

Some Model Assisted Estimators Using Functional Form Calibration Approach

Naila Alam^{1*}, Muhammad Hanif²

* Corresponding Author

1. Kinnaird College for Women, Lahore, Pakistan, naila.alam@kinnaird.edu.pk

2. National College of Business Administration and Economics, Lahore, Pakistan. drhanif@ncbae.edu.pk



Abstract

The Model assisted estimators are approximately design unbiased, consistent, provides robustness and reduce design variance if underlying model reasonably defines the regression relationship. If the model is mis specified, then model assisted estimators might result in an increase of the design variance but remain approximately design unbiased and show robustness against model-misspecification. The well-known model assisted estimators, generalized regression estimators are members of a larger class of calibration estimators. Calibration method generates calibration weights that meet the calibration constraints and have minimum distance from the sampling design weights. By using different distance measures, classical calibration approach generates different calibration estimators but with asymptotically identical properties. Later, the constraint of distance minimization was reduced for studying the properties of calibration estimators by proposing a simple functional form approach. This approach generates calibration weights that prove helpful to control the changes in calibration weights by using different functions of auxiliary variable's values. This paper is an extended work on model assisted approach by using functional form calibration weights. Some new model assisted estimators are considered to get efficient and stabilized regression weights by introducing a control matrix. The asymptotic un-biasedness of the proposed estimators is verified and the expressions for MSE are derived in three different cases. A simulation study is done to compare and evaluate the efficiency of the proposed estimators with some existing estimators of population total.

Key Words: Calibration approach, functional form of calibration approach, auxiliary variables, Model Assisted Estimators

1. Introduction

A well-known model assisted approach, the classical calibration approach by Deville and Särndal (1992) uses auxiliary information to produce efficient estimates of the population parameters. The approach uses weighted sample observations, whereas the weights are obtained such that the distance between these and design weights are minimized under the condition that they have to satisfy a calibration to benchmark constraints and therefore named as calibration weights.

Let y_k is the value of k th observation of the study variable and $\mathbf{x}'_k = (x_{1k}, x_{2k}, \dots, x_{pk})$ for $j = 1, 2, \dots, p$ is the auxiliary vector associated with y_k such that the population totals of the auxiliary variables are known prior to sampling. Let the population total of j th variable is $t_{xj} = \sum_U x_{jk}$ and a vector of population total(s) of p auxiliary variables is denoted by \mathbf{X} , where \sum_U is the sum of all $k \in U$ for $U = \{1, 2, \dots, N\}$ is known prior to estimation of population total of study variable. Also, vector of Horvitz Thompson estimators of population total(s) for auxiliary variables is $\hat{\mathbf{t}}_{x\pi} = \sum_s d_k x_k$, where \sum_s is the sum on $k \in s$.

The classical calibration estimator proposed by Deville and Särndal (1992) for estimating population total of study variable Y is defined as

$$\hat{t}_{yc} = \sum_s w_k y_k \quad (1.1)$$

They named the weights w_k , the calibration weights because these weights have minimum distance from the survey design weights and satisfy a calibration to benchmark constraints:

$$\sum_s w_k x_k = X \quad (1.2)$$

By using different distance functions many linear and nonlinear calibration estimators can be obtained. The chi square distance functions generate linear calibration weights that result in linear calibration estimators. The resulted calibration weights can be written in the form:

$$w_k = d_k(1 + q_k \lambda_s). \quad (1.3)$$

The $d_k = 1/\pi_k$ is the sampling design weight where q_k is any individual observation weight for k th observation. Also, λ_s is a vector of Lagrange multipliers and can be obtained from calibration constraints defined in (1.2). Using value of $\lambda_s = (\sum_s d_k q_k x_k x'_k)^{-1}(X - \hat{t}_{x\pi})$, the calibration estimator (1.1) can also be seen as an approximate general linear regression estimator,

$$\hat{t}_{yc} = \hat{t}_{y\pi} + \hat{b}_{ws}(X - \hat{t}_{x\pi}) \quad (1.4)$$

The $\hat{t}_{y\pi} = \sum_s d_k y_k$ is the Horvitz Thompson estimators for population total of study variable and \hat{b}_{ws} is a vector of order $p \times 1$ defined as $\hat{b}_{ws} = (\sum_s d_k q_k x_k x'_k)^{-1}(\sum_s d_k q_k x'_k y_k)$ with elements $b_{wj} = (\sum_s d_k q_k x_{jk} x'_{jk})^{-1}(\sum_s d_k q_k x'_{jk} y_k)$ where $x'_{jk} = (x_{j1}, x_{j2}, \dots, x_{jn})$.

