Mean Estimation Using Qualitative Prior Information under Stratified Random Sampling Scheme

Abstract

Adapting [14] and [15], we propose modified classes of estimators for the unknown mean of study variable when auxiliary information is available in the form of attribute. The bias and mean square error of the estimators belonging to the classes are obtained and the expressions for the optimum parameters minimizing the asymptotic mean square error (MSE) are given in closed form. An empirical study is carried to judge the merits of the suggested estimators over [2], regression estimator and other known estimators. Both theoretical and empirical study results present the soundness and usefulness of the suggested estimators in practice.

Keywords: Proposed class, Regression estimator, Stratified random sampling scheme, Efficiency comparisons.

1. Introduction

In sample survey, prior information on the finite population is quite often available from previous experience, survey or administrative databases. Wide literature of basic statistics was given by [7]. The sampling literature describes a wide variety of techniques for using auxiliary information to improve the sampling design and/or obtain more efficient estimators. It is well known that when auxiliary information is to be used at the estimation stage, the ratio, product, regression methods are widely employed. Many authors such as [3] and [5] have suggested a difference class of estimator and a combined ratio-product type estimators using stratified random sampling, where the auxiliary information is in the form of auxiliary variates.

In many practical situations instead of existence of auxiliary variables, there exist some auxiliary attributes \( \phi \) (say), which are highly correlated with the study variable \( y \) such as

(a) The height of a person (\( y \)) may depend on gender (\( \phi \))
(b) Amount of milk produced (\( y \)) may depend on particular breed of the cow (\( \phi \))
(c) The yield of wheat crop produced may depend on a particular variety of wheat, etc.

Taking into consideration the point biserial correlation coefficient between auxiliary attribute \( \phi \) and the study variable \( y \), several authors including [4, 6, 8–13] envisaged a large number of improved estimators for the population mean \( \bar{Y} \) of the study variable \( Y \).

There are some situations when in place of one, we have information on two qualitative variables and in support of such type of study, authors including [14–16] have suggested different classes of estimators using two auxiliary attributes in stratified random sampling. Here, we assume that both auxiliary attributes have significant bi-serial correlation with the study variable \( y \) and there is significant phi-correlation between the two auxiliary attributes. In this article, we proposed different classes using single and multi-auxiliary attributes and compared them with some well-known estimators in literature.
Consider a finite population U of size N, stratified into L strata with i th stratum containing \( N_i \) units, where \( i = 1, 2, 3...L \). Let a simple random sample of size \( n_i \) be drawn without replacement from the i th stratum such that
\[
\sum_{i=1}^{L} n_i = n
\]
Let \((y_{ij}, \phi_{kij})\) \( (k = 1, 2) \) denote the values of the study variable \( (y) \) and the auxiliary attributes \( \phi_k \) \( (k = 1, 2) \) respectively in the i th stratum.

Suppose
\[
\phi_{kij} = 1, \text{ if } j^{th} \text{ unit in the stratum } i \text{ possesses the attribute } \phi_k
\]
\[
= 0, \text{ otherwise}
\]
Let \( p_k = \sum_{i=1}^{L} W_i p_{ki} \) and \( P_k = \sum_{i=1}^{L} W_i P_{ki} \) be the sample and population proportions of the units of the auxiliary attributes
where
\[
p_{ki} = \frac{a_{ki}}{n_i}, \quad P_{ki} = \frac{A_{ki}}{n_i}
\]
\[
A_{ki} = \sum_{j=1}^{N_i} \phi_{kij}, \quad a_{ki} = \sum_{j=1}^{N_i} \phi_{kij}
\]
Large sample properties of the classes are obtained according to stratified random sampling scheme. In so doing, we use the following notations in the rest of the article:

\[
\bar{y}_{si} = \sum_{i=1}^{L} W_i \bar{y}_i = \bar{Y}(1 + d_0)
\]
\[
p_{1st} = \sum_{i=1}^{L} W_i p_{li} = P_1(1 + d_1)
\]
\[
p_{2st} = \sum_{i=1}^{L} W_i p_{2i} = P_2(1 + d_2)
\]
Such that, \( E(d_0) = E(d_1) = E(d_2) = 0 \)

Also \( E(d_i) = \frac{\sum_{i=1}^{L} W_i y_{i}^2 g_i S_{p,i}^2}{P_i} = r_{00} \)
\[
E(d_1) = \frac{\sum_{i=1}^{L} W_i y_{i}^2 g_i S_{p,i}^2}{P_1} = r_{10}
\]
\[
E(d_2) = \frac{\sum_{i=1}^{L} W_i y_{i}^2 g_i S_{p,i}^2}{P_2} = r_{20}
\]
\[
E(d_{i}^2) = \frac{\sum_{i=1}^{L} W_i y_{i}^2 g_i S_{p,i}^2}{P_i} = r_{00} \]

Where
\[
S_{yi}^2 = \frac{\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2}{(N_i - 1)}, \quad S_{p,i}^2 = \frac{\sum_{j=1}^{N_i} (\phi_{kij} - P_{ki})^2}{(N_i - 1)}
\]
be the population variance of the study variable and the auxiliary variables respectively in the i th stratum.

And
\[
\frac{\bar{Y}_i}{P_{1st}} = \frac{\sum_{j=1}^{N_i} \frac{y_{ij}}{N_i}}{P_{1st}} = \frac{\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)(\phi_{kij} - P_{ki})}{(N_i - 1)}
\]
\[
\rho_{\phi \phi_i} = \frac{S_{\phi_{kij}}}{S_{\phi_i} S_{\phi_k}}
\]
be the population bi-covariance and point bi-serial correlation between the study variable and the auxiliary attributes, respectively in the i th stratum. Also
\[
S_{\phi_{kij}} = \frac{\sum_{j=1}^{N_i} (\phi_{kij} - P_{li})(\phi_{2ij} - P_{2i})}{(N_i - 1)}
\]
\[
\rho_{\phi_{kij}} = \frac{S_{\phi_{kij}}}{S_{\phi_{kij}} S_{\phi_{kij}}}
\]
be the population bi-covariance and phi-correlation coefficient between the auxiliary attributes in the i th stratum respectively.

