Chapter IV

ON THE IMPROVEMENT OF A GENERALIZED CLASS OF SYNTHETIC ESTIMATORS WITH ITS APPLICATIONS

4.1 INTRODUCTION

4.1.1. In most of the survey situations, the emphasis is laid down on the use of some auxiliary/apriori or supplementary information (already available or may be made available by diverting a part of the resources), for improving the precision of the estimator method. If used intelligibly, this information may provide us with the sampling strategies better than those in which no auxiliary information is used. The information on auxiliary variable may be available in various forms, such as, mean, variance, coefficient of variation, coefficient of skewness, etc. The best use of auxiliary information in the form of population mean of the auxiliary variable, at the estimation stage, was made to develop a general class of estimators for population mean given by

\[ T = \bar{y} + \phi (\bar{X} - \bar{x}) , \quad (4.1) \]

as mentioned in the section 1.1.2. of chapter I. It has been shown there that how different choices of the constant yields usual ratio, usual product, difference and linear regression estimators as members of the class.

Probably the earliest use of auxiliary information in devising the estimation techniques, which gives an improvement in the precision of the estimate, started after 1940 when
Chapter IV. On the improvement of a generalized class of synthetic estimators with its application

Watson (1937), Cochran (1940,42) and Hansen and Hurwitz (1943) developed some sampling strategies based upon auxiliary information. Cochran (1940,42) was the first to consider the ratio, regression estimators. Hansen et al (1953) were the first to discuss the difference type estimator and Robson (1957) was the first to discuss product-type estimator. As a consequence of these estimators, later on a number of authors suggested modification over these estimators. Some of the important works in this direction are due to Singh (1965,67,69), Srivastava (1967,71,81), Reddy (1973,74) and Gupta (1978). In whatever from the auxiliary information in these estimators was utilized, it was always possible to devise suitable ways of using it in obtaining more and more efficient sampling strategies.

4.1.2. While the information on an auxiliary information is assumed to be known, the efficiency of the sampling strategy primarily depends upon the way in which maximum use of such information is made. Some of the commonly used devices are to use weighted means and ratio of weighted means of functions of auxiliary character. Such a use of weighted means of auxiliary information has been demonstrated in chapters II and III in order to develop factor-type synthetic estimator for SAE. Another technique is to define a generalized class of estimators on the basis of a certain parameter which includes many of the estimators as member of the class and, thus, on the basis of the study of the general estimator it is possible to study at a time the properties of several estimators. In this sense, it is a unified approach for dealing a number of estimators. Moreover, one is also able to locate the estimator within the class which possesses minimum variance by minimizing the variance of the estimator with respect to the parameter. In literature, a large number of different classes of ratio, product or regression type estimators is available.

4.1.3. When the auxiliary information is used at the estimation stage, the classical ratio estimator is considered to be most practicable due to the reason that it has good intuitive basis and can be obtained if one adopts the prediction approach for predicting the mean of the unobserved units of the population on the basis of observed units belonging
Chapter IV. On the improvement of a generalized class of synthetic estimators with its application

65
to the sample (Srivastava, 1983). This result has been shown in the following chapter, that is in chapter V. Walsh (1970) has also pointed out that “It would seem worthwhile to develop an estimate which has a ratio form to that of the ratio estimate and variance properties that are not inferior to those attainable with a linear regression estimate.” Consequently, he developed a ratio-type estimator for population mean, given by

\[ t_{W,a} = \bar{y} \left[ \frac{\bar{X}}{(\alpha \bar{x} + (1 - \alpha) \bar{X})} \right] \] (4.2)

where \( \alpha \) is a suitably chosen constant whose appropriate choice is based on the criterion that the estimator has a standard deviation that is approximately as small as that attainable for the linear regression estimator.

4.1.4 In previous chapters, we developed two families of synthetic estimators utilizing the information on an auxiliary variable \( X \) for the estimation of mean of small area with the idea of factor-type estimators. It is clear that while factor-type estimators are developed for estimating the population mean of the study characteristic \( Y \) and hence, utilized the information on the population mean, \( X \), of the auxiliary variable, the similar factor-type estimators are developed, which are synthetic estimators in the sense that these utilize the available information on the mean of the auxiliary variable \( X_a \), of the small domain of interest. Such estimators and other synthetic estimators receive strength from the sample functions based on a sample selected from the entire population. Both the families of synthetic estimators, suggested respectively in chapter II and chapter III, may be considered as ratio-type synthetic estimators as they perform better under the assumption that the study and auxiliary characteristics are positively correlated and also due to the fact that they asymptotically converge either to usual ratio estimator or to ratio-type estimator. Both the estimators were also observed to attain the minimum variance properties of the linear regression synthetic estimator, given by \( t_{lr} = \bar{y} + b \left( \bar{X} - \bar{x} \right) \), where \( b \) has the same meaning as described in section 1.1.2. In fact, Rao (2003) has shown that ratio synthetic estimator \( \tilde{y}_{RS,a} = \tilde{y} \left( \bar{X}_a / \bar{x} \right) \), for population mean \( \bar{Y}_a \) is a particular case of Generalized Regression (GREG) Estimator with a single auxiliary variable \( X \) and
known domain specific auxiliary information.

