ISSN 2229-5518

Study of predictive approach type estimator under Polynomial Regression Model

Dr. Ran Vijay Kumar Singh

Abstract: In the predictive approach, the objective of the statistician is to predict the mean of the unobserved units of the population on the basis of observed units in sample. If the product estimator is used as a predictor for the mean of the unobserved units in the population it is called predictive approach product estimator. In present paper a factor type class of estimator is developed which include predictive approach product estimator (PAPE) as a particular case. The Expression for bias and mean square error have been derived under superpopulation model. The robustness efficiency of proposed estimator under the misspecification in the superpopulation model has been observed. The effect of balancing and approximate balancing of samples on the behaviour of the estimators has been illustrated on real population data set.

Keywords: Bias, Balancing of sample,Mean square error, Predictive approach product estimator, Robustness, Superpopulation model

1 INTRODUCTION

Some of the important contributions in the direction of defining unbiased or almost unbiased product type estimators have been made by Gupta and Adhvarya (1982), Iachan et al (1983), Shrivastava (1983). The fact that product type estimator has superiority over sample mean estimator when the correlation between the study and auxiliary variable in the population is negatively high, led the survey statisticians to focus their attention on the modification of such conventional estimator so that the modified estimators can work efficiently even if the correlation is low. Consequently, a number of modified product estimators came into existence in recent past. Such estimators are generally developed either using one or more unknown constants or introducing a convex linear combination of sample and population means of the auxiliary characteristics with unknown weights. In defining modified estimators based on unknown parameters, the statistician actually develops a class of estimators which includes a number of classical estimators as members of the class and, thus enables him to make a unified study of several estimators. In this paper, a class of factor type estimator is developed which include predictive approach product estimator.

Dr. Ran Vijay Kumar Singh, Senior Lecturer & Head, Dept. of Maths, KSUST, Aliero, Nigeria.

The statistical inference theory for finite population derived from the probability distribution created by the sampler's choice of random sampling plan has been dominated for years. In contrast to this tendency, recent work in finite population sampling theory has supported proposition that many sampling problems can be analysed usefully and realistically as prediction problems under superpopulation probability models. However, the validity of the inference based on modal dependent approach depends on the correct specification of the model ascribing real world situations. It is obvious that there may be two type of misspecification, viz.,(i) misspecification in polynomial regression ; $h(x_k)$ and misspecification in the variance function ; $v(x_k)$. It is therefore desirable to investigate the effect of both of these types of misspecifications on the efficiency of an estimator. The present work is devoted to investigate the effect on the efficiency of developed factor type class of estimator when the superpopulation model is misspecified in its polynomial regression and variance function parts.

2. PROPOSED ESTIMATOR

The proposed one parameter family of factor type estimator is as follow:

$$t_{PF}^* = \bar{y}_s \left[\frac{\Psi_P^* \{\phi(\delta)\}}{\Psi_P^* \{\phi_1(\delta)\}} \right] \qquad (1)$$

Where $\Psi_P^* \{\phi(\delta)\} = 1 + \phi(\delta) \frac{\bar{x}_s}{\bar{X}} - \phi(\delta)$

$\Psi_P^* \{\phi_1(\delta)\} = \{1 - \phi_1(\delta)\} + \phi_1(\delta) \frac{\bar{x}_s}{\bar{X}}$ and

$$\phi(\delta) = \phi_1(\delta) - \phi_2(\delta); \ \phi_1(\delta) = \frac{fB}{(A + fB + C)};$$

$$\phi_2(\delta) = \frac{C}{(A + fB + C)}; \ A = (\delta - 1)(\delta - 2), B = (\delta - 1)(\delta - 4),$$

$$C = (\delta - 2)(\delta - 3)(\delta - 4); \ f = \frac{n}{N}; \ \delta > 0.$$

3 PARTICULAR CASES OF PROPOSED ESTIMATOR t_{PF}^*

It can easily be seen that

for $\delta = 1$, $t_{PF}^* = \bar{y}_s$; (2)

for $\delta = 2$, $t_{PF}^* = \bar{y}_s \left[2 - \frac{\bar{X}}{\bar{x}_s}\right] = t_{RP}$ (3)

for $\delta = 3$, $t_{PF}^* = t_{PAPE}$ and (4)

for $\delta = 4$, $t_{PF}^* = t_P$ (5)

Further as $\delta \to \infty$ $t_{PF}^* = \bar{y}_s$

i.e, t_{PF}^* converges to usual mean estimator $\bar{y}_s$.