2. Estimators in Literature

Koyuncu [14] proposed the combined ratio estimator of population mean in stratified random sampling using auxiliary attribute, given by
\[
\hat{Y}_R = \left[ \frac{\bar{y}_{si}}{P_{1st}} \right] P_i
\]
The bias and MSE of \( \hat{Y}_R \), to the first order of approximation, are respectively given by
\[ B(\hat{Y}_R) = \overline{Y}(r_{20} - r_{110}) \]  
(2)

\[ \text{MSE}(\hat{Y}_R) = \overline{Y}^2(r_{200} + r_{20} - 2r_{110}) \]  
(3)

Following [6], a general combined class in stratified random sampling is defined by Koyuncu [14] as

\[ \hat{Y}_{g(st)} = g(\overline{Y}_st, U_st) \]  
(4)

where \( U_st = \frac{p_{st}}{p_1} \) and \( g(\overline{Y}_st, U_st) \) is a function of \( \overline{Y}_st \) and \( U_st \).

The bias and MSE of \( \hat{Y}_{g(st)} \), to the first order of approximation, are given by

\[ B(\hat{Y}_{g(st)}) = \left[ g_2 r_{200} + g_3 \overline{Y}_{110} + g_4 \overline{Y}^2 r_{200} \right] \]  
(5)

for \( (g_2, g_3, g_4) \) see [14].

\[ \text{MSE}(\hat{Y}_{g(st)}) = \overline{Y}^2 r_{200} + \overline{g}_1^2 r_{200} + 2 \overline{g}_1 \overline{Y}_{110} \]  
(6)

By using optimum value of \( g_1^* = \frac{-\overline{Y}_{110}}{r_{200}} \), the minimum MSE of the estimators in the class \( \hat{Y}_{g(st)} \) is found as

\[ \text{MSE}(\hat{Y}_{g(st)}) = \overline{Y}^2 r_{200} \left[ 1 - \frac{r_{110}^2}{r_{200} r_{200}} \right] \]  
(7)

\[ \mu_{1st}^* = \frac{r_{200}}{r_{200} (1 + r_{200}) - r_{110}} \]

Minimization of \( \text{MSE}(\hat{Y}_{RA}) \) is achieved for the optimum choice of the constants \( \mu_{1st} \) and \( \mu_{2st} \)

or

\[ \text{MSE}(\hat{Y}_{g(st)}) = \overline{Y}^2 r_{200} \left[ 1 - \kappa_{g(st)} \right] \]  
(8)

where \( \kappa_{g(st)} = -\frac{r_{110}^2}{r_{200} r_{200}} \) (say)

A combined regression estimator in stratified random sampling using auxiliary attribute, is given by

\[ \hat{Y}_{\beta_{p1}} = \overline{Y}_{st} + b_{\beta_{p1}} (P_1 - p_{1st}) \]  
(9)

The MSE of \( \hat{Y}_{\beta_{p1}} \), to the first order of approximation, given by

\[ \text{MSE}(\hat{Y}_{\beta_{p1}}) = \overline{Y}^2 \left[ r_{200} + \frac{\beta_{p1}^2 r_{200} - 2 \beta_{p1} r_{110}}{R_1} \right] \]  
(10)

where \( \beta_{p1} = \sum_{i=1}^{L} W_i g_i S_{p1i} / \sum_{i=1}^{L} W_i g_i S_{p1i}^2 \)

A simple general class of biased regression estimators has been proposed by [2]. Thus following [2], we suggested the same in stratified random sampling using auxiliary attribute as

\[ \hat{Y}_{RA} = \mu_{1st} \overline{Y}_{st} - \mu_{2st} (P_1 - p_{1st}) \]  
(11)

with \( \mu_{1st} \) and \( \mu_{2st} \) constants to be properly determined.

The MSE of \( \hat{Y}_{RA} \), up to the first order of approximation, given by

\[ \text{MSE}(\hat{Y}_{RA}) = \overline{Y}^2 \left[ 1 + \mu_{1st}^2 (1 + r_{200}) + \frac{\mu_{2st}^2 r_{200}}{R_1} - 2 \mu_{1st} + 2 \mu_{1st} \mu_{2st} \frac{r_{110}}{R_1} \right] \]  
(12)

Minimization of \( \text{MSE}(\hat{Y}_{RA}) \) is achieved for the optimum choice of the constants \( \mu_{1st} \) and \( \mu_{2st} \)

\[ \mu_{2st}^* = \frac{r_{110} R_1}{r_{110}^2 - r_{200} (1 + r_{200})} \]

Singh [17] proposed a family of estimators for the population mean in stratified random sampling using information on auxiliary attribute, as

\[ \hat{S}_s = \left[ w_1 \overline{Y}_{st} + w_2 (P_1 - p_{1st}) \right] \exp \left[ \frac{\eta_{1st} (P_1 - p_{1st})}{\eta_{2st} (P_1 + p_{1st}) + 2 \lambda_{st}} \right] \]  
(13)

where \( w_1 \) and \( w_2 \) are suitable constants.
MSE of \( \hat{t}_{SS} \) see [17], up to the first order of approximation, given as

\[
\text{MSE}(\hat{t}_{SS}) = \bar{Y}^2 \left[ 1 - \frac{r_{02}^2 + 3r_{02}r_{110} + r_{200}^2 + 2r_{200}r_{110} - 2r_{02}r_{110} - r_{02}r_{200}r_{110}^2}{r_{02}^2 + r_{02}r_{200}^2 + r_{110}^2 + 2r_{02}r_{200}^2} \right] \tag{13}
\]

or

\[
\text{MSE}(\hat{t}_{SS}) = \bar{Y}^2 \left[ 1 - \kappa_{SS} \right] \tag{14}
\]