To study the properties of calibration estimators in general, Estevao and Särndal (2000) proposed a functional form of calibration weights. They proposed weights w_{kCALF} that had mathematical form and by defining two parameters produce different weight systems.

They specified a vector $\mathbf{z}'_k = (z_{1k}, z_{2k}, \dots, z_{pk})$ for every $k \in s$ such that

- (a) $\text{Dim}(\mathbf{x}_k) = \text{Dim}(\mathbf{z}_k) = p$ and
- (b) The matrix $\mathbf{X}'\mathbf{X}$ of order $p \times p$ is non-singular

The components of vector \mathbf{z}_k are functions of \mathbf{x}_k and can be defined as

$$\mathbf{z}'_k = (x_{1k}^{m-1}, x_{2k}^{m-1}, \dots, x_{pk}^{m-1}) \quad \text{for } m \geq 0 \text{ and } x_{jk} > 0 \quad (1.5)$$

Also, for $m = 2$, we have $\mathbf{z}_k = \mathbf{x}_k$

The resulted calibration weights are asymptotically design unbiased.

$$w_{kCALF} = d_k(1 + q_k \lambda_s) \quad (1.6)$$

Where the parameters q_k and \mathbf{z}_k are chosen to satisfy (a) and (b) and the vector $\lambda_s = (\sum_s d_k q_k \mathbf{z}_k \mathbf{z}'_k)^{-1}(X - \hat{t}_{x\pi})$ is determined by the calibration constraints.

The $\mathbf{X} = \sum_s w_k x_k = \sum_U x_k$ vector is the vector of population totals for j auxiliary variables and assumed to be known prior to estimation. The final form of w_{kCALF} can be obtained using λ_s in (1.6), that is

$$w_{kCALF} = d_k + (X - \hat{t}_{x\pi})' R_k \quad (1.7)$$

$$\text{Where } R_k = (\sum_s d_k q_k \mathbf{z}_k \mathbf{z}'_{,k})^{-1} d_k q_k \mathbf{z}_k \quad (1.8)$$

The resulted functional form of calibration estimator can be obtained as

$$\hat{t}_{CALF} = \sum_s w_{kCALF} y_k \text{ and can be written in the form } \hat{t}_{CALF} = \hat{t}_{y\pi} + (\mathbf{X} - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{D}}_k \quad (1.9)$$

$$\text{Where } \hat{\mathbf{D}}_k = (\sum_s d_k q_k \mathbf{z}_k \mathbf{z}_k')^{-1} (\sum_s d_k q_k \mathbf{z}_k' y_k) \quad (1.10)$$

The functional form \hat{t}_{CALF} generates variety of calibration weights when different values of \mathbf{q}_k and \mathbf{z}_k are used, therefore is more convenient and flexible. Also, the functional form of calibration estimator is asymptotically equivalent to instrumental calibration estimators proved by Kim and Park (2014).

Due to the flexibility and formation of the functional form calibration weights, these can helpful to control the undesirable effects of auxiliary variables by using different choices of \mathbf{q}_k and \mathbf{z}_k . The functional form of calibration estimator \hat{t}_{CALF} defined in (1.9) is a linear function of the design weights and the adjustment term. For a strong to moderate linear relationship between the study variable and the auxiliary variable(s) the adjustment term will be equal to the error of the design weights with opposite signs and therefore give minimum mean square error but presence of discrepant or multi-collinear auxiliary variable(s) may cause of negative, inefficient or extreme weights. Estevao and Särndal (2002) have shown that the purpose is not to put the given auxiliary information blindly in estimation but efficient use of auxiliary information is important. The efficient use of auxiliary information in model assisted approaches can result in substantial gain in precision. According to Ståhl et.al (2016), the main advantage of model-assisted estimation is that it does not rely totally on the suitability of the model, but the model only helps to improve the precision of an estimator and in case of failure of assumed model, the resulting estimator remains asymptotically unbiased. The model assisted estimators have been proved efficient and robust under different scenarios. Kim and Rao (2012) used model assisted approach in integrating the data sets from two independently conducted surveys. Breidt (2017) used the model-assisted approach from a complex survey together with auxiliary information to estimate finite population parameters. They reviewed a very broad class of prediction methods including linear models, linear mixed models, nonparametric regression and machine learning techniques. Also, different functions of auxiliary variables have been considered by researchers to improve efficiency of the model assisted estimators. Kumar et.al (2017) considered the case when study variable and auxiliary variable are inversely related and have shown that two calibration approaches provide the different variances.

