1.1 Introduction

It is a well known fact that for estimating the population mean $\mu_y$ of a random variable $Y$, precision of the estimator can be increased when information on an auxiliary variable $X$, highly correlated with $Y$ is readily available on all the units of the population. When the relationship between $Y$ and $X$ is found to be approximately linear but does not pass through the origin, linear regression estimate may be used, which is given by (Cochran, 1977)

$$t_i = \bar{y} + b(\mu_x - \bar{x}) \quad \ldots (1.1)$$

where $b$ is an estimate of the change in $y$ when $x$ is increased by unity, $\bar{y}$ and $\bar{x}$ are the sample means of the variables $Y$ and $X$ respectively and $\mu_x$ is the population mean of $X$.

The regression estimate given in (1.1) require advance knowledge about $\mu_x$, the population mean of the auxiliary variable $X$. When such information is lacking, it is sometimes relatively cheaper to take a large preliminary sample in which $x_i$ alone is measured and used for estimating the population characteristic like mean, total or frequency distribution of $x$ values. The purpose of this sample is to furnish a good estimate of $\mu_x$ or of the frequency distribution of $x_i$. Another independent or sub-sample to observe both $(x_i, y_i)$ meant to estimate $(\bar{x}, \bar{y})$ for using it in the regression estimator. This technique is known as double sampling and was first formulated by Neyman (1938) in connection with the collection of information on strata sizes in a stratified sampling experiment.
1.2 Double sampling with partial information on auxiliary variables

To use the linear regression estimator \( t_1 \) it is usually assumed that population mean \( \mu_x \) is known. However in certain practical situations, \( \mu_x \) is not known a priori, in which case the technique of double sampling is applied. In the first preliminary sample of size \( n' \), we measure only \( x_i \) and use it for the estimation of \( \mu_x \); in the second sample, a random sub-sample of size \( n < n' \), from the preliminary sample we observed both \( x_i \) and \( y_i \). Under double sampling the regression estimate (1.1) becomes

\[
t_2 = \bar{y}_n + b(\bar{x}_n - \bar{x}_{n'})
\]

where \( \bar{x}_{n'} \) is the mean of \( x_i \) in the preliminary sample and \((\bar{y}_n, \bar{x}_n)\) are the means of \( x_i \) and \( y_i \) from the sub-sample and \( b \) is the least square regression coefficient of \( Y \) on \( X \) which can be computed from the sub-sample.

Han (1973) described that the precision of an estimator can be improved if auxiliary variables are used in a regression estimator based on double sampling with partial information on auxiliary variable. Sometimes there are situations where we have partial information about the mean \( \mu_x \) of the auxiliary variable \( X \). In order to utilize the partial information Han (1973) suggested the use of a preliminary test and constructed a preliminary test estimator using double sampling with partial information on the auxiliary variable as follows:

\[
t_3 = \begin{cases} 
(\bar{y}_n - \rho \bar{x}_n) & \text{if } |\bar{x}_{n'}| \leq Z_{\alpha} / \sqrt{n'} \\
(\bar{y}_n + \rho (\bar{x}_{n'} - \bar{x}_n)) & \text{if } |\bar{x}_{n'}| > Z_{\alpha} / \sqrt{n'}
\end{cases}
\]  

\( t_3 \)  \( \cdots \cdots \cdots (1.3) \)
In estimating the population mean $\mu_Y$ of the random variable $Y$, suppose that in addition to information on an auxiliary variable $X$, information on yet another auxiliary variable $Z$ is available. When populations means $\mu_X$ and $\mu_Z$ are not available, one can take a preliminary sample to estimate these by the use of double sampling. In such a situation an estimator using $X$ and $Z$ is being suggested by Mukerjee et al. (1987).

Das (1992), Das and Bez (1995) and Das (2003) suggested some preliminary test estimators for the population mean in double sampling with two auxiliary variables, alternative to the usual regression estimator, when the experimenter has partial information on one and/or both the auxiliary variables.