4 SUPER POPULATION MODEL

Let us consider the superpopulation model under the assumption that $h(x_k)$ is a polynomial of order J. That is;

$$h(x_k) = \delta_0 \beta_0 + \delta_1 \beta_1 x_k + \delta_2 \beta_2 x_k^2 + \delta_J \beta_J x_k^J \qquad (6)$$

where $\delta_j = (j = 0,1,2, \dots, J)$ is zero or one according as the term x_k^J is absent or present respectively in the model. Since the expected value and variance of y_k depend on x_k and are denoted by $h(x_k)$ and $\sigma^2 . V(x_k)$ respectively, we can write;

$$y_K = \sum_{j=0}^{J} \delta_j . \beta_j . x_k^j + \epsilon_k [V(x_k)]^{\frac{1}{2}}; \ k = 1, 2, \ldots N \qquad (7)$$

where $\epsilon_1, \epsilon_2, \dots, \epsilon_k$ are independent random variable each having mean zero variance σ^2 , clearly,then

$$E_\xi (Y_k) = h(x_k) = \sum_{j=0}^{J} \delta_j \ B_j \ x_k^j \qquad (8)$$

$$V(y_k) = V(x_k). E(\epsilon_k^2) = \sigma^2 . V(x_k) \qquad (9)$$

$$and \ \ Cov(y_r, y_k) = 0 \quad (r \neq k)$$

5 BIAS AND MSE OF t_{PF}^* UNDER SUPER POPULATION MODEL

ξ- bias and ξ- MSE of t_{PF}^* under the general polynomial regression model $\xi[\delta_0, \delta_1, \delta_2 \dots, \delta_j : V(x)]$ are as follows:

$$B(t_{PF}^*) = \sum_{j=0}^{J} \delta_j \ \beta_j \ \{ \Psi_P^* (\delta, \bar{X}, \bar{x}_s) \cdot \bar{x}_s^{(j)} - \bar{X}^{(j)} \} \qquad (10)$$

And

$$M(t_{PF}^*) = \left[\sum_{j=0}^{J} \delta_j \ \beta_j \ \Psi_P^* \{(\delta, \bar{X}, \bar{x}_s) \cdot \bar{x}_s^{(j)} - \bar{X}^{(j)} \} \right]^2 + \sigma^2 \left[\frac{1}{n} \cdot \Psi_P^* (\delta, \bar{X}, \bar{x}_s) - \frac{1}{N} \right]^2 \sum_s V(x_k) + \frac{\sigma^2}{N^2} \sum_{\bar{s}} V(x_k) \qquad (11)$$

where $\Psi_P^* (\delta, \bar{X}, \bar{x}_s) = \frac{\Psi_P^* \{\phi(\delta)\}}{\Psi_P^* \{\phi_1(\delta)\}}$

Now the bias and MSE of particular cases of t_{PF}^* are given by

i) For $\delta = 1$,

$B(\, t_{PF}^{*}\,) = B(\bar{y}_s)$; $M(\, t_{PF}^{*}\,) = M(\bar{y}_s)$ (12)

ii) For $\delta = 2$,

$B\,(t_{RP}) = \sum_{j=0}^{J} \delta_j\, \beta_j\, \left\{\left(2 - \frac{\bar{x}_s}{\bar{X}}\right)\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$ (13)

$M(t_{RP}\,) = \left[\sum_{j=0}^{J} \delta_j\, \beta_j\, \left\{\left(2 - \frac{\bar{x}_s}{\bar{X}}\right)\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}\right]^2 + \sigma^2 \left[\frac{1}{n}\left(2 - \frac{\bar{x}_s}{\bar{X}}\right) - \frac{1}{N}\right]^2 \sum_s\, V\,(x_k) + \frac{\sigma^2}{N^2} \sum_{\bar{s}}\, V\,(x_k).$ (14)

iii) For $\delta = 3$,

$B(t_{PAPE}\,) = \sum_{j=0}^{J} \delta_j\, \beta_j \left\{\frac{(N-2n)\bar{X} + n\bar{x}_s}{(N\bar{X} - n\bar{x}_s)}\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$ (15)

