Jour. Ind. Soc. Ag. Statistics
Vol. XXXIX, No. 2 (1987), pp. 181-190

ROBUSTNESS OF SOME RATIO-TYPE ESTIMATORS UNDER SUPER-POPULATION MODEL

RAJU SINGH, B. V. S. SISODIA* and R. B. SINGH
Department of Mathematics, Bihar University, Muzaffarpur

(Received : February, 1986)

SUMMARY

Following Royall and Herson [4] the Robustness of some ratio-type estimators under super-population probability model is discussed. It has been found that the ratio-type estimators dealt with in this paper are optimal under balanced sample and in case of unbalanced sample the ratio-type estimators, particularly, proposed by Sisodia and Dwivedi [5] are more robust than ratio estimator under deviation from the super-population probability model.

Keywords : Robustness; Balanced sample; polynomial regression model; prediction approach; unbalanced sample.

Introduction

When certain features of the assumed super-population probability model are incorrect, the optimal properties of estimator need not hold. Royall and Herson [4] showed that the ratio estimator remained optimal with balanced sample even if the specification of the model was incorrect. In the present paper the effect of misspecification of super-population probability model is studied on some ratio-type estimators under balanced and unbalanced samples. Robustness of ratio and ratio-type estimators is also studied.

*Department of Statistics & Mathematics, Faculty of Basic Science & Humanities, R. A. U., Pusa (Samastipur)-848125.

2. The Models

The population of interest consists of N units labelled 1, 2, . . . , N. Associated with unit K are two numbers (x_k, y_k) with x_k known and y_k fixed but unknown. Let a sample consisting of n units be selected from the population and the y-values associated with the sample units be observed.

The objectives is to estimate the mean $\bar{Y} = \frac{1}{N} \sum_{1}^{N} y_k$

The numbers $y_1, y_2, \ldots, y_N$ whose mean we must estimate are treated as realised values of independent random variables $Y_1, Y_2, \ldots, Y_N$ such that

$$Y_k = h(x_k) + \epsilon_k [V(x_k)]^{1/2} \quad (2.1)$$
$$K = 1, 2, \ldots, N$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are independent random variables each having mean zero and variance σ^2. The expected value and variance of Y_k are $h(x_k)$ and $\sigma^2 V(x_k)$, respectively. The $h(x)$ is a polynomial of order (at most) J, that is

$$h(x) = \delta_0\beta_0 + \delta_1\beta_1 x + \delta_2\beta_2 x^2 + \ldots + \delta_J\beta_J x^J \quad (2.2)$$

where δ's are either zero or one. We refer to the above probability model as

$$\xi[\delta_0, \delta_1, \delta_2, \ldots, \delta_J : V(x)] \quad (2.3)$$

There are two fundamentally different approaches to finite population sampling theory viz. conventional and prediction or model-based (Royall, [3]). We here confine to robustness problems in model based inference when the assumed model is not the true model.

3. Ratio-type Estimators and their Bias and MSE under the Models

To estimate $\bar{Y}$, Sisodia and Dwivedi [5] proposed the ratio-type estimator

$$T_{sd} = (1 - a)\, \bar{y}_s + a\bar{y}_s \left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} \quad (3.1)$$

where a and α are scalar quantities. $\bar{y}_s$ and $\bar{x}_s$ are sample means based on sample s of fixed size n, and $\bar{x} = \frac{1}{N} \sum_{1}^{N} x_k$ If $\alpha = 1$, then T_{sd} reduces to Chakarbarty's estimator say, T_c as

$$T_c = (1 - a)\,\bar{y}_s + a\bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right) \tag{3.2}$$

if $a = 1$, then T_{sd} reduces to Srivastava estimator, say, T_s as

$$T_s = \bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} \tag{3.3}$$

and if $a = 1$ and $\alpha = 1$ together, then T_{sd} reduces to simple ratio estimator, say, T_R as

$$T_R = \bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right) \tag{3.4}$$