The present work is aimed to proceed in accordance to further enhance the work done by Han (1973) and Das (1995) and several other authors to find an appropriate estimator through the use of preliminary test estimation and double sampling procedures.

1.3 The combined regression preliminary test estimator (CRPTE) in double sampling.

It is known that stratified sampling consists of classifying the population units in a certain number of groups called strata and selecting samples independently from each group. The division of population into strata can be done in such a way that the values of the study variable are homogeneous within each stratum, in that the measurement vary little from one unit to another, a precise estimates of any stratum mean can be obtained from a small sample in that stratum. These estimates with best choices of sample sizes can
be combined into a precise estimate for the whole population. When appropriately used, the variance of the estimated mean of the study variable Y under stratification is usually less than that of the variance under simple random sampling (Cochran 1977).

Stratification can also be operationally convenient and economical if the sampling frame is available in the form of sub-frames. Stratification enables that the demarcation of the strata boundaries and the allocation of the total sample size to the strata may be done so as to make the estimator most efficient from the point of view of sampling variability and cost. Though the main advantage of using stratified sampling is the possible increase in efficiency per unit of cost in estimating the population characteristics, the method is also useful in situation when estimators are required with specific margins of errors not only for population as a whole but for certain groups of units.

Cochran(1977) suggested a regression estimate in stratified sampling which he called a combined regression estimator and is given by

\[
\bar{y}_{rc} = \bar{y}_{st} + b(\mu_s - \bar{x}_s),
\]

where

\[
\bar{y}_{st} = \sum_s w_s \bar{y}_s \quad \text{and} \quad \bar{x}_s = \sum_s w_s \bar{x}_s.
\]

In this estimate the whole population is stratified into different classes and samples are selected from each stratum by simple random sampling and the stratum means are combined and used in a regression equation to obtain the desired mean. Here \( b \) is the estimate of combined regression coefficient and \( w_s \) is the stratum weight.
The combined linear regression estimator given by \( \bar{y}_n \), can be utilized under three situations. Firstly when the population mean \( \mu_x \) is known, as a consequence of which the study reduces to usual combined regression method of estimation. Secondly in certain practical situations \( \mu_x \) is not known a priori, in which case the technique of double sampling can be applied wherein a preliminary sample is obtained to estimate \( \mu_x \) and the estimator of \( \mu_y \) is given by

\[
 t_a = \bar{y}_n + b(\bar{x}_n' - \bar{x}_u) , \quad \text{where} \quad \bar{y}_n = \sum_{h} w_h \bar{y}_h \quad \text{and} \quad \bar{x}_n' = \sum_{h} w_h \bar{x}_h
\]

Here \( \bar{x}_n' \) is the value of the mean of \( X \) obtained from the preliminary sample and is utilized to estimate \( \mu_x \). Thirdly when \( \mu_x \) is partially known, then a preliminary test estimation using double sampling procedure can be used.

In the present study, the third case will be considered where partial information about the mean of the auxiliary variable will be used. The first sample is a stratified simple random sample of size \( n \) in which the pair \( (x_n, y_n) \) values are measured from \( n_h \) units drawn from each stratum and consequently estimating of the pair \( (\bar{x}_u, \bar{y}_u) \), with \( n = \sum_h n_h \). The second sample is a larger simple random sample of size \( n' (= n + m) \) is obtained by supplementing \( m \) more independent observations on \( X \) where only \( x_1 \) is measured and evaluates \( \bar{x}_n' \), which is utilized to estimate \( \mu_x \).
In order to utilize the partial information, a preliminary test is done about the hypothesis

\[ H_0 : \mu_x = \mu_o, \quad \text{against} \quad H_1 : \mu_x \neq \mu_o \]

where \( \mu_o \) is the value obtained from the partial information.