$M(t_{PAPE}\,) = \left[\sum_{j=0}^{J} \delta_j\, \beta_j \left\{\frac{(N-2n)\bar{x}_s + n\bar{X}}{(N\bar{X} - n\bar{x}_s)}\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}\right]^2$

$+\, \sigma^2 \left[\frac{(N-2n)\bar{x}_s + n\bar{X}}{n(N\bar{X} - n\bar{x}_s)} - \frac{1}{N}\right]^2 \sum_s\, V\,(x_k) + \frac{\sigma^2}{N^2} \sum_{\bar{s}}\, V\,(x_k)$ (16)

iv) For $\delta = 4$,

$B(t_P\,) = \sum_{j=0}^{J} \delta_J\, \beta_J\, \left\{\frac{\bar{x}_s}{\bar{X}}\,.\;\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$ (17)

$\mathrm{M}(t_P\,) = \left[\sum_{j=0}^{J} \delta_J\, \beta_J \left\{\frac{\bar{x}_s}{\bar{X}}\,.\;\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}\right]^2 + \sigma^2 \left[\frac{\bar{x}_s}{n\bar{X}} - \frac{1}{N}\right]^2 \sum_s\, V(x_k) + \frac{\sigma^2}{N^2} \sum_{\bar{s}}\, V(x_k)$ (18)

5.1 COMPARISON OF $\boldsymbol{t_{PAPE}}$ WITH $\boldsymbol{t_P}$

The biases of t_P and t_{PAPE} are :

$B(t_P\,) = \sum_{j=0}^{J} \delta_J\, \beta_J\, \left\{\frac{\bar{x}_s}{\bar{X}}\,.\;\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$ (18)

$B(t_{PAPE}\,) = \sum_{j=0}^{J} \delta_j\, \beta_j \left\{\frac{(N-2n)\bar{X}_s + n\bar{x}}{N\bar{X} - n\bar{x}_s}\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$ (19)

Now for any j, $\delta_J\, \beta_J\, \left\{\frac{\bar{x}_s}{\bar{X}}\,.\;\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\} < \delta_J\, \beta_J \left\{\frac{(N-2n)\,\bar{X}_s + n\bar{x}}{N\bar{X} - n\bar{x}_s}\, \bar{x}_s^{(j)} - \bar{X}^{(j)}\right\}$,

Since $\frac{\bar{x}_s}{\bar{X}} < \frac{(N-2n)\bar{X}_s + n\bar{x}}{N\bar{X} - n\bar{x}_s}$. Therefore, we have $|B(t_P)| - |B\,(t_{TAPE}\,)| < 0$ (20)

So from (16) and (18) we have

$M(t_{PAPE}\,)$ - $\mathrm{M}(t_P\,) = [B(t_{PAPE}\,)]^2 - [B(t_P\,)]^2$

$$+\sigma^2 \left[\frac{1}{n^2}\left\{\frac{\bar{x}_s}{\bar{X}} + \frac{(N\,-\,2n)\,\bar{x}_s\,+\,n\bar{X}}{N\bar{X}\,-\,n\bar{x}_s}\right.\right.$$

$$\left.\left. -\frac{2n}{N}\right\}\left\{\frac{\bar{x}_s}{\bar{X}} - \frac{(N\,-\,2n)\,\bar{x}_s\,+\,n\bar{X}}{N\bar{X}\,-\,n\bar{x}_s}\right\}\right] \sum_s V\,(x_k)$$

(21)