Gard (2019) explored that model assisted estimators generates the most accurate estimates if relevant auxiliary variables that explain nonresponse and the target variables are available. The efficiency of the estimator depends not only on good and related auxiliary information but also on the methods by which the information has been utilized in estimation. Eric (2020) also used model assisted approach in estimating model parameters in circumstances of a complex survey. Recently, Ben et.al (2021) considered the case of obtaining calibration weights when covariates are high dimensional and especially when interactions between variables exist. They proposed a multilevel calibration weighting system that satisfies strict calibration constraints for main weights and loosen calibration constraints for higher-order interactions. Also, the asymptotic properties of these estimators were developed and assessed.

2. The New Functional Form of Calibration Weights:

In this paper, some new model assisted estimators are proposed using a new functional form of calibration weights that can be used to control unwanted effects of auxiliary variable(s) and to stabilize the calibration weights for precise estimation and prediction. The proposed functional form of calibration estimators is

$$w_{kNCALF} = d_k + \mathbf{C}_p d_k q_k \mathbf{z}_k \boldsymbol{\lambda}_s \quad (2.1)$$

Here the \mathbf{C}_p is a matrix of order $\times p$. We called it control matrix because it will be used to control the effect of auxiliary variables in model assisted estimation. The $\boldsymbol{\lambda}_s$ is obtained from the calibration constraints (1.2) and vector \mathbf{z}_k has been defined in (1.5).

This method like the ridge regression approach is a tradeoff between bias and variance. To stabilize the calibration weights and to prevent them from unwanted effects of the auxiliary variable(s) a control matrix is used, but unlike the ridge approach the penalty factor is not added to the $\mathbf{X}'\mathbf{X}$ matrix; instead the adjustment terms are controlled explicitly by using a control matrix.

We can simplify the new functional form of calibration weights defined in (2.1) using the vector λ_s :

$$w_{k,NCALF} = d_k + \mathbf{C}_p(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})' \mathbf{R}_k \quad (2.2)$$

Where R_k is defined in (1.8). The new functional form of model assisted estimators for population total can be obtained using functional form calibration weights defined in (2.2):

$$\hat{t}_{NCALF} = \sum_s w_{k,NCALF} y_k = \hat{t}_{y\pi} + \mathbf{C}_p(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{D}}_k \quad (2.3)$$

We have defined $\hat{\mathbf{D}}_k$ in (1.10). The different model assisted estimators can be obtained by using different values of \mathbf{C}_p , q_k and \mathbf{z}_k in (2.3) that results in a wider class of regression type estimators.

To obtain the different functional form model assisted estimator, we need to specify in (2.3):

- (i) A matrix \mathbf{C}_p of order $p \times p$ such that the resulted estimator remains asymptotically unbiased under the randomization distribution.
- (ii) Auxiliary variables we want to calibrate,
- (iii) To specify the value of m to get vector of instrumental variables and
- (iv) The value of q_k .

The functional form of the calibration estimators proposed by Estevao & Särndal (2000) and the General Regression Estimator (GREG) estimators are the special cases of the proposed estimator and can be defined as a subset of the proposed functional form calibration estimators, i.e.

$$\hat{t}_{GREG} \subseteq \hat{t}_{CALF} \subseteq \hat{t}_{NCALF}$$

A specific case when \mathbf{C}_p is an identity matrix of order $p \times p$ and $m = 2$ in vector \mathbf{z}_k is the classical calibration estimator proposed by Deville and Särndal (1992) for estimating population total of study variable y .

$$\hat{t}_{NCALF(S)} = \hat{t}_{y\pi} + \mathbf{b}'_{ws}(\mathbf{X} - \hat{\mathbf{t}}_{x\pi}) = \hat{t}_{yc} \quad (2.4)$$

The objective is to find the values that prove helpful in controlling the undesirable variability in calibration weights and to get stabilize weights.

