of the MSE the proposed estimator (T_{p1}) outperforms the sample variance estimator, ratio estimator, product estimator, exponential ratio and exponential product estimators, while the second estimator (T_{p2}), yielded an equal in efficiency with the exponential ratio estimator.

Empirical Study

Real data sets for the empirical study for the

Parameters	Population I	Population II	Population III	Population IV
N	80	104	923	81
n	20	20	180	20
$\bar{Y}$	11.264	625.37	436.4345	33.8346
$\bar{X}$	51.826	13.93	11440.5	112.4568
C_y	0.750	1.866	1.7183	0.297194
C_x	0.354	1.653	1.8645	0.125559
ρ_{yx}	0.941	0.865	0.9543	-0.69079

Population I: Source, Murthy (1967 as cited in Abiodun *et al.*, 2021):

(i) **Auxiliary Variable (X):** Output of 80 factories.
(ii) **Study Variable (Y):** Fixed capital.
(iii) **Context:** This dataset was used to investigate the relationship between the output of factories and their fixed capital, which is a common scenario in industrial economics. The focus is on determining how well the proposed estimator can estimate the population mean of

Science World Journal Vol. 19(No 3) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i3.17

fixed capital using the auxiliary information about factory output.

(iv) **Findings:** With a high correlation coefficient ($\boldsymbol{\rho_{yx}}$ = 0.941), the study found a strong positive relationship between factory output and fixed capital, indicating that higher fixed capital is associated with greater output. This suggests the effectiveness of capital investment in increasing production.

Population II: Source, Kadilar and Cingi (2006):

(i) **Auxiliary Variable ($\boldsymbol{X}$):** Number of apple trees.
(ii) **Study Variable ($\boldsymbol{Y}$):** Level of apple production.
(iii) **Context:** This population was used to analysed agricultural productivity, specifically how the number of apple trees influences the total apple production. This is a typical scenario in agricultural studies where the goal is to maximize yield. The goal is to assess the effectiveness of the proposed estimator in predicting the population mean (total) of apple production using the auxiliary information on the number of apple trees.
(iv) **Findings:** The correlation ($\boldsymbol{\rho_{yx}}$= 0.865) indicates a strong positive relationship between the number of apple trees and apple production. The proposed estimator performed well. The efficiency improvements are in reduction of the MSE and an increase in the PRE when compared to other estimators., affirming the importance of scaling up planting to boost output.

Population III: Source, Koyuncu (2009):

(i) **Auxiliary Variable (X):** Number of students in both primary and secondary schools.
(ii) **Study Variable (Y):** Number of teachers.
(iii) **Context:** This dataset was used in educational studies to explore the relationship between the number of students and the number of teachers. The objective is to use the auxiliary information about the number of students to estimate the population mean or total of the number of teachers. This is essential for policy planning in education systems, particularly in ensuring adequate staffing for student populations.
(iv) **Findings:** The very high correlation ($\boldsymbol{\rho_{yx}}$= 0.9543) implies a strong positive relationship, suggesting that as the number of students increases, the number of teachers must also increase proportionally. This underscores the importance of aligning teacher recruitment with student enrollment to maintain effective education quality. Their estimator exhibited lower MSE and higher PRE compared to others, indicating its efficiency in this context.

Population IV: Source, Gujarati (2004):

(i) **Auxiliary Variable (X):** Top speed (miles per hour)
(ii) **Study Variable:** Average miles per gallon (fuel efficiency)
(iii) **Context:** This population examines the relationship between the top speed of vehicles and their average fuel efficiency. The primary aim is to estimate the population mean of the average miles per gallon using the auxiliary information about the top speed of the vehicles.

Findings: In contrast to Populations I, II, and III, Population IV presents a **negative correlation** ($\boldsymbol{\rho_{yx}}$= -0.69079) between the auxiliary variable (top speed) and the study variable (fuel efficiency). This indicates that as the top speed increases, the average fuel efficiency tends to decrease

Summary:
These populations were used in various studies to analyze the relationships between key variables in different fields such as: industrial productivity, agricultural output, and education. The findings generally show strong positive (negative) correlations between the auxiliary and study variables, indicating that increases in one variable tend to be associated with increases (decreases) in the other. These results highlight the importance of considering such relationships in planning and decision-making within these domains.