Clearly the difference of squares of biases in (21) is always positive. Further, since $\frac{\bar{x}_s}{\bar{X}} < \frac{(N\,-\,2n)\,\bar{x}_s\,+\,n\bar{X}}{N\bar{X}\,-\,n\bar{x}_s}$ the sign of variance term in (21) depends upon the sign of $\left\{\frac{\bar{x}_s}{\bar{X}} + \frac{(N\,-\,2n)\,\bar{x}_s\,+\,n\bar{X}}{N\bar{X}\,-\,n\bar{x}_s} - \frac{2n}{N}\right\}$. If this term is positive then the variance term is negative. The term $\left\{\frac{\bar{x}_s}{\bar{X}} + \frac{(N\,-\,2n)\,\bar{x}_s\,+\,n\bar{X}}{N\bar{X}\,-\,n\bar{x}_s} - \frac{2n}{N}\right\}$ is positive only when $\frac{\bar{x}_s}{\bar{X}} + \frac{\bar{X}}{\bar{x}_s} < 2\left\{\frac{(N-n)^2}{nN} + 1\right\}$. Thus if the variance term in (21) is negative and exceeds $[B(t_{PAPE}\,)]^2 - [B(t_P\,)]^2$ in magnitude, the estimator t_{PAPE} will be more efficient than t_P under the model $\xi\left[\delta_0, \delta_1,\, \delta_2 \ldots, \delta_j : V\,(x)\right]$.

6 EFFICIENCY COMPARISON OF $\boldsymbol{t_{PAPE}}$ WITH $\boldsymbol{t_P}$

As Srivastava (1983) pointed out that under the prediction approach advocated by Basu (1971) the estimator t_{PAPE} is obtained instead of the usual product estimator t_P. However, to the first order of approximation ,that is , assuming sample size n to be large enough , the MSE's of both the estimators under fixed population approach are equal. In present paper both the estimators are particular cases of proposed estimator. Since the MSE and bias of these estimators are obtained under some superpopulation models, it is easy to compare empirically their efficiencies. A theoretical comparison of their MSE's has already been presented under section V .

Table I and table II present the efficiency of t_{PAPE} as compared with the

ISSN 2229-5518

efficiency of t_P under the assumed models when the sample mean $\bar{x}_s$ is smaller and greater than the population mean $\bar{x}_s$ is smaller and greater than the population mean $\bar{x}$. The conclusion given by Srivastava (1983) seems to be verified under the model approach too, since the values in the table are very close to 100. It can, therefore, be concluded that even if the customary product estimator does not have an intuitive basis as set by the predictive approach , it can be used frequently in place of t_{PAPE} because it is relatively easy to define and compute.

7 ROBUSTNESS OF THE PROPOSED ESTIMATOR

In order to examine the effect of misspecification of polynomial regression function and variance function on the MSE of estimator t_{PF} for different value of δ, we have considered the following superpopulation models:

Model I: $\xi\ [0,1 : x^9]$

Model II: $\xi\ [1,1 : x^9]$

Model III: $\xi\ [1,1,1 : x^9]$

Model IV: $\xi\ [1,1,1,1 : x^9]$

where $g = 0,1,2$

the constant g is usually unknown in practice and hence, the variance function should be taken as $V(x) = x^9$ but Cochran (1953) and Brewer (1963) have shown that majority of situations occurring in practice might be covered by assuming that $0 \le g \le 2$. In light of these results, we have taken $g = 0,1$ and 2 for our discussion.

A real population is considered to show the effect of model deviation on the efficiency of t_{PF} . The characteristic under study is weekly food cost of a family while the auxiliary character is weekly income in the population under consideration. The data has been taken from Singh (1987). Following Singh (1987) we have assumed that $\beta_0 = 4, \beta_1 = 2, \beta_2 = -1.5, \beta_3 = 2.4$ and $\sigma^2 = 0.92$. Thus, the models under consideration are:

Model I: $y_k = 2x_k + \epsilon_k x_k^{g/2}$

Model II: $y_k = 4 + 2x_k + \epsilon_k x_k^{g/2}$

Model III: $y_k = 4 + 2x_k - 1.5x_k^2 + \epsilon_k x_k^{g/2}$

Model IV: $y_k = 4 + 2x_k - 1.5x_k^2 + 2.4\, x_k^3 \epsilon_k x_k^{g/2}$

TABLE 1: MSE OF FOR t_{PF} FOR δ= 1 , 2 , 3 , 4 UNDER MODEL I ,II, III, IV when $\bar{x}_s > \bar{X}$