For example, total number of deaths by COVID-19 in Pakistan using sample data can be estimated when confirmed and recovered cases are used as auxiliary variables and their population totals are known prior to estimation but due to high variation within the data (daily observations of confirmed and recovered cases within and after the wave) can cause extreme calibration weights and may effects the estimation of total deaths. In this case, instead of using $m = 2$ in \mathbf{z}_k , some other values of m maybe a good choice to generate optimum weights and estimates. Also, in some cases when multiple auxiliary variables are available for estimation, it may helpful to assign weights to these auxiliary variables according to their correlation with the study variable. An example can be estimation of total population on the basis of household size, employed persons and tax payers from a previous survey but size of a household is highly correlated with the variable “population size” as compared to other variables and therefore may get large weight in estimating total population.

We are free to made different choices of the matrix \mathbf{C}_p for which the calibration weights remains asymptotically unbiased. We consider three different cases for which the resulting calibration estimators remain asymptotically unbiased.

CASE I:

The efficiency of the model assisted estimators depends on the correlated auxiliary variables, used in estimation procedure. In most surveys the complete information about the variance of the study variable is not available and available auxiliary information define only a portion of the variability. The correlation coefficient of determination is a measure of variation that is explained by the auxiliary variables used in a model.

In the first case we consider $C_{p1} = R^2$ a matrix of order 1×1 where R^2 is the adjusted coefficient of determination. The functional form model assisted estimator in this case will be

$$\hat{t}_{NCALF1} = \hat{t}_{y\pi} + R^2(X - \hat{t}_{x\pi})'\hat{D}_k \quad (2.5)$$

The value of adjusted R^2 will assign an appropriate weight to the adjustment term according to the proportion of variability explained by the auxiliary variables. If the value of adjusted $R^2 \approx 0$ the adjustment term will also tend to zero and the auxiliary variable(s) automatically will be excluded from the model and only design based Horvitz Thompson estimator will be used for estimation. For the case when the linear relations of the study variable are perfectly defined by the use of the auxiliary variable(s), the value of $R^2 \approx 1$ and there will be no change in calibration weights.

CASE II:

In today's world bulk of information is available and multiple auxiliary variables are used in estimation to increase precision and accuracy of the estimates. When a large number of auxiliary variables are used in estimation, a case of obtaining inefficient weights may be that some of the less correlated auxiliary variable(s) has large sampling errors and therefore influenced the calibration weights and result in unstable or inefficient weights. To avoid such situations the auxiliary variables can be weighted according to their correlation with the study variable.

Therefore one possible value for control matrix C_p can be:

$$C_{pr} = |r_{kj}| I_{p \times p} \quad (2.6)$$

C_{pr} is a diagonal matrix of order p having partial correlation coefficients values r_{kj} as its diagonal elements, where r_{kj} is the partial correlation coefficient value of X_j with y . The norm is used to avoid the negative sign. The matrix C_{pr} will assign weights to the sampling error(s) of auxiliary variable(s) according to their correlation with the study variable and therefore a more correlated auxiliary variable will get large weight and a less correlated variable will be assigned small weight in construction of the calibration weights. If an auxiliary variable has perfect correlation with the study variable, then its correlation coefficient value will be near to 1 and therefore its residual will remain same. Similarly, if an explanatory variable is not correlated with study variable, the r_{kj} value will be approximately zero and in this case the variable will be excluded from the model. Some of these correlations coefficient's values (of p variables included in the estimation) can be obtained from any previous survey and others can be obtained at the sample level. The estimator of population total in this case will be:

$$\hat{t}_{NCALF2} = \hat{t}_{y\pi} + (X - \hat{t}_{x\pi})'C_{pr}\hat{D}_k \quad (2.7)$$

CASE III:

We consider another case when inefficient weights are occurring because some of the auxiliary variable's values are more disperse as compared to study variable and therefore may have large sampling errors. This may influence the calibration weights and may cause of inefficient or extreme weights. The matrix given below may be useful to assign appropriate weights to auxiliary variables by comparing its variation with study variables i.e.

$$\hat{C}_{pc0} = C V^2_{y.xj} I_{p \times p} \quad \text{for} \quad j = 1, 2, \dots, p \quad (2.8)$$

