A Comparative Study of Generalized Ratio and Regression Estimators with their classical counterparts

Younathan Abdia
Department of Statistics, Forman Christian College, Lahore, Pakistan

Muhammad Qaiser Shahbaz
Department of Statistics, GC University, Lahore, Pakistan

Abstract

A comparative study has been made by using generalized ratio and regression estimators of Brewer, Horvitz and Thompson and Cassel-Sarandal and Wretman estimators. The study also involves ratio and regression estimators along with the mean per unit estimator of equal probability sampling. Ranking is being done to see which population total variances are performing the best.

Key Words: Ratio estimator, Regression estimator, Generalized Ratio estimator, Generalized Regression estimator, Rank of an estimator.

1. Introduction

Estimation of certain population characteristics on the basis of sample information has been a challenging task for survey statisticians. The aim of sample selection is to obtain fairly precise results about population parameters on the basis of sample. The simplest estimator of population total available is that of simple random sampling when no additional information is used. It is often be the case that an auxiliary variable X closely related to the main variable of study Y is also available. In these situations one of the most customary method of estimation is classical ratio method. The estimator of population total in this method is given as:

$$\hat{Y}_R = \hat{R}\,X = \frac{y}{x}X = \frac{\bar{y}}{\bar{x}}X \,. \tag{1.1}$$

An approximate variance for (1.1) is given as

$$V\left(\hat{Y}_R\right) = \frac{N^2\left(1-f\right)}{n}\left\{S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y\right\}. \tag{1.2}$$

The estimator given in (1.1) is best linear unbiased estimator (BLUE) under linear stochastic model

$$\left.\begin{array}{ll} y_i = \beta x_i + \varepsilon_i & , E\left(\varepsilon_i\right) = 0 \\ E\left(\varepsilon_i, \varepsilon_j\right) = & \begin{cases} \sigma_i^2, i = j \\ 0 \;\; , i \neq j \end{cases} \\ \sigma_i^2 = \sigma^2 z_i^{2\gamma} & , \frac{1}{2} \leq \gamma \leq 1 \end{array}\right\} \tag{1.3}$$

Another estimator that utilizes additional information is regression estimator. The regression estimator for population total is given as

$$\hat{Y}_{lr} = N\left(\bar{y} + b\left(\bar{X} - \bar{x}\right)\right). \qquad (1.4)$$

The approximate variance of (1.4) is given as:

$$Var\left(\hat{Y}_{lr}\right) = N^2 \frac{1-f}{n}\left(s_y^2 + b_0^2 s_x^2 - 2b_o s_{xy}\right). \qquad (1.5)$$

Various generalizations of ratio and regression estimators in unequal probability sampling is given by Brewer (1963), Brewer (1975) and Cassel, Sarandal and Wretman (1976).

2. Estimators Compared.

We have compared following methods of estimation in this study:

1. Mean per Unit Method:

   Estimator: $\hat{Y} = N\,\bar{y}$ (2.1)

   Variance: $Var\left(\hat{Y}\right) = N^2 \frac{1-f}{n} S_y^2.$ (2.2)

2. Classical Ratio Method:

   Estimator: $\hat{Y}_R = \frac{y}{x} X = \frac{\bar{y}}{\bar{x}} X.$ (2.3)

   Variance: $Var(\hat{Y}_R) = N^2 \frac{1-f}{n}\left(S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y\right).$ (2.4)

3. Classical Regression Method:

   Estimator: $\hat{Y}_{lr} = N\left(\bar{y} + b\left(\bar{X} - \bar{x}\right)\right)$ (2.5)

   Variance: $Var\left(\hat{Y}_{lr}\right) = N^2 \frac{1-f}{n}\left(s_y^2 + b_o^2 s_x^2 - 2 b_o s_{xy}\right)$ (2.6)

4. Hansen – Hurwitz (1943) Method:

   Estimator: $Y'_{HH} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{p_i}$ (2.7)

   Variance: $Var\left(Y'_{HH}\right) = \frac{1}{n}\sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - Y\right)^2.$ (2.8)

