

Efficient Generalizations of Product Estimator

M. K. Pande

Department of Mathematics, Kumaun University, Nanital (Uttaranchal)

(Received February 11, 2003)

Abstract. : A variant of ratio estimator was proposed by Sisodia and Dwivedi¹ by utilising Coefficient of variation of auxiliary variable. Singh and Upadhyaya² discussed the dual of this estimator as a variant of product estimator. The Coefficient of variation was assumed to be known. In the present paper efficient generalization of product type estimators are proposed by utilising the principal of limited invariance. Proposed estimators are compared with Conventional estimators. The comparisons are computer-aided empirical studies, with Monte-Carlo simulation sampling from hypothetical bivariate normal population.

1. Introduction

The usual ratio and product type estimators, remain invariant with change of scale and /or origin. The literature in sample surveys, is full of their efficient generalizations, using information on auxiliary variable. There exist classes of estimators which are non-invariant with respect to scale and / or origin. Estimators possessing the property of invariance are more popular. They turn out to be more suitable in terms of robustness. However, a class of invariant statistics (estimators) is more restricted (smaller) than a class of statistics (estimators) provided by the idea of limited invariance. The residence of a limited invariant statistics is inside a range of statistics, bounded by two invariant statistics. A limited invariant statistics itself is non-invariant with respect to scale / and or translation. It has one or more design parameter(s). The usual product estimator is $\bar{y} \bar{x} / \bar{X} = t_2$ say. Let $\bar{x}$ and $\bar{y}$ ($=t_1$) be the sample means for auxiliary variable X and main variable Y respectively. Here $\bar{X}$ is the population mean of X . A variant of t_2 can be $\bar{y} (\bar{x} + a) / (\bar{X} + a) = t_3$ say, where 'a' is non-stochastic design parameter t_3 is not scale invariant but its two limits t_1 ($a \rightarrow \infty$) and t_2 ($a \rightarrow 0$) are scale-invariant. As such t_3 is limited scale invariant estimator. Subsequently, we make our point by showing via a computer simulation study, that estimator t_3 and variants, are quite possibly more efficient, than the invariant estimators t_1 and t_2 , though they lack the property of invariance.

The computer study becomes relevant, as exact algebraic expression for Mean Square Error (MSE) of t_3 is not available. If first order $o(n^{-1})$ large sample

Key words : Coefficient of variation, Empirical probability, Mean Square Error (MSE) Simulation

approximation to MSE of t_2 and t_3 is taken, the comparison is algebraically intricate, because it depends upon many unknown population parameters. It then becomes difficult to conclude, as to which estimator is efficient and when. When the sample is that large, so as to justify the first order approximation, the regression estimator is preferable to ratio and product estimator. As such when sample is fairly large, and performance of regression estimator being rather unpredictable, we are motivated to use of estimators t_2 or of type t_3 . Then we have to consider second order $o(n^{-2})$ large sample approximation MSE of these estimators. This turns out to be algebraically more intricate and comparison is just impossible. The only course left is then the computer simulation study. It may be pointed out that asymptotically all estimators are dominated by the regression estimator, which is however more difficult to calculate. It is not easy to decide how far the relative advantage of regression estimator can compensate the difficulty of calculating it in surveys with complicated designs.

Let X, Y denote respectively the main and auxiliary variable with $\bar{X}$ and $\bar{Y}$ denoting their population means. The simple random sample of the size n is taken without replacement and assumed to be drawn from bivariate population which could be regarded, very nearly as normal. Let R be the correlation coefficient between two variables and $C(Y)$ and $C(X)$ be their respective coefficient of variation. Here, auxiliary information is used at the estimation stage. The ratio estimator $(\bar{y}/\bar{x}) \cdot \bar{X}$ and product estimator $(\bar{y} \cdot \bar{x})/\bar{X}$ due to Murthy³ both utilize the auxiliary information. Their use is recommended when $C (= RC(Y)/C(X))$ is greater (less) than 0.5 (-0.5) respectively. Both estimators are scale-invariant. Many efficient invariant ratio-cum-product type estimators, are available in literature, which have been shown to be performing better than sample mean, ratio or product estimators. Significant contribution in this direction have been presented among others by Srivenkataramana⁴, Srivenkataramana and Tracy⁵ and Bandyopadhyay⁶.

Searles⁷ utilized the knowledge of coefficient of variation in estimation. Sisodia and Dwivedi¹ mentioned that coefficient of variation of auxiliary variable is either easily obtainable or is often known. They thus proposed a ratio type estimator $(\bar{y}) \cdot (\bar{X} + C(X)) / (\bar{x} + C(X))$, while inserting the known value of $C(X)$. Their comparisons were based on the first order large sample approximations and were inconclusive. When R is negative above estimator is not useful. In such situations, Singh and Upadhyay² proposed a product type estimator viz. $t_4 = \bar{y}(\bar{x} + C(X)) / (\bar{X} + C(X))$. In the general set up, the estimator $t_3 = \bar{y}(\bar{x} + a) / (\bar{X} + a)$ is proposed, where "a" is nonstochastic design parameter. The limited invariant estimator turns out to be more efficient than its parent estimator $\bar{y} \cdot \bar{x} / \bar{X}$, though lacking the property of invariance. Investigations might reveal that we could gain by considering the class of limited invariant estimators rather than that of invariant estimators. The possibility of using linear combinations of t_1 , t_2 and t_3 may be resorted to study the performance of proposed estimators with respect to conventional estimators.