Under the linear model $\xi(0, 1 : V(x))$, for any variance function $V(x)$, it has been shown by Royall and Herson [4] that ratio estimator T_R is unbiased, i.e. $E_\xi[T_R - \bar{Y}] = 0$ and for the variance function $V(x) = x$,

its variance is $V(T_R) = E_\xi(T_R - \bar{Y})^2 = \dfrac{\sigma^2}{N} \sum\limits_{\bar{s}} x_k \Big/ \sum\limits_{s} x_k \cdot \bar{x}$ (3.5)

where s denotes the non-sampled units.

3.1. *Expected Value and M.S.E. of Ratio-type Estimators*

The bias of ratio-type estimator T_{sd} under the model $\xi(0, 1 : V(x))$ is

$$B(T_{sd}) = E_\xi[T_{sd} - \bar{Y}] = E_\xi\left[(1 - a)\,\bar{y}_s + a\bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} - \bar{Y}\right]$$

$$= \beta_1\left[(\bar{x}_s - \bar{x}) + a\left\{\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} - 1\right\}\bar{x}_s\right] \neq 0 \tag{3.6}$$

Thus the estimator T_{sd} is biased under the linear model $\xi(0, 1 : V(x))$ while the ratio estimator is unbiased. The M.S.E. of T_{sd} under $\xi(0, 1 : V(x))$ is derived as follow :

$$\text{M.S.E.}\ (T_{sd}) = E_\xi(T_{sd} - \bar{Y})^2$$

$$= \beta_1^2\left[(\bar{x}_s - \bar{x}) + a\left\{\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} - 1\right\}\bar{x}_s\right]^2 + \left\{\left(\frac{1}{n} - \frac{1}{N}\right)\right.$$

$$\left. + \frac{a}{n}\left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\alpha} - 1\right)\right\}^2 \sigma^2 \sum_{s} V(x_k) + \frac{\sigma^2}{N^2} \sum_{\bar{s}} V(x_k)$$

For the variance function $V(x_k) = x_k$, it reduces to

$$\text{M.S.E. } (T_{sd}) = \beta_1^2 \left[(\bar{x}_s - \bar{x}) + a \left\{ \left(\frac{\bar{x}}{\bar{x}_s} \right)^{\alpha} - 1 \right\} \bar{x}_s \right]^2 + \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) + \frac{a}{n} \left(\left(\frac{\bar{x}}{\bar{x}_s} \right)^{\alpha} - 1 \right) \right\}^2 \sigma^2 n \bar{x}_s + \frac{\sigma^2}{N^2} (N - n)\, \bar{x}_{\bar{s}} \quad (3.7)$$

which is the sum of variance of the estimator and square of its bias. $\bar{x}_{\bar{s}}$ is the mean of $(N - n)$ non-sampled units.

The estimators T_c and T_s are particular cases of T_{sd} which are obtained by putting $\alpha = 1$ and $a = 1$, respectively in T_{sd}. Hence, the bias and M.S.E. of these estimators will be given by

$$B(T_c) = E_\xi(T_c - \bar{Y}) = \beta_1 \left[(\bar{x}_s - \bar{x}) + a \left(\frac{\bar{x}}{\bar{x}_s} - 1 \right) \bar{x}_s \right] \quad (3.8)$$

$$B(T_s) = E_\xi(T_s - \bar{Y}) = \beta_1 \left[(\bar{x}_s - \bar{x}) + a \left\{ \left(\frac{\bar{x}}{\bar{x}_s} \right)^{\alpha} - 1 \right\} \bar{x}_s \right] \quad (3.9)$$