The \mathbf{C}_{pC0} matrix in this case will be a diagonal matrix having $C V^2_{y.xj} = \frac{C^2_y}{C^2_{xj}}$ as diagonal elements where $C V^2_{y.xj}$ is the ratio of C^2_y (square of the coefficient of variation of study variable) and C^2_{xj} (square coefficient of variation of the j th auxiliary variable). The elements of the control matrix will assign appropriate weight to the sampling error of each auxiliary variable by comparing its variance with study variable. The functional form of model assisted estimator in this case will be:

$$\hat{t}_{NCALF3} = \hat{t}_{y\pi} + (\mathbf{X} - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{C}}_{pC0} \hat{\mathbf{D}}_k \quad (2.9)$$

The Asymptotic unbiased-ness and mean square error (MSE) of the proposed estimators (2.5), (2.7) and (2.9) are derived with the help of following results.

The error associated with adjustment term $\hat{t}_{y\pi}$, $\hat{\mathbf{t}}_{x\pi}$ and $\hat{\mathbf{D}}_k$ is denoted by $\mathbf{e}_x = (\hat{\mathbf{t}}_{x\pi} - \mathbf{X})$, $\mathbf{e}_y = (\hat{t}_{y\pi} - Y)$ and $\mathbf{e}_d = (\hat{\mathbf{D}}_k - \mathbf{D}_N)$ where $\hat{\mathbf{D}}_k$ is an estimator of \mathbf{D}_N . Also, we have the results $E(e_{xj}^2) = \theta X_j^2 C_{xj}^2$, $E(e_y^2) = \theta Y^2 C_y^2$ and $E(e_{xj} e_y) = \theta Y X_j C_{xj} C_y \rho_{y.xj}$ where $\rho_{y.xj} = \frac{s_{y.xj}}{s_{xj} s_y}$.

We assume that \mathbf{e}_x and \mathbf{e}_y are smaller magnitude as compare to \mathbf{X} (auxiliary variable(s) Total) and Y (study variable total), also X_j (population total of j th auxiliary variable) is positive and $\left| \frac{e_{xj}}{X_j} \right| < 1$

Result 1:

The calibration estimator (2.5) is asymptotically unbiased and the bias related to it is

$$Bias_{\hat{t}_{NCALF1}} = -E\{(\hat{R}^2 - R_N^2)(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})'(\hat{\mathbf{D}}_k - \mathbf{D}_N)\}$$

Using results Fuller and Isaki (1982) and Estevao and Sarndal (2000), the bias of the quantities $N^{-1}(\hat{\mathbf{D}}_k - \mathbf{D}_N)$, $N^{-1}(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})$ and $N^{-1}(\hat{R}^2 - R_N^2)$ are of order $(\frac{1}{\sqrt{n}})$ and hence bias of the estimator is of $O(\frac{1}{\sqrt{n}})$ which is negligible when n is large and even for a modest sample size.

The Mean Square Error (MSE) of the estimator (2.5) is:

$$MSE_{\hat{t}_{NCALF1}} = \theta \left[Y^2 C_y^2 + \frac{1}{J} (R^2)^2 \sum_{j=1}^J D_j^2 X_j^2 C_{xj}^2 - \frac{2}{J} Y C_y R^2 \sum_{j=1}^J D_j X_j C_{xj} \rho_{y.xj} \right]$$

Result 2:

The estimator (2.7) is asymptotically unbiased as $N^{-1}(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})$ has bias of order $(\frac{1}{\sqrt{n}})$, also the $\mathbf{C}_{pr} = \text{diag}\{r_{kj}\}$ is a diagonal matrix having r_{kj} at its diagonal such that $0 \leq r_{kj} \leq 1$ for $j = 1, 2, \dots, p$ and hence estimator (2.7) is asymptotically unbiased estimator as the term $N^{-1}(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})' \mathbf{C}_{pr} \hat{\mathbf{D}}_k \rightarrow \mathbf{0}$ as $n \rightarrow \infty$

The MSE expression of estimator (2.7) is derived as:

$$MSE_{\hat{t}_{NCALF2}} = \theta \left[Y^2 C_y^2 + \frac{1}{J} \sum_{j=1}^J D_j^2 X_j^2 C_{xj}^2 \rho_{y.xj}^2 - \frac{2}{J} Y C_y \sum_{j=1}^J D_j X_j C_{xj} \rho_{y.xj}^2 \right]$$