Simulation

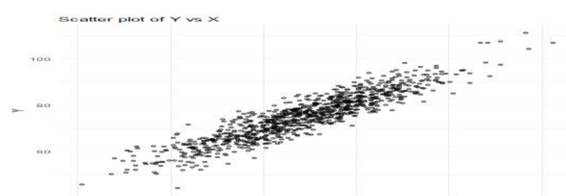
The rnorm function were employed for the simulations using R, a widely used statistical software environment for data analysis and simulations. The following are the steps employed for the data simulation:

Step (I): Set the parameters
Step (II): Function to perform the simulation for a given correlation sign
Step (III): Initialize empty vectors to store results
Step (IV): Loop for replications
Step (VI): Generate the study variable Y.POP
Step (VII): Generate the auxiliary variable X.POP.RATIO based on the correlation sign
Step (VIII): Calculate statistics for this replication
Step (IX): Create a summary table
Step (X): Run simulations for positive and negative correlations
Step (XI): Print the summary tables

The parameters used for the simulation are:

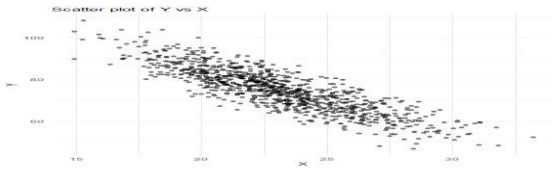
(a) Study variable $N = 1000$ $n = 250$ mean= 75 and standard deviation = 10.

The auxiliary variable is strongly positively correlated with the study variable (ρ_{yx} =0.7923) and the line passes through the origin. $N = 1000$ $n = 250$ mean=75, standard deviation = 10 and number of replications=5000



(ii) The auxiliary variable is strongly negatively correlated with the study variable (ρ_{yx} =-0.7923) and the line passes through the origin. $N = 1000$ $n = 250$ mean=75, standard deviation = 10 and number of replications=5000

Science World Journal Vol. 19(No 3) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i3.17



The objective of the simulation and models used in this analysis was to:

1. Simulate Different Populations: Two scenarios were simulated: one with a positive correlation and another with a negative correlation between the auxiliary and study variables. This helps in examining how the correlation sign affects the efficiency of the estimators.
2. Compare Estimators: The simulation aims to compare different estimators by calculating their Mean Square Error (MSE) and Percentage Relative Efficiency (PRE). The estimators being compared include the basic estimator (MSEY), ratio estimator (MSEYR), product estimator (MSEYP), exponential ratio estimator (MSEYER), exponential product estimator (MSEYEP), and ratio-product-cum exponential estimator (MSETP1).
3. **Evaluate Performance:** The MSE and PRE metrics are used to evaluate the performance of the proposed estimators against other existing estimators. The goal is to identify which estimator provides the most accurate and efficient estimate of the population mean of the study variable under different correlation conditions.

(b) Simulation at different correlation levels.

The different correlation levels are 0.25, 0.50, 0.75, zero, moderate and high, N$= 1000$ $n = 250$, mean=75, standard deviation = 10 and number of replications=5000.

The simulation aims to analyze how changes in correlation levels affect key statistical measures such as sample means, sample correlation, and coefficients of variation. The ultimate goal is to assess the efficiency and reliability of the estimators in estimating the population mean of the study variable across different correlation scenarios.

Therefore, in this research the Percent Relative Efficiency (PRE) is a statistical tool that will be used to measure the efficiency of the proposed and others estimators with respect to mean per unit estimator.

$$\text{PRE} = \frac{Var\ (\bar{y})}{Var\ (\bar{y}_*) \text{ or } MSE(\bar{y}_*)}\ X\ 100$$

for $*$=1,2,3,4,5,6 and$\bar{y}_1 = \bar{y}, \bar{y}_2 = \bar{y}_R$, $\bar{y}_3 = \bar{y}_P$, $\bar{y}_4 = \bar{y}_{ER}$ $\bar{y}_5 = \bar{y}_{EP}$ $\bar{y}_6 = T_{p1}$

Table 1: MSEs and PREs of Proposed and Convectional Estimators for Natural Population