$$\text{M.S.E. } (T_c) = E_\xi(T_c - \bar{Y})^2 = \beta_1^2 \left[(\bar{x}_s - \bar{x}) + a \left(\frac{\bar{x}}{\bar{x}_s} - 1 \right) \bar{x}_s \right]^2 + \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) + \frac{a}{n} \left(\frac{\bar{x}}{\bar{x}_s} - 1 \right) \right\} \sigma^2 n \bar{x}_s + \frac{\sigma^2}{N^2} (N - n)\, \bar{x}_{\bar{s}} \quad (3.10)$$

and

$$\text{M.S.E. } (T_s) = E_\xi(T_s - \bar{Y})^2 = \beta_1^2 \left[(\bar{x}_s - \bar{x}) + \left\{ \left(\frac{\bar{x}}{\bar{x}_s} \right)^{\alpha} - 1 \right\} \bar{x}_s \right]^2 + \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) + \frac{1}{n} \left(\left(\frac{\bar{x}}{\bar{x}_s} \right)^{\alpha} - 1 \right) \right\}^2 \sigma^2 n \bar{x}_s + \frac{1}{N^2} \sigma^2 (N - n)\, \bar{x}_{\bar{s}} \quad (3.11)$$

Also by putting $\alpha = 1$ and $a = 1$ in the estimator T_{sd} the ratio estimator T_R is obtained and hence if we put $\alpha = 1$ and $a = 1$ in the bias and M.S.E. of T_{sd} the bias and the M.S.E. of the estimator T_R is obtained as

$$E_\xi(T_R - \bar{Y}) = 0 \quad \text{and} \quad E_\xi(T_R - \bar{Y})^2 = \frac{1}{N} \sigma^2 \sum_{s} x_k \Big/ \sum_{\bar{s}} x_k \cdot \bar{x}$$

which have already been derived by Royall and Herson (1973).

If the specification of the model is wrong, i.e. the true model is $\xi(\delta_0, \delta_1, \delta_2, \ldots, \delta_J : V(x))$, then bias of the estimator T_{sd} under this general model is

$$E_\xi(\hat{T}_{sd} - \bar{Y}) = E_\xi\left[(1-a)\,\bar{y}_s + a\bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - \bar{Y} \right]$$

$$= (1-a)\,\frac{1}{n}\sum_s\left(\sum_{j=0}^{J}\delta_j\beta_j x_k^j\right) + \frac{a}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha\sum_s\left(\sum_{j=0}^{J}\delta_j\beta_j x_k^j\right)$$

$$-\frac{1}{N}\sum_{1}^{N}\left(\sum_{j=0}^{J}\delta_j\beta_j x_k^j\right)$$

$$= \sum_{j=0}^{J}\delta_j\beta_j\left\{\left(\bar{x}_s^{(j)} - \bar{x}^{(j)}\right) + a\left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right)\bar{x}_s^{(j)}\right\} \quad (3.12)$$

As the M.S.E. is sum of the variance and square of bias, the M.S.E. of the estimator T_{sd} under this general polynomial regression model for the variance function $V(x) = x$ is

$$E_\xi(T_{sd} - \bar{Y})^2 = \left[\sum_{j=0}^{J}\delta_j\beta_j\left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + a\left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right)\bar{x}_s^{(j)}\right\}\right]^2$$

$$+\left\{\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{a}{n}\left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right)\right\}^2\sigma_n^2\bar{x}_s$$

$$+\frac{1}{N^2}\,\sigma^2\,(N-n)\,\bar{x}_{\bar{s}}$$

By putting $\alpha = 1$ we get the bias and M.S.E. of the estimator T_c, respectively, as follow :

$$E_\xi(T_c - \bar{Y}) = \sum_{j=0}^{J}\delta_j\beta_j\left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + a\left(\frac{\bar{x}}{\bar{x}_s} - 1\right)\bar{x}_s^{(j)}\right\} \quad (3.14)$$

and

$$E_\xi(T_c - \bar{Y})^2 = \left[\sum_{j=0}^{J}\delta_j\beta_j\left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + a\left(\frac{\bar{x}}{\bar{x}_s} - 1\right)\bar{x}_s^{(j)}\right\}\right]^2$$