If \( H_0 \) is accepted then \( \mu_o \) is used to replace \( \mu_x \) in the regression estimator \( \bar{y}_{bc} \) and if \( H_0 \) is rejected then the sample mean \( \bar{x}_{n'} \) based on the preliminary sample is used.

We assume that the auxiliary variable \( X \) and the study variable \( Y \) and are jointly normally distributed with parameters given by \((\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)\). The marginal distributions which is the distribution of the study variable \( Y \) and the auxiliary variable \( X \) will also follow normal distribution given as \( Y \sim N(\mu_Y, \sigma_Y^2) \) and \( X \sim N(\mu_X, \sigma_X^2) \). The strata population \((X_s, Y_s)\) being carved out from the parent population are also jointly assumed to follow the bivariate normal distribution with parameters written as \((\mu_{X_s}, \mu_{Y_s}, \sigma_{X_s}^2, \sigma_{Y_s}^2, \rho)\). Also since the relationship between the pair \((X, Y)\) is always maintained even within the stratum the strata correlations are assumed to be equal to the population correlation coefficient \( \rho \). The regression estimator depends on weather the covariance matrix is known or not. If known, one may let \( \sigma_X^2 = \sigma_Y^2 = 1 \) without loss of generality (WLOG).

Since the population is assumed to follow normal distribution, the preliminary sample utilize to collect information on the auxiliary variable for the
estimation of $\bar{x}'$, is also assume to follow normal distribution and therefore

$$\bar{x}' \sim N(\mu_x, \sigma_x^2/n')$$

and under the assumption $\sigma_x^2 = \sigma_y^2 = 1$, $\bar{x}' \sim N(\mu_x, 1/n')$.

Further marginal distributions of $X_h$ and $Y_h$ are also normal given as

$$X_h \sim N(\mu_{x_h}, \sigma_{x_h}^2) \quad \text{and} \quad Y_h \sim N(\mu_{y_h}, \sigma_{y_h}^2).$$

For each stratum, the pair of variables $(X_h, Y_h)$ for every $h$, follows a bivariate normal distribution with mean $(\mu_{x_h}, \mu_{y_h})$ and covariance matrix given by

$$\Sigma_h = \begin{pmatrix} \sigma_{x_h}^2 & \rho \sigma_{x_h} \sigma_{y_h} \\ \rho \sigma_{x_h} \sigma_{y_h} & \sigma_{y_h}^2 \end{pmatrix}$$

The regression estimator depends on whether $\Sigma_h$ is known or not. If $\Sigma_h$ is known, one may let $\sigma_{x_h}^2 = \sigma_{y_h}^2 = 1$, (WLOG).

Also, the stratum means are given by

$$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{ih}}{n_h} \quad \text{and} \quad \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{ih}}{n_h}$$

are linear combinations of normally distributed random variables $(X_h, Y_h)$. Hence it can be easily observed that $\bar{x}_h$ and $\bar{y}_h$ also follow normal distribution with mean and variances given by

$$\bar{x}_h \sim N(\mu_{x_h}, \sigma_{x_h}^2/n_h) \quad \text{and} \quad \bar{y}_h \sim N(\mu_{y_h}, \sigma_{y_h}^2/n_h)$$

i.e. $\bar{x}_h \sim N(\mu_{x_h}, 1/n_h) \quad \text{and} \quad \bar{y}_h \sim N(\mu_{y_h}, 1/n_h)$ under the assumption of $\sigma_{x_h}^2 = \sigma_{y_h}^2 = 1$.

The joint distribution of $(\bar{x}_h, \bar{y}_h)$ is bivariate normal with mean as $(\mu_{x_h}, \mu_{y_h})$ and covariance matrix given by
In certain situations, the experimenter may have partial information about \( \mu_x \). In order to utilize the partial information, one can perform a preliminary test about the hypothesis

\[
H_0 : \mu_x = \mu_0 \quad \text{against} \quad H_1 : \mu_x \neq \mu_0
\]

where \( \mu_0 \) is the value obtained from the partial information and \( \bar{x}_{n'} \) is the value of the mean of \( X \) obtained from the preliminary sample through the use of double sampling.