	POPULATION I		POPULATION II		POPULATION III		POPULATION IV	
ESTIMATOR	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}$	2.67633	100	54993.75	100	2515.07	100	3.80730	100
$\bar{y}_R$	0.895177	298.972	13869.9	396.495	267.644	939.708	6.70916	56.7478
$\bar{y}_P$	5.64996	47.3689	182429	30.1454	10685.0	23.5383	2.26458	168.124
$\bar{y}_{ER} = T_{p2}$	1.63669	163.521	23643.0	232.601	651.042	386.315	5.08834	74.8241
$\bar{y}_{EP}$	4.01408	66.6734	107922	50.9568	5859.73	42.9213	2.86605	132.842
$\boldsymbol{T_{p1}}$	**0.306490**	**873.218**	**13846.1**	**397.180**	**224.625**	**1119.67**	**1.99049**	**191.274**

Note: MSE=Mean Square Error; PRE=Percent Relative Efficiency

Table 2: MSEs and PREs of Proposed and Convectional Estimators for Simulated Population

	POPULATION I		POPULATION I	
ESTIMATOR	MSE	PRE	MSE	PRE
$\bar{y}$	0.2998547	100	0.2999443	100
$\bar{y}_R$	0.1774565	168.97360	1.3025565	23.02736
$\bar{y}_P$	1.3768787	21.77786	0.1614552	185.77564
$\bar{y}_{ER} = \boldsymbol{T_{p2}}$	0.1193274	251.28741	0.6932350	43.26734
$\bar{y}_{EP}$	0.7190385	41.70218	0.1226844	244.48453
$\boldsymbol{T_{p1}}$	**0.1114807**	**268.9745**	**0.1115869**	**268.7990**

Table 3: Simulated result for the MSE and PRE of the proposed estimators compared to Convectional at various positive correlation levels.

	$\rho = 0.25$		$\rho = 0.50$		$\rho = 0.75$	
ESTIMATOR	**MSE**	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$\bar{y}$	0.2523319	100	0.3373012	100	0.2886265	100
$\bar{y}_R$	1.3222589	19.08339	0.3169014	106.43727	0.1550367	186.16656
$\bar{y}_P$	2.432578	10.37302	1.662866	20.28433	1.356284	21.28068
$\bar{y}_{ER} = \boldsymbol{T_{p2}}$	0.3810238	66.2247	0.1639557	205.7270	0.1050731	274.6911
$\bar{y}_{EP}$	0.9361832	26.95326	0.8369379	40.30182	0.7056968	40.89951
$\boldsymbol{T_{p1}}$	**0.20491869**	**123.1376**	**0.16379629**	**205.9273**	**0.09552017**	**302.1629**

In Table 1, which deals with natural populations, the proposed ratio-product-cum-exponential estimator T_{p1} consistently outperforms other estimators, including the simple mean $\bar{y}$, ratio estimator $\bar{y}_R$, product estimator $\bar{y}_P$, exponential ratio estimator $\bar{y}_{ER}$, and exponential product estimator $\bar{y}_{EP}$. This suggests T_{p1} is highly efficient in natural populations, likely due to its ability to integrate multiple estimation techniques that address both the central tendency and variability of the data. However, the second proposed estimator T_{p2}, matches the efficiency of $\bar{y}_{ER}$ but is less efficient than $\bar{y}_R$ and $\bar{y}_P$, indicating that T_{p2} may have limitations in capturing specific aspects of the population's structure that T_{p1} effectively handles. The presence of the scalar in T_{p1} is a key factor in its superior performance and efficiency, as it directly contributes to minimizing the MSE and enhancing the accuracy of the estimator, while the lack of such scalar in T_{p2} limits its efficiency, making it a less reliable choice for estimation in scenarios where minimizing MSE is crucial. When an estimator includes a scalar that is specifically designed to minimize its MSE, it means that the estimator can be fine-tuned to reduce the error associated with the estimates, leading to more precise and reliable results. In the case of T_{p1}, this scalar allows the estimator to adjust its weighting or combination of the auxiliary and study variables in such a way that the variance and bias of the estimator are minimized, resulting in the lowest possible MSE. This characteristic makes T_{p1} an optimum estimator, as it can consistently deliver high efficiency across different populations and correlation structures.

In Table 2, which presents results from simulated data, the findings are consistent across both positive and negative correlations between auxiliary (x) and study (y) variables. Estimator T_{p1} again demonstrates superiority over all other estimators, maintaining its dominance even under varying correlation scenarios. This robustness suggests that T_{p1} effectively adjusts to different correlation structures, making it a versatile tool for estimation. On the other hand, T_{p2} remains on par with $\bar{y}_{ER}$, indicating that while it is effective, it may not be as adaptable as T_{p1}.