Where

\[
\kappa_{SS} = \frac{r_{02}^2 + 3r_{02}r_{110} + r_{200}^2 + 2r_{200}r_{110} - 2r_{02}r_{110} - r_{02}r_{200}r_{110}^2}{r_{02}^2 + r_{02}r_{200}^2 + r_{110}^2 + 2r_{02}r_{200}^2} \tag{15}
\]

(say) and \( \gamma = \frac{\eta_{st} P_1}{2(\eta_{st} P_1 + \lambda_{st})} \). For detail see [13]. There are some situations where in place of one; we have information on two qualitative variables. Using multi-auxiliary information Koyuncu [14] proposed the following estimator

\[
\hat{Y}_K = \bar{Y}_{st} \left[ k_{1st} \left( \frac{P_1}{P_{1st}} \right)^{\alpha_{1st}} \exp \left( \frac{\alpha_{1st}}{P_{1st}} - \frac{P_1}{P_{1st}} \right) \right] + k_{2st} \left[ \frac{P_2}{P_{2st}} \right]^{\alpha_{2st}} \exp \left( \frac{\alpha_{2st}}{P_{2st}} - \frac{P_2}{P_{2st}} \right) \tag{16}
\]

Where \( k_{1st}, k_{2st}, \alpha_{1st}, \alpha_{2st} \) are suitable constants. Also \( a_{st}, b_{st}, c_{st} \) and \( d_{st} \) are either real numbers or the function of the auxiliary attribute.

Under the condition \( k_{1st} + k_{2st} \neq 1 \), bias and MSE of \( \bar{Y}_K \) (see [14]), is given as

\[
\text{MSE}(\bar{Y}_K) = \bar{Y}^2 \left[ 1 - \frac{B_1D_i^2 + A_1E_i^2 - 2C_iE_iD_i}{A_1B_i - C_i^2} \right] \tag{17}
\]

or

\[
\text{MSE}(\bar{Y}_K) = \bar{Y}^2 \left[ 1 - \kappa_K \right] \tag{18}
\]

where

\[
\kappa_K = \frac{B_1D_i^2 + A_1E_i^2 - 2C_iE_iD_i}{A_1B_i - C_i^2} \tag{19}
\]

and

\[
A_1 = 1 + 2a_{st}^2 + 2\gamma_{1st} - 2a_{st}b_{st} - 4\gamma_{1st}a_{st} + \gamma_{1st} \left( k_{020} + r_{200} + 4(a_{st} - \gamma_{1st})r_{110} \right) \tag{20}
\]

\[
B_1 = 1 + 2c_{st}^2 + 2\gamma_{2st} - 2c_{st}d_{st} - 4\gamma_{2st}c_{st} + \gamma_{2st} \left( k_{020} + r_{200} + 4(c_{st} - \gamma_{2st})r_{110} \right) \tag{21}
\]

ISSN: 2455-7021
\[ C_1 = 1 + \left[ 1 + r_{020} \left\{ \frac{a_st(a_st - 2b_u)}{2!} + \gamma_{1st}(\gamma_{1st} + 1) \right\} + (a_st - \gamma_{1st})r_{110} \right] + 2(c - \gamma_{2st})r_{101} \]

\[ + (a_stc_st - \gamma_{1st}c_st - \gamma_{2st}a_st + \gamma_{1st}\gamma_{2st})r_{011} + \left\{ \frac{c_st(c_st - 2d_u)}{2!} - \gamma_{2st}c_st + \gamma_{2st}(\gamma_{2st} + 1) \right\} r_{002} + r_{200} \]

\[ D_1 = 1 + r_{020} \left\{ \frac{a_st(a_st - 2b_u)}{2!} - \gamma_{1st}a_st + \gamma_{1st}(\gamma_{1st} + 1) \right\} + (a_st - \gamma_{1st})r_{110} \]

\[ E_1 = 1 + r_{002} \left\{ \frac{c_st(c_st - 2d_u)}{2!} - \gamma_{2st}c_st + \gamma_{2st}(\gamma_{2st} + 1) \right\} + (c_st - \gamma_{2st})r_{101} \]

### 3. Generalized Class of Estimators

3.1. Inspired by [2], [12] and [14], we define the following estimators of population mean \( \hat{Y} \) as

\[ \hat{Y}^1_p = \pi_{1st} \hat{Y}_{by} + \pi_{2st} \frac{Y_{st}}{n_{1st}P_{1st} + m_{1st}} \]

where \((\pi_{1st}, \pi_{2st})\) are suitably chosen scalars such that MSE of \( \hat{Y}^1_p \) is minimum, \((n_{1st}, m_{1st})\) are either constants or functions of some known population parameters such as coefficient of variation

\[ C_{pi_st} = \sum_{i=1}^{L} P_i \beta_{k(i)}(k=1), \text{ Coefficient of skewness} \]

\[ \beta_{k(pi_st)} = \sum_{i=1}^{L} P_i \beta_{k(i)}(k=1), \text{ Coefficient of kurtosis} \]

\[ T_1 = \beta_{2p1} P_1 \left[ \beta_{2p21} P_{1st} + \rho_{yp} \right]^{-1}; T_2 = \beta_{2p1} P_1 \left[ \beta_{2p21} P_{1st} - \rho_{yp} \right]^{-1}; T_3 = \beta_{1p1} P_1 \left[ \beta_{1p1} P_{1st} - \rho_{yp} \right]^{-1}; \]

\[ T_4 = \beta_{2p21} P_1 \left[ \beta_{2p21} P_{1st} - \rho_{yp} \right]^{-1}; T_5 = P_1 \left[ \rho_{yp} \right]^{-1}; T_6 = P_1 \left[ \rho_{yp} \right]^{-1}; \]

\[ T_7 = \rho_{yp} \left[ \rho_{yp} \right]^{-1} \]