4.1.5. In the present chapter, once again we have focussed our attention towards developing a family of ratio-type synthetic estimator, which might be preferable over other synthetic estimators. Srivastava (1967) proposed a ratio type estimator \( t_S = \bar{y}(\bar{X}/\bar{x})^\lambda \) using the power transformation, where \( \lambda \) is a suitably chosen constant. On the other hand, Walsh (1970) defined another ratio-type estimator, \( t_{W,\alpha} \), given in (4.2), which utilized the concept of using convex linear combination of sample and population means of the auxiliary variable with unknown weights. Motivated with the estimator of Srivastava (1967) and Walsh (1970), we have suggested a family of synthetic ratio-type estimator in the present chapter. Some of the members of the family have been discussed as they are well-known synthetic estimators already studied by other authors. The salient properties of the family have been discussed in details. The optimum estimator belonging to the class has been obtained. The estimator has been compared with the other synthetic estimators on the basis of their MSEs using some empirical data. To support the results obtained for the suggested class, a simulation study has also been made at the end.

4.2 GENERALIZED CLASS OF SYNTHETIC ESTIMATORS

Now we shall suggest a generalized class of synthetic estimator. For this, we shall use same notations as used in chapter II and III. The problem is here to define a synthetic estimator for the estimation of domain mean, \( \bar{Y}_a \), utilizing the domain-specific auxiliary information on the domain mean of the auxiliary variable \( X \). Motivated with the works of Srivastava (1967) and Walsh (1970), we define the generalized family of synthetic estimators as

\[
T_{W,\beta}^a = \bar{y} \left( \frac{\bar{x}_W}{\bar{X}_a} \right)^\beta ; \tag{4.3}
\]
where \( \bar{x}_W = W \bar{x} + (1 - W) \bar{X}_n, \) \( 0 < W < 1; \) \( W \) and \( \beta \) are suitably chosen constants. \( \bar{x}_W \) is basically a convex linear combination of the sample mean \( \bar{x} \), based on a sample of size \( n \) and the known domain mean of the auxiliary variable.

**Remark 1.** The estimator \( T_{W,\beta}^a \) defines a general class of synthetic estimators, which consists a number of synthetic estimators as special cases.

### 4.3 SOME SPECIAL CASES OF \( T_{W,\beta}^a \)

Since \( T_{W,\beta}^a \) is a function of two constants \( W \) and \( \beta \), it reduces to some well-known synthetic estimators for some particular choices of the constants (parameter). The following table presents a list of estimators for different choices of \( W \) and \( \beta \) which belong to the family:

**Various indirect estimators as special cases of the generalized estimator \( T_{W,\beta}^a \)**

<table>
<thead>
<tr>
<th>S.NO.</th>
<th>Choice of ( W )</th>
<th>Choice of ( \beta )</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( \beta )</td>
<td>( \hat{y}<em>a = \bar{y} ) (simple synthetic estimator ( \hat{y}</em>{SS,a} ) or ( \hat{y}_{4,a} ) given in (2.5), chapter II)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>( \beta )</td>
<td>( T_{1,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}_a} \right) ) (generalized synthetic estimator proposed by Tikkiwal and Ghiya(2000))</td>
</tr>
<tr>
<td>3</td>
<td>( r = \frac{n}{n+1} )</td>
<td>1</td>
<td>( T_{2,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}<em>a} \right) ) (product-type synthetic estimator ( T</em>{2,a} ) defined in (3.7), chapter III)</td>
</tr>
<tr>
<td>4</td>
<td>( r = \frac{n}{n-1} )</td>
<td>-1</td>
<td>( T_{0,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}<em>a} \right) ) (ratio-type synthetic estimator ( T</em>{0,a} ) defined in (3.3), chapter III)</td>
</tr>
<tr>
<td>5</td>
<td>( W )</td>
<td>0</td>
<td>( \hat{y}<em>W = \bar{y} ) (simple synthetic estimator ( \hat{y}</em>{SS,a} ) or ( \hat{y}_{4,a} ) given in (2.5), chapter II)</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-1</td>
<td>( T_{1,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}<em>a} \right) ) (ratio synthetic estimator ( T</em>{1,a} = \bar{y}_{RS,a} ) defined in (2.2) chapter II defined by Rao(2003))</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>( T_{1,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}<em>a} \right) ) (product synthetic estimator ( T</em>{2,a} = \bar{y}_{PS,a} ) defined in (2.3) chapter II)</td>
</tr>
</tbody>
</table>