Additionally, the performance of estimators in the second population with negative correlation highlights the nuanced nature of estimator efficiency. Specifically, $\bar{y}_p$ and $\bar{y}_{EP}$ outperform $\bar{y}_R$ and $\bar{y}_{ER}$ respectively, indicating that the choice of estimator should consider the sign and magnitude of the correlation between x and y.

Moreover, the importance of the correlation level between x and y on estimator performance cannot be overstated. As observed in Table 3, increasing the correlation level from low to high leads to a reduction in mean square error (MSE) across the estimators. This improvement in precision corresponds to an increase in percent relative efficiency (PRE), highlighting the critical role that strong correlations play in enhancing the accuracy of estimators. This finding underscores the importance of considering correlation strength when selecting an estimator for practical applications.

Conclusion

In this article optimum and non-optimum estimators were proposed for estimating the population mean of the study variable when the population mean of an auxiliary variable is known in simple random sampling without replacement (SRSWOR). The bias and mean square error expressions of the proposed class of estimators have been obtained up to first degree of approximation. It has been found theoretically as well as empirically that the optimum estimator is superior than the ratio, product, exponential ratio and exponential product estimators whereas the non-optimum estimator is less efficient than the ratio and product estimators, this confirmed the assertion by Kanwai *et al*. (2016) that optimum estimators performed better than the non-optimum estimators. It is apparent that the first proposed estimator demonstrated better performance across varying data conditions, whether the correlation is positive or negative. It is also, worth to note that when the correlation between the study variable and auxiliary variable is high the percent relative efficiency of the estimators increases.

Recommendation

Therefore, when the level of correlation between the auxiliary variable and the study variable is low the efficiency is good; when it is intermediate the efficiency is better and when it is high the efficiency is the best. Therefore, the proposed estimators can be used for estimating the population mean in practice for different

https://dx.doi.org/10.4314/swj.v19i3.17

situations by adjusting constant (ε) based on the characteristics of the auxiliary and study variables.

REFERENCES

Ahmed, A., Abiodun, Y. M., Ishaq, O. O., & Beki, D. O. (2021). An efficient exponential type estimator for estimating finite population mean under simple random sampling. *Anale. Seria Informatică*, *XIX*(fasc.), 46-51.

Audu, O. O., Ishaq, U., Isah, S., Muhammed, K. A., Akintola, A., Rashida, A., & Abubakar, A. (2020). On the class of exponential-type imputation estimators of population mean with known population mean of auxiliary variable. *NIPES Journal of Science and Technology Research, 2*(4), 1-11. https://doi.org/10.37933/nipes/2.4.2020.1

Audu, R. V., & Singh, K. (2020). Exponential-type regression compromised imputation class of estimators.*Journal of Statistics and Management Systems*, 1-15. https://doi.org/10.1080/09720510.2020.1814501

Gujarati, D. N. (2004). *Basic econometrics*. McGraw-Hill Companies.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory* (Vol. 2). John Wiley and Sons.

Kadılar, C., & Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics, 35*(1).

Kanwai, M. E., Asiribo, O. E., & Isah, A. (2016). Use of auxiliary variables and asymptotically optimum estimators in double sampling. *International Journal of Statistics and Probability, 5*(3), 55-62.

Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal,45*(2), 218–225.

Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2007). A general family of estimators for estimating population means using known value of some population parameter(s). *Far East Journal of Theoretical Statistics, 22*(2), 181-191.

Koyuncu, N., & Kadılar, C. (2009). Efficient estimators for the population mean. *Hacettepe Journal*

Perry, P. F. (2007). Improved ratio-cum-product type estimators. *Statistics in Transition, 8*(1), 51-69.

Samiuddin, M., & Hanif, M. (2006). Estimation in two phase sampling with complete and incomplete information. *Proceedings of the 8th Islamic Countries Conference on Statistical Sciences, 13*, 479-495.

Sharma, P., Singh, R., & Kim, J. M. (2013). Study of some improved ratio type estimators using information on auxiliary attributes under second order approximation. *Journal of Scientific Research, 57*, 138-146.

Singh, H. P., & Espejo, M. R. (2003). On linear regression and ratio-product estimation of a finite population mean. *The Statistician, 52*, 59-67. http://dx.doi.org/10.1111/1467-9884.00341

Singh, H. P., & Tailor, R. (2005). Estimation of finite population mean with known coefficient of variation of an auxiliary character. *Statistica, 65*(3), 301-313.