We assume \( |T_i|, d_i | < 1 \), so that the term \( \left[ 1 + T_i d_i \right]^{-au} \) is expandable. Thus by expanding the right hand side of Eq. (20) and neglecting terms of \( d_i \)'s having power greater than two, we have

\[ \hat{Y}^1_p = \pi_{1st} \left[ 1 + d_0 - \frac{b_{yp} d_1}{R_1} \right] + \pi_{2st} \left[ 1 + d_0 - T_h \alpha_0 d_1 - T_h \alpha_0 d_0 d_1 + \frac{\alpha_0 (\alpha_0 + 1) T_h^2}{2} d_1 \right] \]

or

\[ \hat{Y}^1_p - \hat{Y} = \pi_{1st} \left[ 1 + d_0 - \frac{b_{yp} d_1}{R_1} \right] + \pi_{2st} \left[ 1 + d_0 - T_h \alpha_0 d_1 - T_h \alpha_0 d_0 d_1 + \frac{\alpha_0 (\alpha_0 + 1) T_h^2}{2} d_1 \right] - 1 \]
Taking expectations on both sides of Eq. (21), we get the bias of \( \hat{Y}_p^i \) to the first degree of approximation as

\[
B(\hat{Y}_p^i) = AB(\hat{Y}_p^i) + O(v)
\]

where \( AB(\hat{Y}_p^i) \) denotes the asymptotic mean square error and \( O(v) \) denotes the terms of order greater than two.

Squaring both sides of Eq. (21) and neglecting terms of \( d_i^2 \)'s having power greater than two, we have

\[
(\hat{Y}_p^i - Y)^2 = Y^2 \left[ 1 + \pi_{1st}^2 \left( 1 + d_0^2 + \frac{b_{yp}^2 d_i^2}{R_i^2} + 2d_0 - \frac{2b_{yp} d_i^2}{R_i} \right) + \pi_{2st}^2 \left( 1 + d_0^2 + T_h^2 \alpha_{st}^2 d_i^2 \right) - 2T_h \alpha_{st} d_i - 4T_h \alpha_{st} d_i + \alpha_{st}^2 \left( \alpha_{st} + 1 \right) T_h^2 d_i^2 \right]
\]

Analogously, using Eq. (23) the MSE of \( \hat{Y}_p^i \) can be expressed as

\[
\text{MSE}(\hat{Y}_p^i) = \text{AMSE}(\hat{Y}_p^i) + O(v)
\]

where \( \text{AMSE}(\hat{Y}_p^i) \) denotes the asymptotic mean square error

\[
\text{AMSE}(\hat{Y}_p^i) = Y^2 \left[ 1 + \pi_{1st}^2 A_2 + \pi_{2st}^2 B_2 + 2\pi_1 \pi_2 C_2 - 2\pi_1 - 2\pi_2 D_2 \right]
\]

where

\[
A_2 = 1 + r_{200} + \frac{b_{yp}^2 r_{200}}{R_i^2} - \frac{2b_{yp} r_{110}}{R_i}
\]

\[
B_2 = 1 + r_{200} + T_h^2 \alpha_{st}^2 r_{200} - 4T_h \alpha_{st} r_{110} + \alpha_{st}^2 \left( \alpha_{st} + 1 \right) T_h^2 r_{200}
\]

\[
C_2 = 1 + r_{200} - 2T_h \alpha_{st} r_{110} - \frac{b_{yp} r_{110}}{R_i} + \frac{b_{yp} \alpha_{st} T_h r_{200}}{R_i} + \frac{\alpha_{st}^2 \left( \alpha_{st} + 1 \right) T_h^2 r_{200}}{2}
\]

\[
D_2 = 1 - T_h \alpha_{st} r_{110} + \frac{\alpha_{st}^2 \left( \alpha_{st} + 1 \right) T_h^2 r_{200}}{2}
\]

Minimization of \( \text{AMSE}(\hat{Y}_p^i) \) is achieved for the optimum choice of the constants \( \pi_1 \) and \( \pi_2 \)

\[
\pi_1^* = \left[ \frac{B_2 - C_2 D_2}{A_2 B_2 - C_2^2} \right] \quad \pi_2^* = \left[ \frac{A_2 D_2 - C_2}{A_2 B_2 - C_2^2} \right]
\]
which leads to

$$\text{AMSE}(\hat{Y}_p) = \bar{Y}^2 \left[ 1 - \left( \frac{B_2 - 2C_2D_2 + A_2D_2^2}{2A_2B_2 - C_2^2} \right) \right]$$

or

$$\text{AMSE}(\hat{Y}_p) = \bar{Y}^2 \left[ 1 - k_p^1 \right]$$

where \( k_p^1 = \frac{(B_2 - 2C_2D_2 + A_2D_2^2)}{(2A_2B_2 - C_2^2)} \) (say).