Hence, in this set up we propose the following two more non-invariant estimators

$$t_3 = at_1 + (1 - a)t_2$$

$$t_4 = (1 - a)t_1 + at_3$$

Subsequently, we make our point by showing via computers simulation study that the proposed estimators are quite possibly more efficient than the invariant estimators t_1 and t_2 ; though they lack the property of invariance.

2. The Estimator

The following estimators are taken up in the simulation study

- (2.1) Sample mean estimator : $\bar{y} = t_1$
 (2.2) Product estimator : $\bar{y}\bar{x}/\bar{X} = t_2$
 (2.3) $\bar{y}(\bar{x} + a) / (\bar{X} + a) = t_3$
 (2.4) $at_1 + (1 - a)t_3 = t_4$
 (2.5) $(1 - a)t_1 + at_3 = t_5$

Clearly, except t_1 and t_2 , other estimators do not belong to the class of invariant estimators. To the first degree of estimation, the MSE of estimator t_3 is given by :

$$\text{MSE}(t_3) = V_1 \{1 + A_1(A_1 + 2C)C^2(X)\}$$

where $V_1 = C^2(Y)/n$ and $A_1 = \bar{X}/(\bar{X} + a)$.

Clearly t_3 attains minimum value for $A_1 = -C$. We can divide each observation corresponding to auxiliary variable $\bar{X}$ (which is assumed to be known) and carry out the study, with new set of observations. Hence, $\bar{X} = 1$, without loss of generality. Minimisation of $\text{MSE}(t_3)$ with respect to A_1 is same as minimising $\text{MSE}(t_3)$ with respect to 'a' as $da/dA_1 \neq 0$. $\text{MSE}(t_3)$ attains the minimum value for $a = -(1 + \frac{1}{C})$. Replacing C by its guessed value, say 'g', we obtain guessed value of 'a' as $a = -(1 + \frac{1}{g})$ which is used in place of 'a' in expression (2.3) of the proposed estimator. Here, $(g - C)/C$ is the relative error in guessing C denoted by $\text{REG}(C)$. Hence, for t_3 guessed value $a = -(1 + \frac{1}{g})$.

Proceeding similarly, for Guessed value

$$t_4 \quad a = -\frac{(1 + g)}{(1 - g)},$$

$$t_5 \quad a = -\frac{g}{(1 + g)}.$$

The guessed value thus obtained, corresponding to different estimators are used in place of 'a' in expressions (2.3), (2.4) and (2.5).

3 The Simulation Study

The comparison is done for four example levels of $\bar{Y}$ (1.0, 2.0, 3.0, 4.0); two of $\sigma(X)$ (1.0, 2.0), two of $\sigma(Y)$ (2.0, 4.0). Here, $\bar{X} = 1$ is fixed. Simultaneously, five values of R (-0.2, -0.4, -0.8, -0.9) and five of $\text{REG}(C)$ (-0.2, -0.1, 0.0, 0.1, 0.2) are

considered. Hence, we obtain 400 parametric value-combinations, for the bivariate population. We generate, following Box Muller⁸, 100 random samples of size 10, 20 and 50 respectively. First the MSE of estimators are obtained and hence their relative efficiencies (RE'S) over replications of 100 samples are calculated. There are 400×3 (sample sizes : 10, 20 50) occasions when estimators are compared. Let $Ep(.)$ be the empirical probability for an estimator to be the winner. This is found by noting the the number of times a particular estimator attains the maximum efficiency. In the following Table 3.1 we present the results.

Table 3.1 { $Ep(.)$ for t_1, t_2, t_3 }

Estimator	t_1	t_2	t_3
REG (C) - 0.2	0.0499	0.2710	0.6791
- 0.1	0.0730	0.1812	0.7458
0.0	0.0417	0.1791	0.7792
0.1	0.0278	0.1639	0.8083
0.2	0.0427	0.1031	0.8542

From the Table, it is quite apparent that the proposed limited invariant estimator t_3 is probably, more efficient than t_1 and t_2 . It is particularly better when C over-guessed rather than when it is under-guessed.

Next, similar simulation study is carried out for the set of estimators (t_1, t_2, t_4) and (t_1, t_2, t_5). The combined results are presented, in the following Table 3.2 to indicate the performance of t_4 and t_5 vis-a-vis t_1 and t_2 .