Singh, H. P., & Vishwakarma, G. K. (2007). Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian Journal of Statistics, 36*, 217-225.

Singh, P., Singh, R., & Bouza, C. N. (2019). Generalized exponential estimator for estimating population mean using auxiliary variable. *Journal of Scientific Research, 63*, 273-280.

Singh, R., & Kumar, M. (2011). A note on transformations on auxiliary variable in survey sampling. *Modern Assisted Statistics and Applications*, 17-19.

Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2009). Improvement in estimating the population mean using exponential estimator in simple random sampling. *Auxiliary Information and A Priori Values in Construction of Improved Estimators, International Journal of Statistics and Economics, 33*, 217-225.

Singh, R., Mishra, P., Ahmed, A., & Khare, S. (2020). Exponential type estimator for estimating finite population mean. *International Journal of Computational Theoretical Statistics, 7*(1), 37- 41. http://dx.doi.org/10.12785/ijcts/070104

Solanki, R. S., Singh, H. P., & Pal, S. K. (2013). Improved estimation of finite population mean in sample surveys. *Journal of Advanced Computing, 1*(1), 70-78.

Yadav, S. K., & Kadılar, C. (2013). Efficient family of exponential estimators for the population mean. *Hacettepe Journal of Mathematics and Statistics, 42*, 671-677

APPENDICES

APPENDIX A

Use of R Command for Proposed Estimator T_1, T_2 and others Estimators

```
> N<-c(80,104,923,81)
> n<-c(20,20,180,20)
> Y<-c(11.264,625.37,436.4345,33.8346)
> X<-c(51.826,13.93,11440.5,112.4568)
> Cy<-c(0.750,1.866,1.7183,0.297194)
> Cx<-c(0.354,1.653,1.8645,0.125559)
> Pyx<-c(0.941,0.865,0.9543,-0.69079)
> O<-{(1/n)-(1/N)}
> MSEY<-O*Y^2*(Cy)^2
> MSEYR<-O*Y^2*{(Cy)^2+(Cx)^2-2*Cy*Cx*Pyx}
> MSEYP<-O*Y^2*{(Cy)^2+Cx^2+2*Cy*Cx*Pyx}
```

```
> MSEYER<-O*Y^2*{(Cy)^2+((Cx)^2/4)-Cy*Cx*Pyx}
> MSEYEP<-O*Y^2*{(Cy)^2+((Cx)^2/4)+Cy*Cx*Pyx}
> MSETP1<-O*Y^2*Cy^2*{1-(Pyx)^2}
> PREY<-(MSEY/MSEY)*100
> PREYR<-(MSEY/MSEYR)*100
> PREYP<-(MSEY/MSEYP)*100
> PREYER<-(MSEY/MSEYER)*100
>                                  PREYEP<-(MSEY/MSEYEP)*100

> PRETP1<-(MSEY/MSETP1)*100
> MSEY
[1]    2.676326 54993.747732  2515.073817     3.807299
> MSEYR
[1] 8.951777e-01 1.386996e+04 2.676441e+02 6.709157e+00
> MSEYP
[1] 5.649961e+00 1.824285e+05 1.068504e+04 2.264576e+00
> MSEYER
[1]    1.636691 23642.990307   651.042254     5.088336
> MSEYEP
[1] 4.014083e+00 1.079222e+05 5.859739e+03 2.866046e+00
> MSETP1
[1] 3.064902e-01 1.384605e+04 2.246250e+02 1.990491e+00
> PREY
[1] 100 100 100 100
> PREYR
[1] 298.9715 396.4953 939.7084  56.7478
> PREYP
[1]  47.36894  30.14538  23.53828 168.12414
> PREYER
[1] 163.52053 232.60064 386.31499  74.82405
> PREYEP
[1]  66.67342  50.95683  42.92126 132.84154
> PRETP1
[1]  873.2175  397.1800 1119.6765  191.2744
```

