Ratio Estimator for Multiple Regressors

Naila Alam*

Abstract

The ratio estimator is a simple and direct approach of the estimation of population total when a correlated variable with study variable is available. The Classical Ratio Estimator (CRE) for one regressor has been extensively used in the early 18th century and can be extended easily for multiple regressors but unfortunately this topic has not attain much attention and we find lack of related research in literature. Now a day's bulk of information is available for estimation of population parameters to increase the efficiency of the estimator. In this paper a simple form of ratio estimator has proposed that allowed more than one regressors in estimation of population total.

Key Words: Classical ratio estimator, regressors, population total.

This article can be cited as:

Alam, N., (2015). Ratio Estimator for Multiple Regressors, Journal of Arts and Social Sciences. 2 (1). 16-20[1]

Introduction:

The classical ratio estimator was used by Laplace in 1820 for estimating the population of France by using ratio of births during the preceding year. After that Brewer (1963) and Royall (1970) discussed the ratio estimator by using a super-population model. When the relationship between the study variable and auxiliary variable is linear and the regression line passes through the origin, also the variance of the study variable is proportional to the variance of the regressors, then ratio estimators are used to find out the estimate of population total.

*Naila Alam Assistant Prof, Institute of Education and Research, Lahore College for Women University Lahore.
naila.alam1969@gmail.com

Let a finite population $U \;=\{u_1,....u_N\}$, consists of N elements. The $X=\sum_N x_k$ denotes the sum of the N 'observations of regressor variable and similarly $Y=\sum_N y_k$ is the population total of the study variable. Also the sample total of regressor and study variable are respectively $\sum_n x_k$ and $\sum_n y_k$, then ratio of the sample totals will be approximately the same as the ratio of the population total that is

$$\frac{Y}{X} \cong \frac{\sum_n x_k}{\sum_n y_k}.$$

Therefore we can estimate the unknown population total of the study variable by using the law of three that is if AB=CD and D is unknown then D can be found out by using $D = {AB}/{C}$. Similarly, the population total of study variable is,

$$Y \cong \frac{\sum_n x_k}{\sum_n y_k} X .$$

The regression model of y on x is defined as,

$$y_k = b x_k + x_k^{\gamma} e_{ok}$$

If, $w_i = x_i^{-2\gamma}$, then $\gamma = {1}/{2}$ is the basis for the classical ratio estimator.

A simple form of classical ratio estimator for population total is described as,

$$y_r = \frac{\sum_s y_k}{\sum_s x_k} X \quad = bX \tag{1.1}$$

where,

$$b = \frac{\sum_s y_k}{\sum_s x_k} \tag{1.2}$$

Some other chain ratio type estimators are discussed in literature that minimizes the MSE of the ratio estimator for an appropriate value of α that is,

$$Y = \sum_s y_k \left[\frac{\sum_U x_k}{\sum_s x_k} \right]^{\alpha}$$

The CRE proves itself a robust estimator in both model-based and design based methods. It is a biased estimator but bias becomes zero when $N = n$. The sum of residuals of CRE given in (3.4.1) is zero. That is,

$$= \sum_{k=1}^{n} (\mathrm{y}_k - \frac{(\sum_{k=1}^{n} y_k}{\sum_{k=1}^{n} x_k} x_k)$$

$$= \sum_{k=1}^{n} \mathrm{y}_k - \frac{(\sum_{k=1}^{n} y_k}{\sum_{k=1}^{n} x_k} \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} \mathrm{y}_k - \sum_{k=1}^{n} \mathrm{y}_k$$

=0

The MSE of the ratio estimator is defined as,

$$MSE(\overline{y}_r) = \mathrm{E}(\overline{y}_r - \overline{Y})^2$$

$$= V(\overline{y}_r) + [Bias(\overline{y}_r)]^2$$

$$= \frac{1-f}{n}[\{S_y^2 - 2\rho b S_y S_x + b^2 S_x^2\} + \{\frac{S_y S_x}{\overline{X}}(1-\rho)\}^2].$$

The bias of $\overline{y}_r$ approaches to zero when $\rho \to 0$. For $\rho = 1$ (perfect positive correlation) the bias of ratio estimator will be zero.