$$+\left\{\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{a}{n}\left(\frac{\bar{x}}{\bar{x}_s} - 1\right)\right\}\sigma_n^2\bar{x}_s$$

$$+\frac{1}{N^2}\,\sigma^2\,(N-n)\,\bar{x}_{\bar{s}}\,. \quad (3.15)$$

Similarly, the bias and M.S.E. of the estimator T_s is obtained by putting $a = 1$ in (3.12) and (3.13) as follow :

$$E_\xi(T_s - \bar{Y}) = \sum_{j=0}^{J} \delta_j\beta_j \left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + \left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right) \bar{x}_s^{(j)}\right\} \quad (3.16)$$

and

$$E_\xi(T_s - \bar{Y})^2 = \left[\sum_{j=0}^{J} \delta_j\beta_j \left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + \left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right) \bar{x}_s^{(j)}\right\}\right]^2$$
$$+ \left\{\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{1}{n}\left(\left(\frac{\bar{x}}{\bar{x}_s}\right)^\alpha - 1\right)\right\}^2 \sigma_n^2 \bar{x}_s$$
$$+ \frac{1}{N^2} \sigma^2 (N - n)\, \bar{x}_{\bar{s}}. \quad (3.17)$$

Also by putting $a = 1$ and $\alpha = 1$ in equation (3.12) and (3.13) we get the bias and MSE of ratio estimator T_R under the general polynomial regression model $\xi(\delta_0, \delta_1, \delta_2, \ldots \delta_J : x)$ as follow :

$$E_\xi(T_R - \bar{Y}) = \sum_{j=0}^{J} \delta_j\beta_j \left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + \left(\frac{\bar{x}}{\bar{x}_s} - 1\right) \bar{x}_s^{(j)}\right\} \quad (3.18)$$

and

$$E_\xi(T_R - \bar{Y})^2 = \left[\sum_{j=0}^{J} \delta_j\beta_j \left\{(\bar{x}_s^{(j)} - \bar{x}^{(j)}) + \left(\frac{\bar{x}}{\bar{x}_s} - 1\right) \bar{x}_s^{(j)}\right\}\right]^2$$
$$+ \left\{\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{1}{n}\left(\frac{\bar{x}}{\bar{x}_s} - 1\right)\right\}^2 \sigma_n^2 \bar{x}_s$$
$$+ \frac{1}{N^2} \sigma^2 (N - n)\, \bar{x}_{\bar{s}} \quad (3.19)$$

4. Robustness Criterion

If the $\xi(0, 1 : x)$ is the assumed model and $\xi(\delta_0, \delta_1, \ldots, \delta_J : x)$ is the true model, we see in the preceding section that only the bias of the ratio-type estimators is affected while the variance remains the same under misspecification of the model. But if the selected sample is balanced, i.e. $\bar{x}_s^{(j)} = \bar{x}^{(j)}$, $(j = 1, 2, \ldots, J)$ the ratio-type estimators under study including ratio estimator become unbiased and M.S.E. of these estimators reduce, after little simplification, to $\sigma^2 (N - n)/n\ \bar{x}$ whatever the model may be. Thus with the balanced sample the ratio estimator including ratio-type estimators remain optimal in the sense of efficiency whatever the model may be. However, to realise a balanced sample is very difficult in practice if not impossible. It is therefore important to investigate the robustness of the estimators with unbalanced sample when the model is misspecified. To determine the robustness of the estimators, a criterion has to be fixed up. As far as with the ratio-type estimators, it has been seen in the section-3 that only biases are affected with the deviation of the model. We consider change in the amount of M.S.E. with the deviation of the model as a criterion to determine the robustness of the estimator. If the change in the amount of M.S.E. of an estimator with the deviation of the model is nominal, the estimator is said to be robust.

5. Robustness of the Estimators

It is difficult to examine the robustness of the estimators theoretically under the deviation of the super-population probability models. For this purpose we consider the following 3 working models.