5. Horvitz – Thompson (1952) Method:

   Estimator: $Y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}.$ (2.9)

   Variance: $Var\left(Y'_{HT}\right) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j>i}}^{N} \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2.$ (2.10)

6. Brewer (1963, 1975) Ratio Methods:

   Estimator 1: $Y'_{HTR} = \frac{\sum_{i \in s} \frac{y_i}{\pi_i}}{\sum_{i \in s} \frac{x_i}{\pi_i}}$ (2.11)

MSE 1: $MSE(Y'_{HTR}) = Var(Y'_{HT}) + \frac{Y^2}{X^2} Var(X'_{HT}) - 2\frac{Y}{X} Cov(Y'_{HT}, X'_{HT})$ (2.12)

Estimator 2: $Y'_B = \sum_{i\in s} y_i + \frac{\hat{y}_{HT} - y}{\hat{x}_{HT} - x} \sum_{i\notin s} x_i$ (2.13)

7. Generalized Regression (1976) Estimator.
   Estimator: $y'_{CSW} = \sum_{i\in S} \left(Y_i - \hat{\beta}_{CSW} X_i\right)\pi_i^{-1} + \hat{\beta}_{CSW} X$ (2.14)

   Where $\hat{\beta}_{CSW} = \frac{\sum_{i\in S} Y_i X_i \pi_i^{-2}}{\sum_{i\in S} X_i^2 \pi_i^{-2}}$ (2.15)

Recently Shahbaz, et. al. (2002) has conducted similar sort of study in unequal probability sampling.

3. Empirical Study

The populations used for the empirical study are taken from the district census report of Lahore and Gujaranwala for year 1998.

Following variables have been used in the empirical study:

- Population of male who are primary but below matric (X)
- Population of male who are matric and above (Y)
- Total population (Z).

Forty populations have been studied and variances of the estimators stated above have been calculated for each population. After that ranking has been done in order to convert the monotonic relationship to a linear one so that better picture can be visualized. After calculating the ranks the average ranks of various estimators has been obtained for various ranges of coefficient of variation of Z and correlation coefficient between X and Z.

In the following tables we have used following abbreviation:

- Simple random sampling (SRS)
- Hansen Hurwitz estimator (HH)
- Horvitz Thompson estimator (HT)
- Classical ratio (CR)
- Classical Regression (CReg)
- Horvitz Thompson Ratio estimator (HTR)
- Brewer Ratio estimator (BR)
- Cassel Sarndal and Wretman estimator (CSW)

Table 1: Frequency of Ranks

Ranks	SRS	HH	HT	CR	CReg	HTR	BR	CSW
1	0	0	5	0	24	4	0	7
2	2	4	2	13	1	8	0	10
3	2	0	8	5	4	9	0	12
4	1	8	3	8	2	12	0	6
5	1	5	17	2	9	2	3	1
6	1	15	2	12	0	3	5	2
7	3	5	3	0	0	2	25	2
8	30	3	0	0	0	0	7	0
Average	7.15	5.35	4.08	3.88	2.28	3.43	6.90	2.95

Table 2: Average ranks of various estimators of various ranges of Coefficient of variation.

Ranks	SRS	HH	HT	CR	CReg	HTR	BR	CSW
1-10	6.0	5.0	3.8	3.4	2.0	4.4	7.3	4.1
11-20	7.4	5.7	4.4	3.1	1.8	3.6	6.9	3.1
21-30	7.4	5.9	4.6	4.1	1.8	2.9	6.6	2.7
31-40	7.8	4.8	3.5	4.9	3.5	2.8	6.8	1.9

Table 3: Average ranks of various estimators of various ranges of Correlation coefficient.