Now when \( \mu_x \) is partially known, one can let \( \mu_0 = 0 \) (WLOG), so that the hypothesis can be accepted, when

\[
\left| \bar{x}_{n'} - \mu_0 \right| / \text{SE}(\bar{x}_{n'}) \leq Z_a \quad \Rightarrow \quad \bar{x}_{n'} \leq Z_a / \sqrt{n'}
\]

where \( Z_a \) is the 100(1-\( \alpha/2 \))% point of \( N(0,1) \) and \( \alpha \) is the level of significance of the preliminary test.

Under the above assumption the CRPTE in double sampling having partial information on the auxiliary variable \( X \) can be written as

\[
t_s = \begin{cases} (\bar{y} - \rho \bar{x}) & \text{if} \quad |\bar{x}_{n'}| \leq Z_a / \sqrt{n'} \\ (\bar{y} + \rho (\bar{x}_{n'} - \bar{x})) & \text{if} \quad |\bar{x}_{n'}| > Z_a / \sqrt{n'} \end{cases} \quad (1.5)
\]

where \( \bar{y}_{i'} = \sum_h w_h \bar{y}_h \) and \( \bar{x}_{i'} = \sum_h w_h \bar{x}_h \).

and \( b \) from \( \bar{y}_{hc} \) reduces to \( b = \rho (\sigma_y / \sigma_x) = \rho \) under the above assumptions.
1.4 Bias of the CRPTE in double sampling.

To evaluate the bias of $t_b$, we considered that the joint distribution of $(\bar{x}_u, \bar{x}_s, \bar{y}_u)$ which is a multivariate normal with mean $(\mu_x, \mu_x, \mu_y)$ and covariance matrix given by

$$
\Sigma = \begin{pmatrix}
\text{Var}((\bar{x}_u)) & \text{Cov}((\bar{x}_u, \bar{x}_u)) & \text{Cov}((\bar{x}_u, \bar{y}_u)) \\
\text{Cov}((\bar{x}_u, \bar{x}_u)) & \text{Var}((\bar{x}_u)) & \text{Cov}((\bar{x}_u, \bar{y}_u)) \\
\text{Cov}((\bar{y}_u, \bar{x}_u)) & \text{Cov}((\bar{y}_u, \bar{x}_u)) & \text{Var}((\bar{y}_u))
\end{pmatrix}

(1.6)

The derivation of the bias of the estimator $t_b$ involves conditional expectations, the condition being the acceptance or rejection of the hypothesis considered in the preliminary test. Further the expectations can be obtained from the integrals involving probability density functions which are assumed to be normal.

The variance of $\bar{x}_u$ and $\bar{y}_u$ being given by

$$
\bar{x}_h \sim N(\mu_{x_h}, \sigma_{x_h}^2 / n_h) \quad \text{and} \quad \bar{y}_h \sim N(\mu_{y_h}, \sigma_{y_h}^2 / n_h)
$$

i.e.

$$
\bar{x}_h \sim N(\mu_{x_h}, 1 / n_h) \quad \text{and} \quad \bar{y}_h \sim N(\mu_{y_h}, 1 / n_h)
$$

under the assumption $\sigma_{x_h}^2 = \sigma_{y_h}^2 = 1$.