Result 3:

The estimator (2.9) is asymptotically unbiased with bias

$$\text{Bias}_{\hat{t}_{NCALF3}} = -E\{(\hat{\mathbf{D}}_k - \mathbf{D}_N)(\hat{\mathbf{C}}_{pC0} - \mathbf{C}_{pC0})(\mathbf{X} - \hat{\mathbf{t}}_{x\pi})'\}$$

Where $\hat{\mathbf{C}}_{pC0}$ is a matrix of order $p \times p$ estimating the matrix \mathbf{C}_{pC0} . The diagonal elements of $\hat{\mathbf{C}}_{pC0}$ matrix are $C V^2_{y.xj} = \frac{C^2_y}{C^2_{xj}}$. The CV has bias of $O(\frac{1}{n})$ and hence the estimator has bias of $O(\frac{1}{n^2})$.

The MSE of estimator (2.9) is derived as:

$$\text{MSE}_{\hat{t}_{NCALF3}} = \theta C_y^2 \left[Y^2 + C_y^2 + \frac{1}{J} \sum_{j=1}^J D_{Nj}^2 \frac{X_j^2}{C_{xj}^2} - \frac{2}{J} Y C_y \sum_{j=1}^J D_{Nj} \frac{X_j}{C_{xj}} \rho_{y.xj} \right]$$

3-Simulation Study:

We examine and compare the performance of the proposed estimators \hat{t}_{NCALF1} , \hat{t}_{NCALF2} and \hat{t}_{NCALF3} defined in (2.5), (2.7) and (2.9) respectively through a simulation study with classical Horvitz Thompson estimator ($\hat{t}_{y\pi}$) and with the special case of the functional form calibration estimator (\hat{t}_{CALF}) proposed by Estevao & Särndal (2000, 2002) defined in (1.9). The efficiency is compared through bias and Mean square Error. A population consists of 1000 normally distributed values of dependent variable (Y) is generated. The correlated auxiliary information is generated in two cases, in first case only one moderately linearly related variable (\mathbf{X}_1) with the study variable is generated such that $m = 2$ or $\mathbf{z}_k = \mathbf{x}_k$. The Bias and MSE are calculated on repeated sampling of 500 samples each of size 100. Similarly the process is repeated for two auxiliary variables; one auxiliary variable is generated taking ($0.65 < \rho < 0.85$) such that the generated variable is moderately correlated with study variable whereas second variable is generated by fixing the value of correlation coefficient between (0.25 to 0.30). In particular, the purpose is to study the impact of including less correlated auxiliary variables in model assisted estimation.

The population total of 1000 values $\sum_{i=1}^{1000} Y_i = 9970.055$ is estimated using the five estimators. The table below is showing the results of simulation study.

Table 1: Table showing results of simulation study for comparison of the proposed estimators:

Estimators	One Auxiliary variable			Two auxiliary variables		
	Estimate	Bias	MSE	Estimate	Bias	MSE
$\hat{t}_{y\pi}$	9990.50	12.44514	143.577	9990.50	12.44514	143.557
\hat{t}_{CALF}	9961.371	-11.68338	136.5013	9962.956	-11.324	135.867
\hat{t}_{NCALF1}	9978.995	10.94009	135.6836	9978.395	10.356	135.145
\hat{t}_{NCALF2}	9978.208	9.15232	134.86228	9977.410	8.452	133.856
\hat{t}_{NCALF3}	9979.395	9.8999	135.80934	9978.02	9.324	134.986