Ranks	SRS	HH	HT	CR	CReg	HTR	BR	CSW
1-10	6.2	5.7	4.6	3.2	2.0	4.1	6.8	3.4
11-20	7.0	5.3	3.9	3.7	1.8	3.6	7.0	3.7
21-30	7.4	5.7	4.4	4.1	2.0	3.1	6.8	2.5
31-40	8.0	4.7	3.4	4.5	3.3	2.9	7.0	2.2

Table 4: Regression Summaries for Ranks of Various Estimators for Model $Rank(Estim) = \beta_0 + \beta_1 [Rank(CV)] + \beta_2 [Rank(\rho)] + \varepsilon$

	Estimator							
	SRS	**HH**	**HT**	**CR**	**CReg**	**HTR**	**BR**	**CSW**
β_0	5.867	5.784	4.637	2.818	1.344	4.415	6.986	4.149
p-Value	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000
β_1	0.007	0.016	0.023	0.078	0.057	-0.059	-0.047	-0.077
p-Value	0.842	0.649	0.531	0.026	0.125	0.080	0.005	0.020
β_2	0.055	-0.038	-0.051	-0.026	-0.011	0.010	0.042	0.018
p-Value	0.143	0.298	0.178	0.440	0.755	0.751	0.010	0.568
F	3.506	1.027	1.185	4.010	2.283	3.128	4.599	5.203
p-Value	0.040	0.410	0.317	0.027	0.116	0.056	0.016	0.010

4. Conclusions

Table 1 shows that classical regression estimator outperform all other estimators and is followed by generalized regression estimator. Whereas the performance of simple random sampling is worst of all of the estimators and other estimators performed moderately. In twenty four populations the variance of classical regression estimator variance is minimum. That is in sixty percent of the populations classical regression estimator is the best estimator. As the average of the ranks also shows that the classical regression estimator has the lowest average followed by generalized regression estimator, whereas the simple random sampling estimator has the maximum average ranks. From all this it can be seen that classical regression estimator performs best among all other estimators.

Table 2 shows the average ranks of various estimators for various ranges of coefficient of variation for variable Z. From this table it can be seen that classical regression estimator performs best when the coefficient of variation for the auxiliary variable Z is least or moderate. When the coefficient of variation is high then generalized regression estimator performs well.

Further, table 3 gives the average ranks of various estimators for various ranges of correlation coefficient between X and Z. This also show that when the correlation between X and Z is low or moderate classical regression estimator perform well and when the correlation coefficient is high then the generalized regression estimator performs best.

Table 4 gives the regression summaries for rank of various estimators for model. The coefficient β_0 gives the average rank for the estimator when the effect of independent variables is zero. If the effect of the independent variables is zero then classical regression estimator performs best followed by classical regression estimator, whereas Brewer ratio estimator performs worst. The coefficient β_1 shows the partial effect of coefficient of variation on the average rank of an estimator. This coefficient is positive for most of the estimators except the Horvitz – Thompson ratio, Brewer ratio and Cassel-Sarndal-Wretmena estimator. This indicates that the average rank of these estimators will decrease with an increase in the coefficient of variation. From this we can say that these estimators should be used for populations having larger coefficient of variation. The coefficient β_2 shows the partial effect of correlation coefficient on the average rank of an estimation method. If the correlation coefficient is high then Horvitz-Thompson estimator performs well and if the correlation coefficient is low then mean per unit estimator are good. Overall we can see that the regression model is significant.

References

1. Brewer, K. R. W., (1963), *A model of systematic sampling with unequal probabilities*, Aust. J. Stat. 5, 5 – 13.

2. Brewer, K. R. W., (1975), *A simple procedure for sampling πpswor*, Aust. J. Stat. 17(3), 166 – 172.
3. Cassel, C. M., Sarndal, C. E. and Wretman, J. H., (1976), *Some results on generalized difference estimation and generalized regression estimation for finite populations*, Biometrika, 63, 615–620.
4. Hansen, M. H. and Hurwitz, W. N., (1943), *On the theory of sampling from a finite population*, Ann. Math. Stat. 14, 333 – 362.
5. Horvitz, D. G. and Thompson, D. J., (1952), *A generalization of sampling without replacement from a finite universe*, J. Amer. Stat. Assoc. 47, 663 – 685.
6. Shahbaz, M. Q., Hanif, M. and Samiuddin, M., (2002), *On some estimators in unequal probability sampling*, J. Statist. Theory and Appl. Vol. 1 (4), 239 – 252.