We would like to mention here that the proposed class \( \hat{Y}_p \) may reduce to some well-known class of population mean \( \bar{Y} \) by putting different values of \( \left[ \pi_{1st}, \pi_{2nd}, \bar{y}_a, n_{1st}, m_{1st}, \alpha_{st} \right] \), i.e.,

\[
\hat{Y}_p \rightarrow \hat{Y}_h : \text{for} \quad \left[ \pi_{1st}, \pi_{2nd}, \bar{y}_a, n_{1st}, m_{1st}, \alpha_{st} \right] = [0,1,1,0,1]
\]

\[
\hat{Y}_p \rightarrow \hat{Y}_p : \text{for} \quad \left[ \pi_{1st}, \pi_{2nd}, \bar{y}_a, n_{1st}, m_{1st}, \alpha_{st} \right] = [0,1,1,0,-1]
\]

\[
\hat{Y}_p \rightarrow \hat{Y}_{\text{reg}} : \text{for} \quad \left[ \pi_{1st}, \pi_{2nd}, \bar{y}_a, n_{1st}, m_{1st}, \alpha_{st} \right] = [0,1,1,0,-1]
\]

Further motivated by Koyun [14], we suggest another efficient estimator using two auxiliary attributes, given as

$$\hat{y}_p = \pi_{1st} \hat{y}_{by} \exp \left( \frac{P_1 - p}{P_1 + p} \right) + \pi_{2nd} \hat{y}_{by} \left[ \frac{n_{2st}p_2 + m_{2st}}{n_{2st}p_2 + m_{2st}} \right] \alpha_{st}$$

where \( \hat{y}_{by} = \bar{y} + b_y(P - p_{2nd}) \) is the well-known regression estimator that depends on study variable \( y \) and auxiliary attribute \( p_{2nd} \). Also \( \left( \pi_{1st}, \pi_{2nd} \right) \) are suitably chosen constants such that MSE of \( \hat{y}_p \) is minimum, \( n_{2nd}, m_{2nd} \) are either constants or function of some known population parameters such as coefficient of variation, kurtosis, skewness and correlation coefficient as already defined in section 3.1 (except here we take value k=1,2). Some members of proposed class of estimators \( \hat{y}_p \) for different values of \( \left( n_{2nd}, m_{2nd}, \pi_{1st}, \pi_{2nd}, \alpha_{st} \right) \) have been summarized in Table 4 in appendix.

Expressing Eq. (27) in terms of \( d_i \)'s, we have

$$\hat{Y}_p^2 = \pi_{1st} \left( \bar{Y}(1 + d_0) - b_{yp1}P_1d_1 \right) \exp \left( \frac{-d_1}{2} + \frac{d_1^2}{4} \right) + \pi_{2nd} \left( \bar{Y}(1 + d_0) - b_{yp2}P_2d_2 \right) \left( 1 + T_h d_2 \right)^{\alpha_{st}}$$

We assume that \( |T_h d_2| < 1 \), so that the term \( \left( 1 + T_h d_2 \right)^{\alpha_{st}} \) is expandable. Thus by expanding the right-hand side of Eq. (28) and neglecting the terms of \( d_i \)'s having power greater than two, we have

$$\hat{Y}_p^2 = \bar{Y} \pi_{1st} \left[ 1 + d_0 - \frac{d_1}{2} + \frac{3d_1^2}{8} - \frac{d_0 d_1}{2} + b_{yp1} \left( d_1 - \frac{d_1^2}{2} \right) \right] + \pi_{2nd} \bar{Y} \left( 1 + d_0 - \alpha_{st} T_h d_2 \right)$$

Or

$$\hat{Y}_p^2 - \bar{Y} = \bar{Y} \pi_{1st} \left[ 1 + d_0 - \frac{d_1}{2} + \frac{3d_1^2}{8} - \frac{d_0 d_1}{2} + b_{yp1} \left( d_1 - \frac{d_1^2}{2} \right) \right] + \pi_{2nd} \bar{Y} \left( 1 + d_0 - \alpha_{st} T_h d_2 \right)$$

$$- \alpha_{st} T_h d_0 d_2 + \frac{\alpha_{st} \left( \alpha_{st} + 1 \right)}{2} T_h^2 d_2^2 - \frac{b_{yp1}}{R_2} \left( d_2 - \alpha_{st} T_h d_2 \right)$$

$$\hat{Y}_p^2 - \bar{Y} = \bar{Y} \pi_{1st} \left[ 1 + d_0 - \frac{d_1}{2} + \frac{3d_1^2}{8} - \frac{d_0 d_1}{2} + b_{yp1} \left( d_1 - \frac{d_1^2}{2} \right) \right] + \pi_{2nd} \bar{Y} \left( 1 + d_0 - \alpha_{st} T_h d_2 \right)$$

$$- \alpha_{st} T_h d_0 d_2 + \frac{\alpha_{st} \left( \alpha_{st} + 1 \right)}{2} T_h^2 d_2^2 - \frac{b_{yp1}}{R_2} \left( d_2 - \alpha_{st} T_h d_2 \right) - \bar{Y}$$

$$\left[ \pi_{1st}, \pi_{2nd}, \bar{y}_a, n_{1st}, m_{1st}, \alpha_{st} \right] = [1,0,-,-,-]$$
where $R_2 = \frac{Y}{\bar{p}_2}$ and $T_h$ has the following possible values

\begin{align*}
T'_1 &= P_2 \left[ \rho_{p2} - \rho_{p2}^2 \right]^{-1};
T'_2 &= P_2 \left[ \rho_{p2} - \rho_{p2}^2 \right]^{-1};
T'_3 &= \beta_{1p1} P_2 \left[ \rho_{p2} - \rho_{p2}^2 \right]^{-1};
T'_4 &= \beta_{1p1} P_2 \left[ \rho_{p2} + \rho_{p2}^2 \right]^{-1};
T'_5 &= \beta_{1p2} P_2 \left[ \beta_{1p2} P_{20} - \beta_{1p2}^2 \right]^{-1};
\end{align*}

Taking expectation on both sides of Eq. (30) we get the bias of $\hat{Y}_p$ to the first degree of approximation as

\begin{align*}
B(\hat{Y}_p) &= AB(\hat{Y}_p) + O(v) \\
AB(\hat{Y}_p) &= Y \left[ \frac{1}{8} + \frac{r_{110}}{2} + \frac{b_{yp1} r_{020}}{2 R_1} \right] + \left[ 1 - \alpha'_{st} T_h r_{020} + \frac{\alpha'_{st} (\alpha'_{st} + 1)}{2} T_h^2 r_{020} \right. \\
&\left. + \frac{b_{yp1} \alpha'_{st} T_h r_{020}}{R_2} \right] - 1
\end{align*}