Also when the samples are selected with proportional allocation then the stratum weight is given by $W_h = (N_h / N) = (n_h / n)$

Thus $\sum_h W_h^2 / n_h = \sum_h W_h^2 / n \sum_h W_h = (1/n) \sum_h W_h = (1/n) \sum_h W_h = 1$ (as $\sum_h W_h = 1$)

Therefore the above covariance matrix in (1.6) reduces to

$$
\Sigma = \begin{pmatrix}
1/n' & 1/n' & \rho/n' \\
1/n' & 1/n' & \rho/n' \\
\rho/n' & \rho/n' & 1/n
\end{pmatrix}

(1.7)
The Bias of an estimator is defined as

\[ \text{Bias} \left( t_n \right) = E \left( t_n \right) - \mu_y \]

where \( E(.) \) is the expectation

\[ \text{Bias} \left( t_n \right) = E \left\{ \bar{y}_u - \rho \bar{x}_u \right\} \quad \text{if} \quad \left| \bar{x}_u \right| \leq Z_n / \sqrt{n'} \]
\[ + E \left\{ \bar{y}_u + \rho (x_{u'} - \bar{x}_u) \right\} \quad \text{if} \quad \left| \bar{x}_u \right| > Z_n / \sqrt{n'} - \mu_y \]

After routine derivation we get

\[ \text{Bias} \left( t_n \right) = -\rho \mu_x \left\{ \Phi \left( A \right) - \Phi \left( B \right) \right\} + \rho \left( 1 / \sqrt{n'} \right) \left\{ \varphi \left( A \right) - \varphi \left( B \right) \right\} \]

where \( \Phi(.) \) is the cumulative distribution function of \( N(0,1) \) and \( \varphi(.) \) is its density function and \( A = Z_n - \sqrt{n'} \mu_x \), \( B = -Z_n - \sqrt{n'} \mu_x \).

The values of \( \text{Bias} \left( t_n \right) \) can be easily computed for different values of \( \mu_x \). We notice that the bias is symmetric about \( \mu_x = 0 \), hence we need to consider only the case when \( \mu_x \geq 0 \). In order to get an idea about the behavior of the bias function with respect to \( \mu_x \), \( \text{Bias} \left( t_n \right) \) is computed for a set of values of \( n, n', \alpha \) and \( \rho \). As \( \mu_x \) increases, the \( \text{Bias} \left( t_n \right) \) increases to a maximum and then decreases to zero. This establishes the utility of the present study that the utilization of partial information and preliminary test of the auxiliary variable reduces the bias of the proposed estimator.

The above analytical method used for computing of the bias of the proposed estimator involves the evaluation of mathematical expectation of the random variables and consequently results in the computation of integrals. However, sometimes computation of integrals analytically may become cumbersome. Therefore an alternative method for the evaluation of the bias is
also sought with the help of numerical techniques. In the present study, attempt is made to evaluate the bias of the constructed estimator \( \hat{t} \), using numerical integration with programmes written in Fortran 77. When the results obtained numerically and that by analytical methods are compared, it is found that the bias function shows a similar pattern and difference wherever it exists is very negligible.

### 1.5 Mean square error of the CRPTE in double sampling.

The precision or a measure of the closeness of the sample estimates to the census count taken under identical conditions is judged in sampling theory by the variances of the estimators concerned. Here reliance is placed on the fact that with a small variance the probability of large deviations from the census count will be small. The general principle is to use estimators which will give the highest concentration of the sample estimates (in the sense of probability) around the valued aimed for. With unbiased estimators the method used for judging the degree of concentration is the variance of the estimators.

It may happen sometime that the degree of concentration of the sample around the valued aimed at is higher for the distribution of a biased estimator than for an unbiased one. In such a situation the biased estimator is preferable to the unbiased one. However in order to compare a biased estimator with an unbiased estimator, or two estimators with different amounts of bias, variance is not a satisfactory criterion, since it measures deviation from the expected value of the estimator, which is not the same as the population value. A useful
criterion is the mean square error (MSE) of the estimate, measured from the population value that is being estimated.