APPENDIX B

Simulated Populations for the positive and negative correlation

```
> # Set the parameters
> population_size <- 1000
> replications <- 5000
> sample_size <- 250
>
> # Function to perform the simulation for a given c
orrelation sign
> simulate_population <- function(correlation_sign) {
+   # Initialize empty vectors to store results
+   mean_y <- numeric(replications)
+   mean_x <- numeric(replications)
+   corr_xy <- numeric(replications)
+   cv_y <- numeric(replications)
+   cv_x <- numeric(replications)
+
+   # Loop for replications
+   for (rep in 1:replications) {
+     # Generate the study variable Y.POP
+     Y.POP <- abs(rnorm(population_size, 75, 10))
+
+     # Generate the auxiliary variable X.POP.RATIO based on
the correlation sign
+     multiplier <- ifelse(correlation_sign ==
"positive", 0.65, -0.65)
+     offset <- ifelse(correlation_sign == "negative", 100, 0)
+     X.POP.RATIO <- abs(rnorm(population_size,
 multiplier * Y.POP + offset, 5))
+
+     # Calculate statistics for this replication
+     mean_y[rep] <- mean(Y.POP)
+     mean_x[rep] <- mean(X.POP.RATIO)
+     corr_xy[rep] <- cor(X.POP.RATIO, Y.POP)
+     cv_y[rep] <- sd(Y.POP) / mean(Y.POP)
+     cv_x[rep] <- sd(X.POP.RATIO) / mean(X.POP.RATIO)
+   }
+
+   # Create a summary table
+   summary_table <- data.frame(
+     Mean_Y = mean(mean_y),
+     Mean_X = mean(mean_x),
+     Correlation = mean(corr_xy),
+     CoV_Y = mean(cv_y),
+     CoV_X = mean(cv_x)
+   )
+
+   return(summary_table)
+ }
>
> # Run simulations for positive and negative correlations
> summary_positive <- simulate_population("positive")
> summary_negative <- simulate_population("negative")
>
> # Print the summary tables
> cat("Results for Positive Correlation:\n")
Results for Positive Correlation:
> print(summary_positive)
    Mean_Y   Mean_X Correlation     CoV_Y     CoV_X
1 74.99577 48.74642   0.7922133 0.1332642 0.1681382
> cat("\nResults for Negative Correlation:\n")

Results for Negative Correlation:
> print(summary_negative)
    Mean_Y   Mean_X Correlation     CoV_Y     CoV_X
1 75.00181 51.25115  -0.7926169 0.1333417 0.1600401
> population_size=N<-c(1000,1000)
> sample_size=n<-c(250,250)
> Mean_Y=Y<-c(74.99749,75.00285)
> Mean_X =X<-c(48.74707,51.24791)
> CoV_Y=Cy<-c(0.1333055,0.1333159)
> CoV_X=Cx<-c(0.1681877,0.1600053)
> Correlation =Pyx<-c(0.7926018,-0.7924486)
> O<-{(1/n)-(1/N)}
> MSEY<-O*Y^2*(Cy)^2
> MSEYR<-O*Y^2*{(Cy)^2+(Cx)^2-2*Cy*Cx*Pyx}
> MSEYP<-O*Y^2*{(Cy)^2+Cx^2+2*Cy*Cx*Pyx}
> MSEYER<-O*Y^2*{(Cy)^2+((Cx)^2/4)-Cy*Cx*Pyx}
> MSEYEP<-O*Y^2*{(Cy)^2+((Cx)^2/4)+Cy*Cx*Pyx}
> MSETP1<-O*Y^2*Cy^2*{1-(Pyx)^2}
> PREY<-(MSEY/MSEY)*100
> PREYR<-(MSEY/MSEYR)*100
> PREYP<-(MSEY/MSEYP)*100
```

Science World Journal Vol. 19(No 3) 2024 https://dx.doi.org/10.4314/swj.v19i3.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

```
> PREYEXR<-(MSEY/MSEYER)*100
> PREYEXP<-(MSEY/MSEYEP)*100
> PRETP1<-(MSEY/MSETP1)*100
> MSEY
[1] 0.2998547 0.2999443
> MSEYR
[1] 0.1774565 1.3025565
> MSEYP
[1] 1.3768787 0.1614552
> MSEYER
[1] 0.1193274 0.6932350
> MSEYEP
[1] 0.7190385 0.1226844
> MSETP1
[1] 0.1114807 0.1115869
> PREY
[1] 100 100
> PREYR
[1] 168.97360  23.02736
> PREYP
[1]  21.77786 185.77564
> PREYER
[1] 251.28741  43.26734
> PREYEP
[1]  41.70218 244.48453
> PRETP1
[1] 268.9745 268.7990
```