Model I : $E_\xi(Y) = 2X$
Model II : $E_\xi(Y) = 4 + 2x$ and
Model III : $E_\xi(Y) = 4 + 2x - 1.5x^2$

Let the population of interest be consist of $N = 11$ units, and let sample size be $n = 3$. Without loss of generality we assume that $\bar{x} = 6$ and two unbalanced sample means say, $\bar{x}_s = 9$ $(> \bar{x})$ and $\bar{x}_s = 2$ $(< \bar{x})$. Both the situations of unbalanced sample i.e., sample mean greater than or less than the population mean have been considered.

Tables 1 and 2 show the figures of difference of M.S.E. of the estimators under different models for both the situations of unbalanced sample, i.e. $\bar{x} > \bar{x}_s$ and $\bar{x} < \bar{x}_s$, respectively. Four values of a, i.e. $\frac{1}{4}$, $\frac{1}{2}$, $\frac{2}{3}$ and 1 and four values of α, i.e. 0.5, 1, 1.5 and 2 are considered for the purpose of illustration.

If the model I is the assumed model and $\bar{x} > \bar{x}_s$ while Model II is the true model, it is obvious from the Table-1 that the differences in M.S.E. of ratio-type estimator T_{sd} under model I and II are small for all values of $a(0 < a < 1)$ and $\alpha = 0.5$ to 1.5. Such little differences will have negligible effect on the efficiency of the estimator T_{sd} even if model-I is used instead of model-II. Thus, the estimator T_{sd} is robust for model-I and II. It may be noted that $T_{sd} = T_c$ when $\alpha = 1$ implying thereby the T_c is also robust for model-I and II.

When model III is true, the differences in M.S.E. of the estimators under model I and III and model II and III are considerably high. This

TABLE 1—DIFFERENCE OF M.S.E. UNDER DIFFERENT MODEL OF DIFFERENT ESTIMATORS (FOR $\bar{x} > \bar{x}_s$)

Estimator	a / α	$\frac{1}{4}$			$\frac{1}{2}$		
		Model I—II	*Model I—III*	*Model II—III*	*Model I—II*	*Model I—III*	*Model II—III*
T_{sd}	0.5	10.212	—2882.148	—2892.832	17.067	—2912.238	—2929.305
	1.0	20.000	—3269.100	—3289.100	16.000	—3233.00	—3249.00
	1.5	14.318	—3014.931	—3029.249	—76.936	—3145.936	—3069.00
	2.0	−64.000	—3136.000	—3072.000	—512.00	—3298.84	—2786.84

Estimator	a / α	$\frac{2}{3}$			1		
		Model I—II	*Model I—III*	*Model II—III*	*Model I—II*	*Model I—III*	*Model II—III*
T_{sd}	0.5	19.874	—2931.468	—2951.342	21.126	—2969.667	—2990.793
	1.0	00.054	—3203.226	—3203.28	—64.00	—3134.880	—3070.88
	1.5	—196.322	—3214.93	—3018.608	—576.049	—3308.812	—2732.763
	2.0	—1023.626	—3342.359	—2318.733	—2560.000	—3268.000	—708.000

At $\alpha = 1$, and $a = 1$ $T_{sd} = T_R'$ at $\alpha = 1$, $T_{sd} = T_c'$ at $a = 1$, $T_{sd} = T_s$.