### 4.4 DESIGN - BIAS AND MSE OF \( T_{W,\beta}^a \)

4.4.1. Since it is evident that the estimator \( T_{W,\beta}^a \) is not an unbiased estimator, the bias and MSE of the estimator will be obtained using large sample approximation. We know that under simple random sampling without replacement scheme, we have \( E(\bar{y}) = \bar{Y} \) and \( E(\bar{x}) = \bar{X} \), therefore, we write

\[
\bar{y} = \bar{Y} (1 + e_1), \quad \bar{x} = \bar{X} (1 + e_2),
\]

where \( e_1 \) and \( e_2 \) are sampling errors.
where obviously

\[ E(e_1) = E(e_2) = 0 \quad \text{and} \quad E(e_1^2) = V(e_1) = \frac{N-n}{Nn}C_Y^2 = V_{20} \]

\[ E(e_2^2) = V(e_2) = \frac{N-n}{Nn}C_X^2 = V_{02} \quad \text{and} \quad E(e_1e_2) = \frac{N-n}{Nn}C_{XY} = V_{11}. \]

The expression (4.3), then can be expressed as

\[
T_{w,\beta}^n = \bar{Y}(1 + e_1) \left( \frac{W \bar{X} (1 + e_2) + \bar{X}_a - W \bar{X}_a}{\bar{X}_a} \right)^\beta \\
= \bar{Y}(1 + e_1) \left( \frac{D + W \bar{X}e_2}{\bar{X}_a} \right)^\beta \tag{4.4}
\]

where \( D = W \bar{X} + (1 - W) \bar{X}_a \).

Now (4.4) can be written as

\[
T_{w,\beta}^n = \bar{Y}(1 + e_1) \left( \frac{D}{\bar{X}_a} \right)^\beta \left( 1 + \frac{W \bar{X}}{D}e_2 \right)^\beta. \tag{4.5}
\]

Obviously, \( \left| \frac{W \bar{X}}{D}e_2 \right| < 1 \), for all choice of \( W \in (0,1) \). Therefore, expanding the term \( \left( 1 + \frac{W \bar{X}}{D}e_2 \right)^\beta \) and retaining terms up to the second power of \( e_1 \) and \( e_2 \) (that is, the terms up to the order \( O(n^{-1}) \)), we get

\[
T_{w,\beta}^n = \bar{Y} \left( \frac{D}{\bar{X}_a} \right)^\beta \left[ 1 + \beta \left( \frac{W \bar{X}}{D} \right) e_2 + \frac{\beta(\beta - 1)}{2} \left( \frac{W \bar{X}}{D} \right)^2 e_2^2 + e_1 + \beta \left( \frac{W \bar{X}}{D} \right) e_1 e_2 \right]. \tag{4.6}
\]

Hence, taking expectation of both the sides of (4.6) applying the large sample approximations and realizing that \( B[T_{w,\beta}] = E[T_{w,\beta}^n] - \bar{Y}_a \), we get

\[
B[T_{w,\beta}^n] = \left\{ \bar{Y} \left( \frac{D}{\bar{X}_a} \right)^\beta - \bar{Y}_a \right\} \\
+ \bar{Y} \left( \frac{D}{\bar{X}_a} \right)^\beta \left[ \frac{\beta(\beta - 1)}{2} \left( \frac{W \bar{X}}{D} \right)^2 V_{02} + \beta \left( \frac{W \bar{X}}{D} \right) V_{11} \right] \tag{4.7}
\]
Remark 2. Tikkiwal and Ghiya (2000) defined the estimator
\[
\bar{y}_{\text{syn},a} = \bar{y} \left( \frac{\bar{x}}{X_a} \right)^{\beta}
\]  
which may be considered a special case of \( T^a_{W,\beta} \) for \( W = 1 \) (which is listed in Table 4.1). Thus, \( T^a_{W,\beta} \) may be looked upon as an extension of \( \bar{y}_{\text{syn},a} \). The bias of the estimator \( \bar{y}_{\text{syn},a} \) was reported by Tikkiwal and Ghiya (2000) as
\[
B[\bar{y}_{\text{syn},a}] = \bar{Y} \left( \frac{\bar{X}}{X_a} \right)^{\beta} \left[ 1 + \frac{N - n}{Nn} \left\{ \frac{\beta(\beta - 1)}{2} C_X^2 + \beta C_{XY} \right\} \right] - \bar{Y}_a
\]  
which can be deduced from (4.7) by putting \( W = 1 \). Further, if the synthetic assumption \( \bar{Y}_a (X_a)^{\beta} \approx \bar{Y} (X)^{\beta} \) is satisfied for all \( \beta \) and \( W = 1 \), then the bias of \( T^a_{W,\beta} \) reduces to
\[
B[T^a_{W,\beta}] = \bar{Y}_a \left( \frac{N - n}{Nn} \right) \left[ \frac{1}{2} C_X^2 (\beta^2 - \beta) + \beta C_{XY} \right]
\]  
Remark 3. Choosing appropriate values of the parameters \( W \) and \( \beta \) in the expression (4.7), one can deduce the bias of different synthetic estimators which are listed in table 4.1.