2- A Ratio Estimator for Multiple Regressors:

The ratio estimator given in (1.1) can be extended easily for multiple regressors. Now a day's bulk of information is available for estimation process to increase the efficiency of the estimator, therefore a functional form of ratio estimator has proposed when multiple regressors are available prior to estimation.

Let we have p regressors,

$X_1, X_2, \ldots\ldots X_p$ related to the study variable 'y with known population totals, and a vector of 'p regressors $\mathbf{x}'_k = \left[x_{k1} \;\; x_{k2} \;\; \ldots \;\; x_{kp} \right]$ is associated with y_k, where $j = 1, 2, \ldots\ldots \mathrm{p}$.

Then a new form of the ratio estimator for multiple regressors is defined as,

$$Y_{rc} = \frac{\sum_{n} y_k}{\sum_{n} x_{kc}} X_c = \frac{\hat{Y}}{\hat{X}_c} X_c \qquad (2.1)$$

Where,

$$X_c = w_1 X_1 + w_2 X_2 + \ldots\ldots.. + w_p X_p \text{ and,} \qquad (2.2)$$

$$\hat{X}_c = w_1 \hat{X}_1 + w_2 \hat{X}_2 + \ldots\ldots.. + w_p \hat{X}_p \qquad (2.3)$$

Also, $\hat{X}_j = \sum_{n} x_{jk}$ and $X_j = \sum_{N} x_{jk}$ for $j = 1, 2, \ldots.. p$

The $w' = \left[w_1, w_2, \ldots\ldots.., w_p \right]$ is the vector of weights obtained as,

$$w_1 = \frac{\sum_{n} y_k}{\sum_{n} x_{1k}}$$

$$w_2 = \frac{\sum_{n} y_k}{\sum_{n} x_{2k}}$$

.

.

.

$$w_p = \frac{\sum_n y_k}{\sum_n x_{pk}}$$

The estimator defined in (2.1) allows more than one regression variables to use in the estimation of total via ratio estimator.

The MSE of the ratio estimator defined in (2.1) is described as,

$$MSE(y_{rc}) = [\{S_y^2 - 2\rho b S_y S_{xc} + b^2 S_{xc}^2\} + \{\frac{S_y S_{xc}}{\bar{X}}(1-\rho_c)\}^2].$$

Where ρ_c is the correlation between X_c and 'y ,defined as,

$$\rho_c = \frac{\sum_{k=1}^{N}(y_k - \bar{y})(x_{kc} - \bar{x}_c)}{\sqrt{\sum_{k=1}^{N}(y_k - \bar{y})^2 (x_{kc} - \bar{x}_c)^2}}$$

The estimator (2.1) will be an unbiased estimator of the population total if

ρ_c =1. That is perfect positive correlation between 'y and X_c.

References:

1. Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2004). Weighting sample data subject to independent controls. *Survey Methodology*, 30, 35-44.
2. Knaub, J.R., Jr. (1997), "Weighting in Regression for Use in Survey Methodology. *InterStat, April 1997, http://interstat.statjournals.net/. (Note*

shorter, more recent version in ASA Survey Research Methods Section proceedings, 1997.)

3. Knaub, J.R., Jr. (2003), "Applied Multiple Regression for Surveys with Regressors of Changing Relevance: Fuel Switching by Electric Power Producer. *InterStat, May 2003, http://interstat.statjournals.net/. (Note another version in ASA Survey Research Methods Section proceeding*s, 2003.)

4. Krapavickaitė, D. and Plikusas, A. (2005). Estimation of a ratio in the finite population. *Informatica*, 16, 347-364.

5. Park,S. and Kim, J.K. (2014). Instrumental-Variable calibration estimation in survey sampling. *Statistica Sinica*, 24, 1001-1015.

6. Rao, J.N.K. (1994). Estimating totals and distribution function using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

7. Särndal, C.E., Swensson, B. and Wretmen, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator. *Biometrika*, 76, 527-637.