The results of the simulation study show that the estimator \hat{t}_{CALF} under-estimate the total of the study variable as compared to $\hat{t}_{y\pi}$ (Horvitz Thompson's estimator) and the three considered cases of the functional form model assisted estimators. However, when only one auxiliary variable is used, the decrease in bias and Mean Square Error (MSE) can be observed when the proposed functional form model assisted estimators are used to estimate population total under the condition when response variable follow a normal distribution. An interesting result can be observed that by choosing control matrix $\mathbf{C}_{pr} = r_{ij} \mathbf{I}_{p \times p}$, where \mathbf{C}_{pr} is a diagonal matrix (consists of partial correlations of X_j with y as diagonal elements) yields optimum results among all estimators considered for comparison. Also, for the case of two auxiliary variables when one is weakly correlated with the response variable the \hat{t}_{NCALF2} again produces the minimum bias and MSE, the reason may be that control matrix assign weights to the sampling errors of the auxiliary variables that are proportional to their correlation with the study variable and hence second auxiliary variable get less weight as compared to the first variable in generation of calibration weights and in

estimation. Also, we can examine that the other two proposed estimators are also more efficient than Horvitz Thompson ($\hat{t}_{y\pi}$) and functional form calibration estimator proposed by Estevao & Särndal (2000) (\hat{t}_{CALF}) but yield slightly larger bias and MSE than \hat{t}_{NCALF2} when the data is generated through normal distribution.

4-Conclusion:

The utility of a given estimator depends on what auxiliary data are available and how it is used in estimation. The auxiliary variables having large variances and discrepant values may effects estimation process. To control the unwanted effects of auxiliary variables and for optimal use of the available information, a new functional form of calibration estimator is proposed to estimate population total which result in a family of model assisted estimators. Also, the bias and MSE expressions of the three functional form calibration estimators are derived and a simulation study is conducted to assess the performance of the proposed estimators. The result demonstrates that the proposed estimators perform well as compared to Horvitz Thompson and functional form calibration estimators proposed by Estevao and Sarndal (2000). As the results are obtained under a single condition, when random error follow a normal distribution and auxiliary variables are moderately or weakly correlated with the study variables, therefore, more work may be needed to develop specific guidelines for various type of auxiliary information available for estimation. A further study can explore performance of different cases of the proposed functional form under various other probability distributions. The important question is that for more than two auxiliary variables which case of control matrix may result better. Also, for auxiliary variables having large variances what will be the role of control matrix and which estimator will perform well.

References:

1. Ben-Michael, Eli & Feller, Avi & Hartman, Erin. (2021). Multilevel calibration weighting for survey data. <https://arxiv.org/abs/2102.09052>.
2. Breidt, F. & Opsomer, Jean. (2017). Model-Assisted Survey Estimation with Modern Prediction Techniques. *Statistical Science*. 32. 190-205. 10.1214/16-STS589.
3. Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, 87, 376-382.
4. Eric V. Slud1, 2 (2020). Model-Assisted Estimation of Mixed-Effect Model Parameters in Complex Surveys, Research Report by *Center for Statistical Research & Methodology Research and Methodology Directorate U.S. Census Bureau Washington, D.C.* 20233
5. Estevao, V.M. and Särndal, C.E. (2000). A functional form approach to calibration. *J. Off. Statist.*, 16, 379-399.
6. Estevao, V.M. and Särndal, C.E. (2002). The ten cases of auxiliary information for calibration in two phase sampling. *J. Off. Statist.*, 18, 233-255.
7. Fuller, W.A. and Isaki, C.T. (1982). Survey design under the regression super population model. *J. Amer. Statist. Assoc.*, 77, 89-96.
8. Gard, R. (2019). Design-Based And Model-Assisted Estimators Using Machine Learning Methods - Exploring the k-Nearest Neighbor method applied to data from the Recreational Fishing Survey, *Master thesis, 15 ECTS Örebro University, School of Business*.
9. Kim, J. K. and Rao, J. N. K. (2012). Combining data from two independent surveys: a model-assisted approach, *Biometrika* 99: 85–100.
10. Kumar. S., Sisodia, B.V.S., Singh, D. and Basak, P. (2017). Calibration approach based estimation of finite population total in survey sampling under super population model when study variable and auxiliary variable are inversely related, *Journal of Reliability and Statistical Studies*; ISSN (Print): 0974-8024, (Online): 2229-5666 Vol. 10, Issue 2, 83-93.
11. Park, S. and Kim, J.K. (2014). Instrumental-Variable calibration estimation in survey sampling. *Statistica Sinica*, 24, 1001-1015.
12. Ståhl, G., Saarela*, S., Schnell, S., Holm, S., Breidenbach, J., Sean P. H., Paul, L., Magnussen, S., Næsset, E., McRoberts³ and Gregoire, T. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation, *Forest Ecosystem*, DOI 10.1186/s40663-016-0064-9.