4.4.2. Further, let us obtain the expression for MSE of the estimator \( T^a_{W,\beta} \). By definition
\[
M \left[ T^a_{W,\beta} \right] = E \left[ T^a_{W,\beta} - \bar{Y}_a \right]^2 = E \left[ T^a_{W,\beta} \right]^2 - 2\bar{Y}_a E \left[ T^a_{W,\beta} \right] + \bar{Y}_a^2
\]  
Now, from (4.5), we have
\[
(T^a_{W,\beta})^2 = \left[ \bar{Y} (1 + e_1) \left( \frac{D}{X_a} \right)^{\beta} \left( 1 + \frac{W\bar{X}}{D} e_2 \right)^{\beta} \right]^2
\]
\[
= \bar{Y}^2 \left( \frac{D}{X_a} \right)^{2\beta} \left[ 1 + 2\beta \left( \frac{W\bar{X}}{D} \right) e_2 + \beta (2\beta - 1) \left( \frac{W\bar{X}}{D} \right)^2 e_2^2 + 4\beta \left( \frac{W\bar{X}}{D} \right) e_1 e_2 + 2e_1 + e_1^2 \right]
\]
expanding the expression on right hand side and retaining terms upto the second power of $e_1$ and $e_2$. Taking expectation of the expression (4.12) and applying the results of the large sample approximation, we get

$$E[T^a_{W,\beta}]^2 = \bar{Y}^2 \left( \frac{D \bar{X}}{X_a} \right)^{2\beta} \left[ 1 + \beta (2\beta - 1) \left( \frac{W \bar{X}}{D} \right)^2 V_{02} + 4\beta \left( \frac{W \bar{X}}{D} \right) V_{11} + V_{20} \right]. \tag{4.13}$$

Further, taking expectation of the expression (4.6), we get

$$E[T^a_{W,\beta}] = \bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} \left[ 1 + \frac{\beta (\beta - 1)}{2} \left( \frac{W \bar{X}}{D} \right)^2 V_{02} + \beta \left( \frac{W \bar{X}}{D} \right) V_{11} \right]. \tag{4.14}$$

Hence, substituting the values of $E[T^a_{W,\beta}]^2$ and $E[T^a_{W,\beta}]$ from (4.13) and (4.14) respectively in the expression (4.11), we obtain

$$M[T^a_{W,\beta}] = \left\{ \bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} - \bar{Y}_a \right\}^2 + \bar{Y}^2 \left( \frac{D \bar{X}}{X_a} \right)^{2\beta} V_{20} + \beta \bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} \left[ (2\beta - 1) \bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} - (\beta - 1) \bar{Y}_a \right] V_{02} + 2\beta \bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} \left[ 2\bar{Y} \left( \frac{D \bar{X}}{X_a} \right)^{\beta} - \bar{Y}_a \right] V_{11} \tag{4.15}$$

**Remark 4.** It can be seen from (4.15) that when $W = 1$, the $M[T^a_{W,\beta}]$ reduces to $M[\bar{y}_{sym,a}]$, where $\bar{y}_{sym,a}$ is due to Tikkiwal and Ghiya (2000), given in (4.8). The expression of $M[\bar{y}_{sym,a}]$ obtained from (4.15) for $W = 1$ is

$$M[\bar{y}_{sym,a}] = \bar{Y}^2 \left( \frac{\bar{X}}{X_a} \right)^{2\beta} \left[ 1 + \frac{N - n}{Nn} \left\{ (2\beta^2 - \beta) C^2_X + C_Y^2 + 4\beta C_{XY} \right\} \right] - 2\bar{Y}\bar{Y}_a \left( \frac{\bar{X}}{X_a} \right)^{\beta} \left[ 1 + \frac{N - n}{Nn} \left\{ \frac{\beta (\beta - 1)}{2} C^2_X + \beta C_{XY} \right\} \right] + \bar{Y}^2, \tag{4.16}$$

which tallies with the expression of $M[\bar{y}_{sym,a}]$ reported by Tikkiwal and Ghiya (2000).

**Remark 5.** The MSE of other synthetic estimators which belong to the proposed family and listed in Table 4.1 can easily be obtained by choosing appropriate values of $W$ and
\( \beta \). For example, letting \( W = 0 \), we have \( D = \bar{X}_a \) and hence for all \( \beta \), \( M \left[ T_{W,\beta}^a \right] \) reduces to

\[
M \left[ T_{0,\beta}^a \right] = (\bar{Y} - \bar{Y}_a)^2 + \frac{N - n}{Nn} \bar{Y}^2 C_Y^2,
\]

which is MSE of simple synthetic estimator \( \bar{y}_{ss,a} \). The same MSE will be obtained for the estimator \( T_{W,0}^a \) for all \( W \). Similarly, another important estimator which should be dealt with here is ratio synthetic estimator \( \bar{y}_{RS,a} \) discussed by Rao (2003). From table 4.1, it is seen that \( T_{W,\beta}^a = \bar{y}_{RS,a} \) when \( W = 1 \) and \( \beta = -1 \). Therefore, letting \( W = 1 \) and \( \beta = -1 \) in the expression (4.15) we get