(31)

Squaring both sides of Eq. (30) neglecting terms of $d_i$'s having power greater than two and then taking expectations, we have

\begin{align*}
\text{MSE}(\hat{Y}_p) &= \text{AMSE}(\hat{Y}_p) + O(v) \\
\text{AMSE}(\hat{Y}_p) &= \bar{Y}^2 \left[ 1 + \pi' \rho_{1st} A_3 + \pi' \rho_{2st} B_3 - 2 \pi' C_3 - 2 \pi' D_3 + 2 \pi' E_3 \right]
\end{align*}

where $\text{AMSE}(\hat{Y}_p)$ denotes the asymptotic mean square error

\begin{align*}
A_3 &= \left[ 1 + r_{020} + r_{020} + \frac{b_{yp1} r_{020}}{2 R_1} \right. \\
&\left. + \frac{2 b_{yp1} r_{020}}{R_1} - \frac{2 b_{yp1} r_{020}}{R_1} \right] \\
B_3 &= \left[ 1 + r_{020} + \alpha'_{st} T_h^2 r_{020} + \frac{b_{yp1} r_{020}}{R_1} - 4 \alpha'_{st} T_h r_{010} + \alpha'_{st} (\alpha'_{st} + 1) T_h^2 r_{002} + \frac{4 b_{yp1} \alpha'_{st} T_h r_{002}}{R_2} \right. \right. \\
&\left. \left. - \frac{2 b_{yp1} r_{010}}{R_2} \right] \\
C_3 &= \left[ 1 + \frac{3 r_{020}}{8} - \frac{r_{110}}{2} + \frac{b_{yp1} r_{020}}{2 R_1} \right] \\
D_3 &= \left[ 1 + \alpha'_{st} (\alpha'_{st} + 1) T_h^2 r_{002} - \alpha'_{st} T_h r_{010} + \frac{b_{yp1} \alpha'_{st} T_h r_{002}}{R_2} \right] \\
E_3 &= \left[ 1 + \frac{3 r_{020}}{8} + \frac{\alpha'_{st} (\alpha'_{st} + 1)}{2} T_h^2 r_{020} + \frac{b_{yp1} r_{020}}{2 R_1} + \frac{b_{yp1} \alpha'_{st} T_h r_{002}}{R_2} - r_{110} - 2 \alpha'_{st} T_h r_{010} \right]
\end{align*}
Minimization of AMSE($\hat{Y}_p^2$) is achieved for the optimum choice of constants $\pi'_1$ and $\pi'_2$

\[
\pi'_1 = \left[ \frac{B_1C_3 - D_1E_3}{A_1B_3 - E_3^2} \right] \quad \pi'_2 = \left[ \frac{A_1D_1 - C_1E_3}{A_1B_3 - E_3^2} \right]
\]

which leads to

\[
AM_{\min}(\hat{Y}_p^2) = \nabla^2 \left[ 1 - \frac{A_1B_1^2C_1^2 - A_1D_1^2E_3^2 + A_1^2B_1D_3^2 - B_1C_3^2E_3^2 + 2C_1D_1E_3^3 - 2A_1B_1C_1D_3E_1}{(A_1B_3 - E_3^2)^2} \right]
\]

(33)

or

\[
AM_{\min}(\hat{Y}_p^2) = \nabla^2 \left[ 1 - \kappa_p^2 \right]
\]

(34)

where $\kappa_p^2 = \frac{A_1B_1^2C_1^2 - A_1D_1^2E_3^2 + A_1^2B_1D_3^2 - B_1C_3^2E_3^2 + 2C_1D_1E_3^3 - 2A_1B_1C_1D_3E_1}{(A_1B_3 - E_3^2)^2}$ (say)

and $AM_{\min}(\hat{Y}_p^2)$ shows the minimum asymptotic mean square error of proposed estimator $\hat{Y}_p^2$.

4. Efficiency Comparison

From Eqs. (3), (7), (9), (11), (14) and (26), we have

\[
AMSE(\hat{Y}_p^1) \leq MSE(\hat{Y}_R), \text{ if } \left[ 1 - \kappa_p^1 \right] - (r_{200} + r_{202} - 2r_{110}) < 0
\]

(35)

for $\kappa_p^1 > 0$

\[
AMSE(\hat{Y}_p^1) \leq MSE(\hat{Y}_{g(st)}), \text{ if } \left[ 1 - \kappa_p^1 \right] - \left[ 1 + \mu_{1a}^2 \left( 1 + r_{200} \right) + \frac{\mu_{1a}^2}{R_{1}^2} r_{202} - 2\mu_{1a}^2 + 2\mu_{1a}^2 \frac{r_{110}}{R_{1}^2} \right] \leq 0
\]

(36)

\[
AMSE(\hat{Y}_p^2) \leq MSE(\hat{Y}_{g(st)}), \text{ if } \left[ 1 - \kappa_p^2 \right] \leq \kappa_K
\]

(37)

\[
AMSE(\hat{Y}_p^1) \leq MSE(\hat{Y}_{ss}), \text{ if } \left[ 1 - \kappa_p^1 \right] - \left[ 1 - \kappa_{ss} \right] < 0 \text{ or } \kappa_p^1 > \kappa_{ss}
\]

(38)

or $\kappa_p^1 > \kappa_{ss}$

From Eqs. (18) and (34), we have

\[
AMSE(\hat{Y}_p^2) \leq MSE(\hat{Y}_{K}), \text{ if } \left[ 1 - \kappa_p^2 \right] \leq \kappa_K
\]

(39)

Equations (35) to (40) are the conditions under which proposed classes $(\hat{Y}_p^1, \hat{Y}_p^2)$ have less mean square error, i.e., more efficient than simple mean estimator $\hat{Y}_{st}$, usual ratio estimator $\hat{Y}_R$, combined regression.
estimator \( \hat{Y}_{p|p} \), Rao class \( \hat{Y}_{RA} \), Singh et al. \( \hat{Y}_{SS} \) and Koyunsu generalized class of estimators \( (\hat{Y}_{g(a)}, \hat{Y}_{k}) \).