δ	g	Models			
		I	II	III	IV
1	0	779.0	779.0	1.007×10^7	3.478×10^{11}
	1	789.3	7893	1.007×10^7	3.478×10^{11}
	2	1661	1661	$1,661\times 10^7$	3.478×10^{11}
2	0	3116	3188	2.461×10^7	7.070×10^{11}
	1	3130	3202	2.461×10^7	7.070×10^{11}
	2	4332	6001	2.461×10^7	7.070×10^{11}
3	0	3928	4029	2.915×10^7	8.145×10^{11}
	1	3943	4045	2.915×10^7	8.145×10^{11}
	2	5234	5336	2.916×10^7	8.145×10^{11}
4	0	339.8	339.8	3.778×10^6	1.139×10^{11}
	1	3478	3478	3.778×10^6	1.139×10^{11}
	2	8859	8859	3.778×10^6	1.139×10^{11}

TABLE 2: MSE OF FOR t_{PF} FOR δ= 1 , 2 , 3 , 4 UNDER MODEL I ,II, III, IV when $\bar{x}_s < \bar{X}$

δ	g	Models			
		I	II	III	IV
1	0	338.8	338.8	3.778×10^6	1.139×10^{11}
	1	347.9	347.9	3.778×10^6	1.139×10^{11}
	2	885.9	885.9	3.778×10^6	1.139×10^{11}
2	0	1359	1402	7.905×10^6	1.840×10^{11}
	1	1365	1408	7.905×10^6	1.840×10^{11}
	2	1770	1813	7.905×10^6	1.840×10^{11}
3	0	1161	1195	7.185×10^6	1.724×10^{11}
	1	1168	1201	7.185×10^6	1.724×10^{11}
	2	1591	1625	7.186×10^6	1.724×10^{11}

ISSN 2229-5518

4	0	1192	1227	7.297×10^6	1.742×10^{11}
	1	1198	1233	7.297×10^6	1.742×10^{11}
	2	1618	1654	7.298×10^6	1.742×10^{11}

Looking at the tables 1 and 2, it can be said that t_{PF}^* has smallest MSE for $\partial = 1$ at which it is $\bar{y}_s$. When the sample mean $\bar{x}_s$ exceeds the population mean $\bar{X}$, t_{RP} is a better choice, next to $\bar{y}_s$ over t_{PAPE} and t_p but when $\bar{x}_s < \bar{X}$, t_{PAPE} is preferable over t_p and t_{RP}. However, for practical purposes, the MSE's of t_p and t_{PAPE} can be treated to be almost equal. This is so because the estimator t_{PAPE} converses to t_p as the population size increases to infinity.

As for as the robustness of t_{PF}^* is concerned, it can be seen that the change in MSE of the estimator under models I and II for g = 0 and 1is almost negligible. Similarly for g = 2, the estimator is again robust under tses models. The estimator t_{PF}^* is not robust under the model III and IV but the MSE of the estimator for a fixed value of δ under these models is not at all affected by the variation in variance function.

8 CONCLUDING REMARKS

In the present work a factor type class of estimator is developed which generates both the estimators t_P and t_{PAPE}. The estimators t_P and t_{PAPE} is obtained through predictive approach and is virtually a product estimator. When compared under the variations in superpopulation models, it is observed that the properties of both the estimators are more or less same.

REFERENCES

[1] Brewer, K. R. W.. "Ratio estimation andfinite populations; Some results deducible from the assumption of an underlying stochastic process", *Australian journal of Statistics* 5: 93-105,1963.

[2] Cochran, W.G.. "Sampling techniques", *John Willy and Sons*, Inc. New York, I edition, 1953..

[3] Ray, S. K. and Sahai, A. , "A note on the ratio and product type estimators", *Annals of the institute of Mathematical Statistics*, 31:141-144,1979.

[4] Reddy, V. N., "A note on the efficiency of ratio estimator under a superpopulation model", *Sankhya*, 37c: 95-100,1975.

[5] Royall, R. M. and Herson, J., "Robust estimation in finite population I", *Journal of American Statistical Association* 68: 880-889.1973a.

[6] Royall, R. M. and Herson, J., "Robust estimation in finite population II", *Journal of American Statistical Association* 68: 890-893.1973b.

[7] Sahai, A. and Ray, S. K., "An efficient estimators using auxiliary information", *Metrica* 27: 271 – 275, 1980.

[8] Shukla, D., "On some Factor type estimator for population parameters in Sample Surveys", Unpublished Ph.D. thesis, Banaras Hindu University, Varanasi", India, 1988.