\[
M \left[ T_{1,-1}^a \right] = M \left[ \bar{y}_{RS,a} \right] = \left\{ \bar{Y} \left( \frac{X_a}{X} \right) - \bar{Y}_a \right\}^2 + \frac{N - n}{Nn} \bar{Y} \left( \frac{X_a}{X} \right) \left[ \bar{Y} \left( \frac{X_a}{X} \right) \left\{ 3C_x^2 + C_Y^2 - 4C_{XY} \right\} - 2\bar{Y}_a \left\{ C_X^2 - C_{XY} \right\} \right].
\]

(4.18)

4.5 OPTIMIZING \( M \left[ T_{W,\beta}^a \right] \)

4.5.1. The \( M \left[ T_{W,\beta}^a \right] \) is a function of the parameters \( W \) and \( \beta \). Therefore, in order to optimize \( M \left[ T_{W,\beta}^a \right] \) so as to obtain minimum MSE and corresponding optimum values of \( W \) and \( \beta \), it is necessary to obtain first derivatives of \( M \left[ T_{W,\beta}^a \right] \) with respect to \( W \) and \( \beta \) and equate them to zero, so that two simultaneous equations \( \frac{\partial M \left[ T_{W,\beta}^a \right]}{\partial W} = 0 \) and \( \frac{\partial M \left[ T_{W,\beta}^a \right]}{\partial \beta} = 0 \) are obtained for yielding \( W_0 \) and \( \beta_0 \), the optimum value of \( W \) and \( \beta \) respectively. However, since the expression of \( M \left[ T_{W,\beta}^a \right] \) is a complex function of \( W \) and \( \beta \), the explicit solutions \( W_0 \) and \( \beta_0 \) are not easy to obtain. We have, therefore, obtained the optimum value of \( \beta \) and \( W \) under certain assumptions.

4.5.2. Let us assume that \( \bar{X}_a = \bar{X} \). Then \( D = \bar{X}_a \). We then have

\[
\beta_0 = \frac{W (\bar{Y} - \bar{Y}_a) V_{02} + 2 (\bar{Y}_a - 2\bar{Y}) V_{11}}{2W (2\bar{Y} - Y_a) V_{02}}.
\]

(4.19)
and

\[ W_0 = \frac{(\bar{Y}_a - \bar{Y}) V_{11}}{\{ \beta (2\bar{Y} - \bar{Y}_a) + (\bar{Y}_a - \bar{Y}) \} V_{02}}. \]

(4.20)

Using (4.19) and (4.20) the optimum values of \( \beta \) and \( W \) can easily be obtained.

However, the above values of \( \beta \) and \( W \) are obtained under a very crude and unrealistic assumption, which rarely occurs.

4.5.3. It is easy to observe that when \( \beta = -1 \), we have

\[ T_{W,-1}^a = \hat{\bar{y}} \left( \frac{\bar{X}_a}{x_W} \right), \]

(4.21)

which is a generalized ratio-type synthetic estimator and can be considered a synthetic version of Walsh (1970) estimator \( t_{W,a} \), given in (4.2). Further if \( W = r = n/(N+n) \) then it reduces to \( T_{r,r}^a \) defined in chapter III. Thus, \( T_{W,-1}^a \) may be looked upon as a generalization of the estimator \( T_{r,r}^a \). The MSE of \( T_{W,-1}^a \) can be deduced from the expression (4.15) as

\[
M \left[ T_{W,-1}^a \right] = \left\{ \bar{Y} \left( \frac{\bar{X}_a}{D} \right) - \bar{Y}_a \right\}^2 + \bar{Y}^2 \left( \frac{\bar{X}_a}{D} \right)^2 V_{20} \\
- \bar{Y} \left( \frac{\bar{X}_a}{D} \right) \left( \frac{W \bar{X}}{D} \right)^2 \left[ 2\bar{Y}_a - 3\bar{Y} \left( \frac{\bar{X}_a}{D} \right) \right] V_{02} \\
- 2\bar{Y} \left( \frac{\bar{X}_a}{D} \right) \left( \frac{W \bar{X}}{D} \right) \left[ 2\bar{Y} \left( \frac{\bar{X}_a}{D} \right) - \bar{Y}_a \right] V_{11}
\]