Formally, \( \text{MSE}(t) = \text{Var}(t) + \{\text{Bias}(t)\}^2 \)

To obtain the MSE of \( t_s \), we notice that

\[
\text{MSE}(t_s) = \text{var}(t_s) + \{\text{Bias}(t_s)\}^2 = E(t_s) - \{E(t_s)\}^2 + \{\text{Bias}(t_s)\}^2
\]

After routine derivations

\[
\text{MSE}(t_s) = \left\{ \frac{(1 - \rho^2)}{n} + \frac{\rho^2}{n'} \right\} + \left\{ \rho^2 \left\{ A\phi(A) - B\phi(B) \right\} - \rho^2 \left( \frac{1}{n'} - \mu^2 \right) \right\} \text{E}(A - B) \quad \ldots (1.9)
\]

It is seen that the analytical method of determining the MSE involves evaluating the mathematical expectations of the random variables like \( E(\tilde{x}_s) \), \( E(\tilde{x}_s^2) \), \( E(\tilde{x}_s \tilde{y}_s) \) and \( E(\tilde{x}_s \tilde{y}_s) \). The derivation of these expectations is done using moment generating function and also involves the application of single and double integration technique. In the process of evaluation which involves bivariate frequency distributions, a tedious substitution of variables is necessary to simplify the integrals. The above expectation is finally obtained by differentiating under the integral sign. As a consequence of the complexity involved in analytical deductions and the availability of numerical techniques, the MSE is evaluated using the numerical methods. In the numerical methods, the use of moment generating function and the substitutions involved can be avoided. The results of MSE obtained numerically show the similar pattern with that of the one derived by analytical methods for increasing values of the mean \( \mu_s \) of the auxiliary variable. The differences in the values of MSE between analytical and numerical methods of computations are minimal.
1.6 Relative efficiency

The study of the mean square error of an estimator will not be completed unless it is compared with other estimators. Without the use of real life data, relative efficiency of an estimator can be obtained analytically as the ratio of the variance or mean square error of one estimator to that of the mean square error of the proposed estimator. If the relative efficiency is greater than 1, it can be concluded that the proposed estimator is more efficient in comparison to the other estimator.

In the present study, the mean square error of the proposed estimator is compared with other estimator $t_4$ and conclusion is drawn through the relative efficiency. Under similar assumptions a routine analysis gives

$$\text{MSE } (t_4) = \frac{1}{n}(1 - \rho^2) + (1/n')\rho^2$$

Therefore the relative efficiency of $t_4$ to $t_4$ is given by

$$e_1 = \frac{\text{MSE}(t_4)}{\text{MSE}(t_4)}$$

In order to get an idea about the behavior of the relative efficiency function with respect to $\mu_2$, $e_1$ is compute for a set of values of $n, n', \alpha$ and $\rho$. It is found in general that $e_1$ has a maximum at $\mu_2 = 0$. This establishes the utility of the present study that the utilization of partial information and preliminary test increases the efficiency of the estimator. Graphically as $\mu_2$ is increases $e_1$ decreases to a minimum and then increases to unity. It is found that $e_1$ is very close to 1 at $\mu_2 = 1$. 
1.7 Optimum allocation.

In planning of a sample survey, a stage is always reached at which an important decision must be made about the size of the sample. Too large a sample implies a waste of resources, and too small a sample diminishes the precision of the estimators. Thus an optimum size of the sample is required so as to balance precision and cost involved in the survey. The optimum allocation of sample sizes are attained either by minimizing precision against a given cost or minimizing cost against given precision.

In obtaining optimum allocation of sample sizes for the proposed estimator, we consider a simple linear cost function $C$ given by

$$ C = c'n' + cn $$

where $c$ is the cost per unit of observing the variable $y$ and $c'$ is the cost per unit of observing the variable $x$, assuming that the cost per unit is the same for all strata.

In general the values of $\mu_x$ are unknown, the experimenter has partial information about it. When $\mu_x = 0$, the mean square error of $t_s$ is least and the relative efficiency is largest. Thus it would be reasonable to let $\mu_x = 0$ in the $MSE(t_s)$ and obtain the values of $n$ and $n'$ under the optimum situation of minimizing precision against a given cost.