TABLE 2—DIFFERENCE OF M.S.E. UNDER DIFFERENT MODEL OF DIFFERENT ESTIMATORS (FOR $\bar{x} < \bar{x}_s$)

Estimator	a / α	$\frac{1}{4}$ Model I—II	$\frac{1}{4}$ Model I—III	$\frac{1}{4}$ Model II—III	$\frac{1}{2}$ Model I—II	$\frac{1}{2}$ Model I—III	$\frac{1}{2}$ Model II—III
T_{sd}	0.5	1.828	—1811.962	—1813.790	2.995	—1445.755	—1448.75
	1.0	2.945	—1510.124	—1513.069	3.572	—935.947	—939.519
	1.5	3.343	—1284.079	—1287.422	2.612	—601.588	—604.200
	2.0	3.537	—1112.751	—1116.288	0.968	—382.526	—383.494

Estimator	a / α	$\frac{2}{3}$ Model I—II	$\frac{2}{3}$ Model I—III	$\frac{2}{3}$ Model II—III	1 Model I—II	1 Model I—III	1 Model II—III
T_{sd}	0.5	3.474	—1225.376	—1228.850	3.395	—836.666	—840.061
	1.0	—19.814	—629.025	—609.211	—1.768	—195.720	—193.952
	1.5	—0.182	—287.701	—287.519	—12.160	1.076	13.236
	2.0	—4.144	—105.037	—100.893	—22.800	—53.388	—30.708

N. B. : At $\alpha = 1$ and $a = 1$, $T_{sd} = T_R$, at $\alpha = 1$, $T_{sd} = T_c$ · at $a = 1$ · $T_{sd} = T_s$.

shows that when the true model is quadratic and the assumed model is linear the efficiency of ratio type estimators including ratio estimator will be badly affected and, therefore, these estimators are not robust in these situations. It may also be pointed out that even if we could call ratio-estimator $T_R(T_{sd} = T_R$ when $\alpha = 1$ and $a = 1)$ as a robust in a situation where model II is true while model I is the assumed model, it will comparatively be less robust than T_{sd}.

Table 2 described the case where $\bar{x} < \bar{x}_s$. When model II is considered to be true while model I is the assumed model, one can observe from Table 2 that the differences in M.S.E. of the ratio-type estimators including ratio-estimator under model I and II are small. Thus ratio-type estimators including ratio-estimator could be considered as robust for model I and II. If model III is true the results are almost similar to those of Table 1 $(\bar{x} > \bar{x}_s)$. It may, however, be seen from Table 2 that the estimator $T_s(T_{sd} = T_s$ when $a = 1)$ for $\alpha = 1.5$ could be viewed as robust for all the models I, II and III because the differences in M.S.E. are not considerably substantial to affect the efficiency of the T_s under the deviation of the model.

It is noted that with appropriate choice of unbalanced sample and reasonable values of a and α, the ratio-type estimator T_{sd} could be considered as robust for the model I and II. The estimator T_{sd} is more robust than ratio estimator T_R for model I and II when unbalanced sample is $\bar{x}_s < \bar{x}$. The ratio-type estimators and ratio-estimator are almost equally robust for model I and II in case $\bar{x}_s > \bar{x}$. The only estimator which could be said as robust for all the models I, II and III is T_s at $\alpha = 1.5$.

REFERENCES

[1] Chakrabarty, R. P. (1979): Some ratio-type estimators, *J. Indian Soc. Agric. Statistics*, **31** : 49-62.

[2] Royall, R. M. (1970) : On finite population sampling theory under certain Linear Regression Models, *Biometrika*, 57 (2) : 377-87.

[3] Royall, R. M. (1971) : Linear regression Models in finite population sampling theory. *In* Godambe, V. P. and Sprott, D. A. (eds.), *Foundations of Statistical Inference*, Torento : Holt, Rinehart and Wiston of Canada Ltd.

[4] Royall, R. M. and Herson, H. J. (1973) : Robust estimation in finite population I, *Journal of the American Statistical Association*, 68 (344) : 880-89.

[5] Sisodia, B. V. S. and Dwivedi, V. K. (1981) : A class of ratio-cum-product-type estimators, *Biom. J.*, **23** (2) : 133-39.

[6] Sukhatme, P. V. and Sukhatme, B. V. (1976) : *Sampling Theory of Surveys with Application*, Indian Society of Agricultural Statistics, New Delhi—India.