(4.22)

which is a function of \( W \) only. Therefore, \( M \left[ T_{W,-1}^a \right] \) can be optimized with respect to \( W \) in order to get optimum value of \( W \) and corresponding minimum MSE. The equation \( \frac{\partial M [T_{W,-1}^a]}{\partial W} = 0 \) is obtained as

\[
2\bar{Y}_a D' - 2 (\bar{Y} \bar{X}_a) \left( \frac{D' \bar{X}_a}{D^2} \right) - 2 (\bar{Y} \bar{X}_a) \left( \frac{D' \bar{X}_a}{D} \right) V_{20} + 6 (\bar{Y} \bar{X}_a) \bar{X}^3 W V_{02} \\
- 12 (\bar{Y} \bar{X}_a) \bar{X}^2 W^2 \left( \frac{W D'}{D^3} \right) V_{02} - 4\bar{Y}_a \bar{X}^2 \left( \frac{W}{D} \right) V_{02}
\]
Chapter IV. On the improvement of a generalized class of synthetic estimators with its application

\[ +6 \bar{Y}_a \bar{X}^2 W^2 \left( \frac{D'}{D^2} \right) V_{02} - 4 \left( \bar{Y} \bar{X}_a \right) \left( \frac{\bar{X}}{D} \right) V_{11} + 2 \left( \bar{Y}_a \bar{X} \right) V_{11} \]

\[ +12 \left( \bar{Y} \bar{X}_a \right) \bar{X} W \left( \frac{D'}{D^2} \right) V_{11} - 4 \left( \bar{Y}_a \bar{X} \right) W \left( \frac{D'}{D} \right) V_{11} = 0, \tag{4.23} \]

where \( D' = \frac{\partial D}{\partial W} = (\bar{X} - \bar{X}_a) \).

The equation (4.23) can be satisfied so as to find \( W_0 \), the optimum value of \( W \), which minimizes the MSE of the estimator \( T_{W,-1}^a \). Needless to say that the estimator \( T_{W,-1}^a \) can be compared with some of the synthetic ratio-type estimators for observing its performance over other estimators. The minimum MSE of \( T_{W,-1}^a \) can be computed by substituting the value of \( W \), obtained from (4.23), in (4.22).

**Remark 6.** The equation (4.23) reveals the fact that it may yield more than one optimum values of \( W \). However, since \( W \in (0,1) \), only those values of \( W \) will be considered which lies in the range \((0,1)\).

**Remark 7.** The estimator \( T_{W,-1}^a \), as a member of the family generated by \( T_{W,\beta}^a \), is comparable directly with some of the existing ratio type/indirect estimators for observing its performance through some numerical study, since the expressions of MSEs of the estimators are not theoretically comparable. However, the estimator \( T_{W,-1}^a \) is a function of the weight \( W \), where \( 0 < W < 1 \), the comparison may be made either selecting some extreme values of \( W \) or selecting the optimum value of \( W \) (say, \( W_0 \)). In the next section, we have presented a direct comparison of MSEs of different synthetic estimators based upon an empirical data.

### 4.6 COMPARISON OF DIFFERENT ESTIMATORS

#### 4.6.1. We observe that the generalised class of synthetic estimators, \( \bar{y}_{sym,a} = \bar{y} \left( \frac{x}{\bar{x}} \right)^\beta \)

proposed by Tikkiwal and Ghiya (2000) reduces to the form \( \bar{y} \left( \frac{x}{\bar{x}} \right) \) when \( \beta = -1 \). In fact, it is the ratio synthetic estimator \( \bar{y}_{RS,a} \) or \( T_{1,a} \) discussed in chapter II and also proposed
by Rao (2003). Therefore, the members of the family $T_{W,\beta}$, proposed in this chapter, may be compared with $\bar{y}_{RS,a}$ for $\beta = -1$. The estimators, $T_{D,a}$ (Direct ratio estimator), $\bar{y}_{SS,a}$ (simple synthetic estimator) may also be taken into account for the comparison purpose. The estimator $T_{W,\beta}$ has been considered for $\beta = -1$ and $W = 0.1, 0.9$ and $W_0$ for the comparison purpose.

### 4.7 EMPIRICAL DATA AND EFFICIENCY COMPARISON

#### 4.7.1

We have considered the data of MU284 population given in Appendix B in Sarndal et al. (1992) for the illustration purpose. As mentioned in earlier chapter, the data relate to the population of Sweden divided into 284 municipalities spread over four regions. The data show values on several variables for all the 284 municipalities. Since some of the regions have very small number of municipalities, these might be considered as small domains in comparison of the population of Sweden. If the problem is to estimate the mean of the study variable for these small domains, indirect estimators might be helpful.

We consider the following two variables $Y$ and $X$, $Y$ being study and $X$ being auxiliary variable:

- $Y$: The total number of seats in municipal council
- $X$: The number of conservative seats in municipal council.

We have considered only the east, south and central regions (region indicators 1, 2, 3, 6, 7 and 8) of sizes 25, 48, 32, 41, 15 and 29 respectively as six small domains. Here $N = 190$.

The population values as well as domain-specific values are already depicted in Table 2.1 of chapter II. The absolute differences between the ratio $\frac{\bar{Y}}{\bar{X}}$ and $\frac{\bar{Y}}{\bar{X}}$ for all the domains are shown in Table 2.2 which provide an idea how closely the synthetic assumption are met in these domains. We take $n = 19$ (a 10 percent sample from the population). In order to compute MSE of the direct estimator, $T_{D,a}$, an equivalent size sample has been considered to be drawn from each domain, that is, the criterion $\frac{n_a}{N_a} = \frac{n}{N}$ is satisfied for selecting the sample size $n_a$. 