5. Empirical Study

To evaluate the performance of proposed estimators \( (\hat{Y}_{p|1}, \hat{Y}_{p|2}) \) in comparison to other estimators, we have considered the same data set as considered by Koyuncu [14]. Here the apple production amount in 1999 is specified as study variable and apple production amount less than 1000 tons in 1998 as auxiliary attributes in 854 villages of Turkey (Source: Institute of Statistics, republic of Turkey).

The data set we have considered here is firstly stratified into six different regions of Turkey (as 1: Aegean 2: Central Anatolia 3: Black Sea 4: East and southeast Anatolia 5: Mediterranean 6: Marmara) and from each stratum; we have randomly selected the samples whose sizes are computed by using Neyman allocation method. The summary of the data is given in Table 5 in appendix.

The MSE values for existing estimators \( (\hat{Y}_{R}, \hat{Y}_{g(a)}, \hat{Y}_{p|p}, \hat{Y}_{RA}, \hat{Y}_{SS}) \) and adapted estimator \( (\hat{Y}_{p|1}, \hat{Y}_{p|2}) \) are given in Table 1. When we examine Table 1, we observe that \( (\hat{Y}_{p|1}, \hat{Y}_{p|2}) \) have the smallest MSE values among \( (\hat{Y}_{p|1}, \hat{Y}_{p|2}) \) and all existing estimators defined in Sec. 2 including \( \hat{Y}_{K} \). Taking usual regression estimator \( \hat{Y}_{p|p} \) as base, we have also given PRE values for some efficient estimators [among \( (\hat{Y}_{p|1}, \hat{Y}_{p|2}) \)] and existing estimators in Table 1. For computing the PRE value, we use the following expression

\[
\text{PRE}(\chi) = \frac{\text{MSE}(\hat{Y}_{p|p})}{\text{MSE}(\hat{Y})} * 100
\]

where \( \chi = \hat{Y}_{p|1}, \hat{Y}_{p|2}, \hat{Y}_{R}, \hat{Y}_{g(a)}, \hat{Y}_{RA}, \hat{Y}_{SS} \) and \( \hat{Y}_{K} \).

### Table 1. MSE Values for Adapted and Existing Estimators

<table>
<thead>
<tr>
<th>Estimators</th>
<th>MSE/ PRE Estimators</th>
<th>MSE/ PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{Y}_{p</td>
<td>1} )</td>
<td>490041/79.14 ( \hat{Y}_{p</td>
</tr>
<tr>
<td>( \hat{Y}_{g(a)} )</td>
<td>367247/105.6 ( \hat{Y}_{p</td>
<td>1} )</td>
</tr>
<tr>
<td>( \hat{Y}_{p</td>
<td>p} )</td>
<td>387835.5/100* ( \hat{Y}_{p</td>
</tr>
<tr>
<td>( \hat{Y}_{RA} )</td>
<td>371116.6/104.50 ( \hat{Y}_{p</td>
<td>p} )</td>
</tr>
<tr>
<td>( \hat{Y}_{SS} )</td>
<td>346120.0/112.05 ( \hat{Y}_{p</td>
<td>p} )</td>
</tr>
<tr>
<td>( \hat{Y}_{K} )</td>
<td>248035.4/156.36** ( \hat{Y}_{p</td>
<td>p} )</td>
</tr>
<tr>
<td>( \hat{Y}_{p</td>
<td>1} )</td>
<td>309538.5/125.29* ( \hat{Y}_{p</td>
</tr>
<tr>
<td>( \hat{Y}_{p</td>
<td>2} )</td>
<td>341079.6/113.48 ( \hat{Y}_{p</td>
</tr>
<tr>
<td>( \hat{Y}_{p</td>
<td>3} )</td>
<td>344776.8/112.48 ( \hat{Y}_{p</td>
</tr>
</tbody>
</table>

* corresponds to the MSE/PRE of the estimators using single auxiliary attribute; ** for two auxiliary attributes

Table 2 establishes the conditions obtained in Sec. 4 empirically. It shows that all conditions are satisfied for the given data set.

### Table 2. Empirical Study of Theoretical Conditions Explained in Sec. 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>Condition</th>
<th>Condition</th>
<th>Condition</th>
<th>Condition</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(35)</td>
<td>(36)</td>
<td>(37)</td>
<td>(38)</td>
<td>(39)</td>
<td>(40)</td>
</tr>
<tr>
<td>-0.02102 &lt; 0</td>
<td>-0.006722 &lt; 0</td>
<td>0.00912 &lt; 0</td>
<td>-0.00717 &lt; 0</td>
<td>0.963947 &gt; 0.959686</td>
<td>0.978962 &gt; 0.959093</td>
</tr>
</tbody>
</table>
6. Conclusion

Section 4 provides the theoretical conditions under which proposed classes \( \hat{Y}_1, \hat{Y}_2 \) are more efficient than other estimators. Table 2 establishes those conditions empirically and it shows that all theoretical conditions obtained in Sec. 4 are satisfied for the given data set (see Sec. 5). Table 1 exhibits mean square error and percent relative efficiency of proposed and existing classes with respect to usual regression estimator \( \hat{Y}_{hyp} \). It can be observed that using single auxiliary attribute \( \hat{Y}_1 \) has highest percent relative efficiency and respectively \( \hat{Y}_3 \) for two auxiliary attributes. Thus, it can be concluded that if conditions obtained in Sec. 4 are satisfied the proposed classes \( \hat{Y}_1, \hat{Y}_2 \) are much efficient for application point of view in comparison to other methods.