APPENDIX C

Simulated population at different levels of correlation between the auxiliary variable and the study variables

```
> # Set the parameters
> pop_size <- 1000
> num_replications <- 5000
> sample_size <- 250
> correlation_levels <- c(0.25, 0.50, 0.75)
>
> # Initialize variables to store results
> results <- matrix(NA, nrow = length(correlation_levels),
ncol = 6)
> colnames(results) <- c("Mean(Y.POP)", "Mean
(X.POP.RATIO.1)",
"Sample Correlation", "CoV(Y.POP)", "CoV(X.POP.RATIO.1)",
"Correlation Level")
>
> # Data Simulation for population One
> for (corr_level in correlation_levels) {
+  # Generate the study variable (Y.POP)
+  Y.POP <- abs(rnorm(pop_size, 75, 10))
+
+  # Generating the first Auxiliary Variable positively
correlated with Y.POP
+  X.POP.RATIO.1 <- rnorm(pop_size, mean =
corr_level * Y.POP, sd = 5)
+
+  # Correlations between study and auxiliary
variables
+  CORR1 <- cor(X.POP.RATIO.1, Y.POP)
+
+  # Sample indices
+  POP.Indices <- sample(1:pop_size, sample_size,
replace = TRUE)
+
+  # Samples for study variable (Y.POP) and
auxiliary variable (X.POP)
+  y.1 <- Y.POP[POP.Indices]
+  x.1 <- X.POP.RATIO.1[POP.Indices]
+
+  # Calculate statistics
+  mean_y <- mean(y.1)
+  mean_x <- mean(x.1)
+  sample_corr <- cor(x.1, y.1)
+  cov_y <- sd(y.1) / mean_y
+  cov_x <- sd(x.1) / mean_x
+
+  # Store results in the matrix
+  result_row <- c(mean_y, mean_x, sample_
corr, cov_y, cov_x, corr_level)
+  results[corr_level * 4, ] <- result_row
+ }
>
> # Print the summary results in an m x n table
> print(results)
   Mean(Y.POP)   Mean(X.POP.RATIO.1) Sample
Correlation  CoV(Y.POP)      CoV(X.POP.RATIO.1)
```

CORRELATION

```
 [1,]    74.91910      18.46877          0.4334744       0.1224145
 0.3106598           0.25
 [2,]    76.34668       8.33745          0.7172110       0.1388859
0.1931821          0.50  [3,]    75.83838      57.22973          0.8179564
 >
> N<-1000
> n<-250
> Y<-c(74.91910,76.34668,75.83838)
> X<-c(18.46877,8.33745,57.22973)
> Pyx<-c(0.4334744,0.7172110,0.8179564)
> Cy<-c(0.1224145,0.1388859,0.1293357)
> Cx<-c(0.3106598,0.1931821,0.1645222)
> O<-{(1/n)-(1/N)}
> MSET11<-O*Y^2*Cy^2
> MSET12<-O*Y^2*{Cy^2+Cx^2-2*Cy*Cx*Pyx}
> MSET13<-O*Y^2*{Cy^2+Cx^2+2*Cy*Cx*Pyx}
> MSET14<-O*Y^2*{Cy^2+(Cx^2/4)-Cy*Cx*Pyx}
> MSET15<-O*Y^2*{Cy^2+(Cx^2/4)+Cy*Cx*Pyx}
> MSET16<-O*Y^2*Cy^2*{1-(Pyx)^2}
> PRET11<-(MSET11/MSET11)*100
> PRET12<-(MSET11/MSET12)*100
> PRET13<-(MSET11/MSET13)*100
> PRET14<-(MSET11/MSET14)*100
> PRET15<-(MSET11/MSET15)*100
> PRET16<-(MSET11/MSET16)*100
```

https://dx.doi.org/10.4314/swj.v19i3.17

```
> MSET11
[1] 0.2523319 0.3373012 0.2886265
> MSET12
[1] 1.3222589 0.3169014 0.1550367
> MSET13
[1] 2.432578 1.662866 1.356284
> MSET14
[1] 0.3810238 0.1639557 0.1050731
> MSET15
[1] 0.9361832 0.8369379 0.7056968
> MSET16
[1] 0.20491869 0.16379629 0.09552017
> PRET11
[1] 100 100 100
> PRET12
[1]  19.08339 106.43727 186.16656
> PRET13
[1] 10.37302 20.28433 21.28068
> PRET14
[1]  66.2247 205.7270 274.6911
> PRET15
[1] 26.95326 40.30182 40.89951
> PRET16
[1] 123.1376 205.9273 302.1629

>
```