4.7.2. We have considered the following estimators for comparison purpose:

(i) **Direct estimator (Direct ratio estimator):**

\[ T_{D,a} = \bar{y}_a \left( \frac{\bar{X}_a}{\bar{x}_a} \right) \]  \hspace{1cm} (4.24)

with

\[ M[T_{D,a}] = \frac{N_a - n_a}{N_a n_a} \left[ S_{Y_a}^2 + R_{N_a}^2 S_{X_a}^2 - 2R_{Y_a;X_a} S_{Y_a} S_{X_a} \right]. \]  \hspace{1cm} (4.25)

(ii) **Simple synthetic estimator:**

\[ \bar{y}_{SS,a} = \bar{y} \]  \hspace{1cm} (4.26)

with

\[ M[\bar{y}_{SS,a}] = \frac{N - n}{N n} S_Y + (\bar{Y} - \bar{Y}_a)^2. \]  \hspace{1cm} (4.27)

(iii) **Indirect estimators:**

(a) **Ratio synthetic estimator:**

\[ \bar{y}_{RS,a} = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}} \right) \]  \hspace{1cm} (4.28)

with

\[ M[\bar{y}_{RS,a}] = \left\{ \bar{Y} \frac{\bar{X}_a}{X} - \bar{Y}_a \right\}^2 + \frac{N - n}{N n} \left( \bar{Y} \frac{\bar{X}_a}{X} \right) \]

\[ \left[ \left( \frac{\bar{Y} \bar{X}_a}{X} \right) \left\{ 3C_X^2 + C_Y^2 - 4C_{XY} \right\} - 2\bar{Y}_a \left\{ C_X^2 - C_{XY} \right\} \right] \]  \hspace{1cm} (4.29)

(b) **Ratio-type synthetic estimators** \( T_{0.1,-1}^a, T_{0.9,-1}^a \) and \( T_{W_0,-1}^a \):

which can be deduced from (4.3) with corresponding MSEs from the expression (4.15).

The value \( W = W_0 \), the optimum value of \( W \) is obtained using the equation (4.23).
4.7.3. Using the empirical data, we have compared the values of MSEs of the above mentioned estimators on the basis of the Percent Relative Efficiency (PRE) with respect to $T_{D,a}$. The results are shown in the following table:

(bracketed figures show the PRE of the estimators with respect to $T_{D,a}$)

MSE and PRE of estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Domain 1</th>
<th>Domain 2</th>
<th>Domain 3</th>
<th>Domain 6</th>
<th>Domain 7</th>
<th>Domain 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{D,a}$</td>
<td>51.14</td>
<td>56.09</td>
<td>26.99</td>
<td>63.83</td>
<td>134.76</td>
<td>61.92</td>
</tr>
<tr>
<td></td>
<td>(100.00)</td>
<td>(100.00)</td>
<td>(100.00)</td>
<td>(100.00)</td>
<td>(100.00)</td>
<td>(100.00)</td>
</tr>
<tr>
<td>$\bar{y}_{SS,a}$</td>
<td>18.57</td>
<td>6.52</td>
<td>13.08</td>
<td>7.80</td>
<td>48.90</td>
<td>63.07</td>
</tr>
<tr>
<td></td>
<td>(275.39)</td>
<td>(806.28)</td>
<td>(206.35)</td>
<td>(818.33)</td>
<td>(275.58)</td>
<td>(98.18)</td>
</tr>
<tr>
<td>$\bar{y}_{RS,a}$</td>
<td>1848.17</td>
<td>24.21</td>
<td>57.88</td>
<td>71.56</td>
<td>372.74</td>
<td>283.98</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
<td>(231.68)</td>
<td>(46.63)</td>
<td>(89.20)</td>
<td>(36.15)</td>
<td>(21.80)</td>
</tr>
<tr>
<td>$T_{aW,\beta}$ with $\beta = -1$, $W = 0.1$</td>
<td>7.11</td>
<td>4.62</td>
<td>8.82</td>
<td>6.02</td>
<td>73.31</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>(719.27)</td>
<td>(1214.07)</td>
<td>(306.01)</td>
<td>(1000.30)</td>
<td>(183.82)</td>
<td>(870.89)</td>
</tr>
<tr>
<td>$T_{aW,\beta}$ with $\beta = -1$, $W = 0.9$</td>
<td>1183.66</td>
<td>19.23</td>
<td>41.05</td>
<td>52.35</td>
<td>211.66</td>
<td>140.33</td>
</tr>
<tr>
<td></td>
<td>(4.32)</td>
<td>(291.68)</td>
<td>(65.74)</td>
<td>(121.93)</td>
<td>(63.67)</td>
<td>(44.12)</td>
</tr>
<tr>
<td>$T_{W,-1}$ at opt $W$</td>
<td>5.73</td>
<td>3.46</td>
<td>3.97</td>
<td>3.79</td>
<td>71.72</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>(892.50)</td>
<td>(1621.10)</td>
<td>(679.85)</td>
<td>(1684.17)</td>
<td>(187.90)</td>
<td>(2517.07)</td>
</tr>
</tbody>
</table>