References

Appendix

In the following Table 3, we have given some family of estimators of proposed class $\left(\hat{Y}_1^1, \hat{Y}_1^2\right)$ as

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}_1^{ip}$</td>
<td>$\alpha_m$, $n_{1st}$, $m_{1st}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{1p}$</td>
<td>$1$, $-\beta_{2p_1}$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{2p}$</td>
<td>$1$, $1$, $-\rho_{yp}^2$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{ip}$</td>
<td>$1$, $\beta_{2p_1}$, $-\rho_{yp}^2$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{1p}$</td>
<td>$1$, $-\beta_{1p_1}$, $\rho_{yp}^2$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{1p}$</td>
<td>$1$, $-\beta_{2p_1}$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{1p}$</td>
<td>$1$, $-1$, $\rho_{yp}^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}_2^{ip}$</td>
<td>$\alpha_m$, $n_{2st}$, $m_{2st}$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{1p}$</td>
<td>$1$, $1$, $-\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{2p}$</td>
<td>$1$, $1$, $-\rho_{yp}^2$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{ip}$</td>
<td>$1$, $-\beta_{2p_1}^2$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{1p}$</td>
<td>$1$, $-\beta_{2p_1}$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{2p}$</td>
<td>$1$, $-\beta_{2p_1}^2$, $1$</td>
</tr>
<tr>
<td>$\hat{Y}_2^{1p}$</td>
<td>$1$, $-\beta_{2p_1}$, $\rho_{yp}^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}_1^{ip}$</td>
<td>$\alpha_m$, $n_{1st}$, $m_{1st}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{2p}$</td>
<td>$1$, $1$, $-\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{ip}$</td>
<td>$1$, $\beta_{2p_1}$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{2p}$</td>
<td>$1$, $-\beta_{1p_1}$, $\rho_{yp}^2$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{ip}$</td>
<td>$1$, $-\beta_{2p_1}$, $\rho_{yp}$</td>
</tr>
<tr>
<td>$\hat{Y}_1^{1p}$</td>
<td>$1$, $-1$, $\rho_{yp}^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5.Parameters of the Population under Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_1 = 106$</td>
</tr>
<tr>
<td>$N_4 = 171$</td>
</tr>
<tr>
<td>$n_1 = 13$</td>
</tr>
<tr>
<td>$n_4 = 95$</td>
</tr>
<tr>
<td>$S_{y1} = 6425.0872$</td>
</tr>
<tr>
<td>$S_{y4} = 28643.42$</td>
</tr>
<tr>
<td>$\bar{Y}_1 = 1536.773585$</td>
</tr>
<tr>
<td>$\bar{Y}_4 = 5588.012$</td>
</tr>
<tr>
<td>$S_{p_1}(1) = 0.432298$</td>
</tr>
<tr>
<td>$S_{p_1}(4) = 0.501254$</td>
</tr>
<tr>
<td>$P_{1}(1) = 0.24528$</td>
</tr>
<tr>
<td>$\text{P}_1(1)$</td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td>0.48538</td>
</tr>
<tr>
<td>$\text{S}_1(1)$</td>
</tr>
<tr>
<td>0.4205</td>
</tr>
<tr>
<td>$\text{S}_1(4)$</td>
</tr>
<tr>
<td>0.49085</td>
</tr>
<tr>
<td>$\text{P}_2(1)$</td>
</tr>
<tr>
<td>0.22641</td>
</tr>
<tr>
<td>$\text{P}_2(1)$</td>
</tr>
<tr>
<td>0.39766</td>
</tr>
<tr>
<td>$\rho_{\text{yp}_1}(1)$</td>
</tr>
<tr>
<td>0.3588</td>
</tr>
<tr>
<td>$\rho_{\text{yp}_1}(4)$</td>
</tr>
<tr>
<td>0.19113</td>
</tr>
<tr>
<td>$\rho_{\text{yp}_2}(1)$</td>
</tr>
<tr>
<td>0.38094</td>
</tr>
<tr>
<td>$\rho_{\text{yp}_2}(4)$</td>
</tr>
<tr>
<td>0.22694</td>
</tr>
<tr>
<td>$\rho_{\text{p}\text{p}_2}(1)$</td>
</tr>
<tr>
<td>0.8442</td>
</tr>
<tr>
<td>$\rho_{\text{p}\text{p}_2}(4)$</td>
</tr>
<tr>
<td>0.7649</td>
</tr>
<tr>
<td>$\beta_2(\text{p}_{1(1)})$</td>
</tr>
<tr>
<td>0.56846</td>
</tr>
<tr>
<td>$\beta_2(\text{p}_{1(4)})$</td>
</tr>
<tr>
<td>2.02028</td>
</tr>
<tr>
<td>$\beta_1(\text{p}_{1(1)})$</td>
</tr>
<tr>
<td>1.20109</td>
</tr>
<tr>
<td>$\beta_1(\text{p}_{1(4)})$</td>
</tr>
<tr>
<td>0.05902</td>
</tr>
<tr>
<td>$\beta_2(\text{p}_{2(1)})$</td>
</tr>
<tr>
<td>0.24603</td>
</tr>
<tr>
<td>$\beta_2(\text{p}_{2(4)})$</td>
</tr>
<tr>
<td>1.84369</td>
</tr>
<tr>
<td>$\beta_1(\text{p}_{2(1)})$</td>
</tr>
<tr>
<td>1.32626</td>
</tr>
<tr>
<td>$\beta_1(\text{p}_{2(4)})$</td>
</tr>
<tr>
<td>0.42192</td>
</tr>
</tbody>
</table>