Table 3.2 { $Ep(.)$ for (t_1, t_2, t_4), (t_1, t_2, t_5)}

Estimator	t_1	t_2	t_4	t_1	t_2	t_5
REG (C) - 0.2	0.0967	0.3321	0.5712	0.0687	0.2801	0.6512
- 0.1	0.0536	0.3541	0.5923	0.0368	0.2512	0.7120
0.0	0.1771	0.3016	0.5213	0.0219	0.1761	0.8020
0.1	0.2779	0.2909	0.4312	0.0687	0.2201	0.7112
0.2	0.1308	0.2879	0.5813	0.0318	0.2306	0.8012

The estimator t_4 attains highest $Ep(.)$ (viz. 0.5923) when $REG(C) = -0.1$. The proposed estimator is more efficient than t_1 and t_2 over entire range of $REG(C)$ which is considered t_2 comes relatively nearer to t_4 for $REG(C) = 0.1$. The underguess is more favourable than overguess of C . The Table also depicts that t_5 generally performs better, as compared to t_1 and t_2 . From the table it is evident that $Ep(.)$ for t_5 is higher than that of t_1, t_2 . Highest $Ep(.)$ is attained when there is no error in guessing C . The overguess appears to be more favourable than underguess for proposed estimator to be efficient.

Next, we undertake comparisons of all proposed non-invariant estimators for cases when $C < 0.0$. We consider six sub-ranges, namely : Range 1 : $C < -1.0$, Range 2 : $-1.0 \leq C < -0.8$, Range 3 : $-0.8 \leq C < -0.6$, Range 4 : $-0.6 \leq C < -0.4$, Range 5 : $-0.4 \leq C < -0.2$, Range 6 : $-0.2 \leq C < -0.0$. Relative frequency of winning is considered for all proposed non-invariant estimators, along with the sample mean and product estimator. The result for the case $C < -1.0$ is depicted in the Table 3.3 below.

Table 3.3 { $Ep(.)$ for $(t_1, t_2, t_3, t_4, t_5)$ }
 $C < -1.0$

Estimator	t_1	t_2	t_3	t_4	t_5
REG (C) -0.2	0.0	0.90305	0.0	0.0964	0.0
-0.1	0.0	0.9111	0.0444	0.0	0.0444
0.0	0.0	0.8740	0.0	0.0518	0.0740
0.1	0.0	0.7935	0.1269	0.0	0.0793
0.2	0.0	0.6582	0.0915	0.0582	0.1915

From the above Table 3.3 for the range $C < -1.0$ it is apparent that the conventional estimator t_2 is the most probable winner. We prepare similar tables for other cases. It is found that in the case of Range $-1 \leq C < -0.8$, the estimator t_5 is the best when C is highly overguessed. For the next Range $-0.8 \leq C < -0.6$ the most probable winner is the estimator t_5 . The product estimator t_2 is almost written off in the competition. In the range $-0.6 \leq C < -0.4$, the estimator t_3 turns out to be the most probable winner, irrespective of the value of $REG(C)$. Here again the estimator t_2 is almost completely written off. Considering the Range $-0.4 \leq C < -0.2$ the estimator t_3 turns out to be the most probable winner. It is worth noting that the conventional estimator t_2 is completely written off, whereas the usual unbiased estimator surfaces up in the competition very nicely. For the Range $-0.2 < C \leq 0$, it is the estimator t_3 which is most probably unbeaten. It is observed that the usual unbiased estimator comes up quite nicely in the competition, particularly when C is underguessed.

For $C < 0$, considering the Table 3.4 the look reveals that the estimator t_3 is the most probable winner consistently for all example values $REG(C)$

Table 3.4 {Ep (.) for (t_1, t_2, t_3, t_4, t_5) }
 $C < -1.0$

Estimator	t_1	t_2	t_3	t_4	t_5
REG (C) - 0.2	0.1296	0.2754	0.2837	0.1712	0.1337
- 0.1	0.1024	0.2119	0.2691	0.1565	0.1482
0.0	0.1033	0.2283	0.2626	0.2158	0.1908
0.1	0.0891	0.2058	0.3224	0.2016	0.1818
0.2	0.2920	0.1825	0.3033	0.4116	0.2200

References

1. B. V. S. Sisodia and V. K. Dwivedi, A modified ratio estimator using coefficient of variation of auxiliary variable, *Jour. Ind. Soc. Agri. Stat.*, **33(2)** (1981) 13-18.
2. H. P. Singh and L. N. Upadhyaya, A dual to modified ratio estimator using coefficient of variation of auxiliary variable, *Proc. Nat. Acad. Sci. India*, **56(a)IV**, (1986) 336-340.
3. M. N. Murthy, Product method of estimation, *Sankhya (A)*, **26** (1964) 69-74.
4. T. Srivenkataramana, A dual to ratio estimator in sample surveys, *Biometrika*, **67** (1980) 199-204.
5. T. Srivenkataramana and D. S. T. Tracy, Extending product method of estimation to positive correlation cases in surveys, *Aust. J. Stat.*, **23(1)** (1981) 95-100.
6. S. Bandyopadhyay, On the efficiency of product method of estimation, *Sankhya*, **B56** (1994) 222-227.
7. D. T. Searles, The utilization of a known coefficient of variation in the estimation procedure, *J. Amer. Stat. Assoc.*, **59** (1964) 1225-1226.
8. G. E. P. Box and M. E. Muller, A note on the generation of normal deviates, *Ann. Math. Statist.*, (1958) 610-611.