4.7.4 Concluding Remarks on the Tabulated Values

On the basis of results shown in the above table, the following conclusions are evident:

(a) The optimum estimator $T_{W_0,-1}$ is uniformly more efficient than other synthetic estimators considered. Moreover, the estimator is highly efficient than the direct estimator $T_{D,a}$ for all the domains except the domain 7, where it is moderately efficient.

(b) As $0 < W < 1$, therefore, it is desirable to see the performance of the estimator $T_{W,-1}$ for values of the parameter $W$ other than $W_0$. We have, therefore, considered the cases when $W = 0.1$ and 0.9. It is clear that while $T_{0.1,-1}$ for $W = 0.1$ is considered it is also more efficient than $T_{D,a}$ over all the domains, while $T_{W,-1}$ for $W = 0.9$ seems to be efficient only for domains 2 and 6.
(c) The fact mentioned in (b) seems to be trivial. Since $W$ and $(1 - W)$ are respectively the weights attached to $\bar{x}$ and $\bar{X}_a$ in the linear combination $\bar{x}_W$, a larger value of $W$ means giving more weight to the sample mean $\bar{x}$, rather than to known value $\bar{X}_a$, the domain-specific mean of the variable $X$. If $W$ is comparatively smaller one believes more on the known mean $\bar{X}_a$ in comparison to borrowed sample mean $\bar{x}$. It is also clear from the optimum value, $W_0$, which generally occurred in the range $(0.1, 0.35)$, that is, closer to 0.1.

(d) The ratio synthetic estimator $\bar{y}_{RS,a}$ is generally not preferable over $T_{D,a}$ except for domain 2. This is perhaps due to the fact that for domain 2, the synthetic assumption $\frac{\bar{Y}_a}{\bar{Y}} = \frac{\bar{X}_a}{\bar{X}}$, as mentioned in chapter II, is closely met, which is a pre-requisite for the ratio-synthetic estimator to be more efficient.

In a nutshell, it can be concluded that giving more weightage to the known population mean $\bar{X}_a$ in comparison to sample mean $\bar{x}$, the estimator $T_{W,\beta}$ can be made more efficient than other synthetic/direct estimators which have been considered in the present work.

4.8 A SIMULATION STUDY

In the previous section, we made a comparative study of several synthetic estimators on the basis of their MSEs, which have been computed on the basis of some known population parameters. However, it is criticised that such a comparison does not provide a clear picture of the situation, since the study does not take into account the possible samples which could be drawn from the population and sample values. Due to this fact, simulation study of estimators is preferred depending upon as many samples as possible and consequently on actual sample values.

We have already mentioned the process of conducting simulation study of estimators
in previous two chapters. A similar study is conducted in the present chapter also for the estimator \( T_{W,\beta} \) and other comparable estimators.

We select \( \beta = -1 \) and \( W = W_0 \) for the simulation study. We select 500 independent samples of size 19 each from the population of 190, selected randomly and the ARB and SRSE of Direct Estimator \( T_{D,a} \), generalised synthetic ratio estimator \( \bar{y}_{RS,a} \) (same as \( \bar{y}_{sym,a} \) for \( \beta = -1 \)) and the estimator \( T_{W,\beta}^a \) for \( \beta = -1 \) and \( W = W_0 \) were computed on the basis of estimates obtained from these 500 samples.

The values of ARB and SRSE for the estimators are depicted in the Table 4.3. The simulation study reveals the fact that, on the average, values of the estimates, obtained through the estimator \( T_{W,\beta}^a \), computed over a large number of samples, are much closer to the domain-specific mean \( \bar{Y}_a \). The highest difference is observed for domain 7, perhaps due to not meeting the synthetic assumption in this domain.

### 4.9 CONCLUSIONS

Concluding the results obtained for the suggested estimator \( T_{W,\beta}^a \) in the above paragraphs, it can be concluded that the estimator and some of its members do perform better than some other synthetic estimators which have been members of the families discussed in the previous chapter. However, the estimator seems to be worse than these estimators in
Chapter IV. On the improvement of a generalized class of synthetic estimators with its application 79
certain small domains if optimality condition is not observed. This is quite obvious since
the estimator assigns weights to the two means, namely, $\bar{x}$ and $\bar{X}_a$ and if weighting is not
appropriately made, the suggest strategy may not be preferable. Hence, the performance
of the estimator heavily depends upon the weight $W$. Due to complex nature of the MSE
expression, the problem of finding optimum choice of the other parameter, $\beta$ is complex
and hence, has not been discussed here.