For a specific cost $C^*$, routine mathematical derivation gives

$$ M_{opr}(t_s) = \left[ \sqrt{Kc} + \sqrt{c'K'} \right] / C^* $$
In a similar way the optimum allocation for the estimator $t_4$ is given by

$$M_{opt}(t_4) = \frac{(\sqrt{Kc} + \sqrt{K''c'})}{C}$$

where

$$K = 1 - \rho^2 \quad K' = \rho^2 \{\alpha + 2Z_o \varphi(Z_o)\} \quad \text{and} \quad K'' = \rho^2$$

Analytically it can be seen that $(\alpha + 2Z_o \varphi(Z_o))$ is a decreasing function of $Z_o$ with a maximum equal to unity at $Z_o = 0$. Therefore we can conclude that $M_{opt}(t_5) \leq MSE_{opt}(t_4)$ with equality holding for $Z_o = 0$ in which case the two estimators coincide.

Thus we have proved that under certain conditions, mean square error of a CRPTE in double sampling is smaller than the mean square error of combined regression estimator under double sampling.

1.8 Empirical Studies and Conclusion

Finally, in the last chapter empirical studies were made to show the applications of the proposed estimator as compared with other estimator. The empirical work is carried out with the help of both real life data as well as simulated data. Real life data were extracted from Rapid household survey – Reproductive child health (RHS-RCH project, phase 1, 1998). The data provides district wise demographic indicators for the Empowered Action Groups (EAG) States. From the data two distinct characteristics of the population were identified, namely complete child immunization and female literacy rate. Here the percent of complete child immunization is taken to be the dependent variable (Y) and female literacy rate is considered as an auxiliary variable (X). In the present analysis, the primary sample unit is set at district level and after
eliminating those districts in which data were not available partly or wholly the total size of the population reduces to \( N = 158 \). Since each state is governed by a distinct political and cultural system, which can have far reaching consequences on the economic, social status of the population within each state. This in turn can have an impact on the demographic characteristics of each state differently. So the data, on the selected variables can be homogeneous within each state and heterogeneous between states. Hence for the purpose the present study of the EAG states, each state is considered as a stratum. The data corresponding to different states and also the combined data are tested for normality and it was found that both \( X \) and \( Y \) follow normal distribution.

The partial information on \( X \) is obtained from census 1991 as female literacy rate is computed for the EAG states and given as \( \mu_0 = 26.4 \% \). After routine calculation, the estimates of the percentage of complete child immunization for the EAG states by the use of the combine regression preliminary test estimator in double sampling is 45.29\%. As partial checks, the true population value of the mean of the auxiliary variable is 45\%. Thus when reliable partial information on the mean of the auxiliary variable is not obtained as in present case, the hypothesis is rejected and as a result \( \ddot{x}_w \) is utilized in the estimation of \( \mu_x \) and consequently the proposed estimator reduces to the usual combined regression estimator under double sampling.

An attempt is also made to compute \( t_5 \) through the use of simulated data set. Bivariate normally distributed population data for different strata were
generated with the help of STATA 8.0 in which the input parameters are 
$(\mu_n, \mu_n, \sigma_n, \sigma_n, \rho)$. Four strata each of sizes $N_1=35$, $N_2=40$, $N_3=50$ and $N_4=45$ were considered and the stratification was done according to the mean of the value of the study variable $Y$. Further it is assume that there exist partial information about the mean of the auxiliary variable from certain sources and given by $\mu_0 = 75.0$. Again routine calculation for the estimate $\mu$, the CRPTE in double sampling is 135.4. As partial checks, the true population value of the mean of the auxiliary variable is 73. Thus when a reliable partial information of the mean of the auxiliary variable is available, then the $MSE(\mu)$ is smaller than the $MSE(\mu)$ and consequently this increases the efficiency of the proposed estimator.

Thus we see that the empirical study also supports the analytical work of the present study that under the stated assumptions the CRPTE in double sampling is more efficient than the usual combined regression estimator, when a reliable information about the mean of the auxiliary variable is available.
References


