The Chain Ratio Estimator and Regression Estimator with Linear Combination of Two Auxiliary Variables

Jingli Lu*
College of Sciences, Inner Mongolia University of Technology, Hohhot, China

Abstract

In sample surveys, it is usual to make use of auxiliary information to increase the precision of the estimators. We propose a new chain ratio estimator and regression estimator of a finite population mean using linear combination of two auxiliary variables and obtain the mean squared error (MSE) equations for the proposed estimators. We find theoretical conditions that make proposed estimators more efficient than the traditional multivariate ratio estimator and the regression estimator using information of two auxiliary variables.

Citation: Lu J (2013) The Chain Ratio Estimator and Regression Estimator with Linear Combination of Two Auxiliary Variables. PLoS ONE 8(11): e81085. doi:10.1371/journal.pone.0081085

Editor: Raya Khanin, Memorial Sloan Kettering Cancer Center, United States of America

Received July 26, 2013; Accepted October 9, 2013; Published November 18, 2013

Copyright: © 2013 Jingli Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors has no support or funding to report.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: lujingli2004@163.com

Introduction

The use of supplementary information provided by auxiliary variables in survey sampling was extensively discussed [1–10]. The ratio estimator and regression estimator are among the most commonly adopted estimators of the population mean or total of study variable of a finite population with the help of two auxiliary variables when the correlation coefficient between the two variables is positive. It is well known that these estimators are more efficient than the usual estimator of the population mean based on the sample mean of a simple random sampling.

In this study, we proposed a new chain ratio estimator and regression estimator using linear combination of two auxiliary variates, and obtain the mean squared error (MSE) equations for the two proposed estimators. The proposed estimators, the traditional multivariate ratio estimator and the regression estimator using information of two auxiliary variables were compared at theoretical conditions. And we obtained the satisfactory results.

Materials and Methods

The existed estimators

The classical ratio estimator and regression estimator for the population mean $\bar{Y}$ of the variate of interest $y$ using one auxiliary information are defined by

$$\bar{y}_r = \frac{\bar{Y} \cdot \bar{x}}{\bar{x}}$$

(1)

$$\bar{y}_{reg} = \bar{y} + b(\bar{Y} - \bar{x})$$

(2)

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ are the sample mean of the auxiliary variate $x$ and the variate of interest $y$ respectively. Here

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

(3)

where $n$ is the number of units in the sample[11], and $b = \frac{Syx}{Sx^2}$ is the regression coefficient for of $Y$ on $X$. $S_y^2$ and $S_x^2$ are the population variances of the $y_i$ and $x_i$ respectively. $Syx$ is the population covariance between $y_i$ and $x_i$ [11].

The MSE of the classical ratio estimator is

$$MSE(\bar{y}_r) \approx \frac{1-f}{n} (S_y^2 + R^2 S_x^2 + 2RS_{yx})$$

(4)

where $f = \frac{n}{N}$; $N$ is the number of units in the population; $R = \frac{\bar{Y}}{\bar{x}}$ is the population ratio, $\bar{Y}$ and $\bar{x}$ are the population means of the $y_i$ and $x_i$ respectively.

The MSE of the regression estimator is

$$MSE(\bar{y}_{reg}) = \frac{1-f}{n} S_y^2 (1 - \rho_{yx}^2)$$

(5)

where $\rho_{yx} = \frac{Syx}{S_y S_x}$ is the population correlation coefficient between $y_i$ and $x_i$.

Kadilar and Cingi[12] proposed the chain ratio estimator using one auxiliary information for $\bar{Y}$ as

$$\bar{y}_{cr} = \bar{y} + z(\bar{Y} - \bar{x})$$

(6)

where $z$ is real number.
MSE of this estimator is given as follows:

\[
\text{MSE}(\hat{y}_{i1}) \leq \frac{1-f}{n} \left( S_t^2 + \hat{\rho}_1^2 R_1^2 S_{x_1}^2 + \hat{\rho}_2^2 R_2^2 S_{x_2}^2 \right. \\
\left. - 2 \hat{\rho}_1 R_1 S_{xy_1} - 2 \hat{\rho}_2 R_2 S_{xy_2} + 2 \hat{\rho}_1 \hat{\rho}_2 R_1 R_2 S_{xy_{12}} \right)
\]  

(10)

where \( \hat{\rho}_{XY} \) denote respectively the sample and the population means of the variable \( X_i \), \( i=1,2 \); \( b_1 = \frac{S_{xy_1}}{S_{x_1}} \) and \( b_2 = \frac{S_{xy_2}}{S_{x_2}} \) are the regression coefficients of on \( X_1 \) and \( X_2 \), respectively, here \( S_{x_1}^2 \) and \( S_{x_2}^2 \) are the variances of \( X_1 \) and \( X_2 \), respectively, and \( S_{xy_1} \) and \( S_{xy_2} \) are the covariances between \( Y \) and \( X_1 \), \( Y \) and \( X_2 \), respectively, \( \hat{\rho}_1 \) and \( \hat{\rho}_2 \) are the weights that satisfy the condition respectively: \( \hat{\rho}_1 + \hat{\rho}_2 = 1 \) and \( \hat{\rho}_1 \hat{\rho}_2 = 1 \).

The MSE of this traditional multivariate ratio estimator is given by

\[
\text{MSE}(\bar{y}_{MR}) \leq \frac{1-f}{n} \left( S_t^2 + \hat{\rho}_1^2 R_1^2 S_{x_1}^2 + \hat{\rho}_2^2 R_2^2 S_{x_2}^2 \right. \\
\left. - 2 \hat{\rho}_1 R_1 S_{xy_1} - 2 \hat{\rho}_2 R_2 S_{xy_2} + 2 \hat{\rho}_1 \hat{\rho}_2 R_1 R_2 S_{xy_{12}} \right)
\]  

(11)

The minimum MSE of \( \bar{y}_{MR} \) can be shown to be:

\[
\text{MSE}_{\text{min}}(\bar{y}_{MR}) = \frac{1-f}{n} \left( S_t^2 + \hat{\rho}_1^2 R_1^2 S_{x_1}^2 + \hat{\rho}_2^2 R_2^2 S_{x_2}^2 \right. \\
\left. + \hat{\rho}_1 \hat{\rho}_2 R_1 R_2 S_{xy_{12}} \right)
\]  

(12)

The optimum values of \( \hat{\rho}_1 \) and \( \hat{\rho}_2 \) are given by

\[
\hat{\rho}_1^* = \frac{b_1^2 S_{x_1}^2 + b_1 S_{xy_1} - b_2 S_{xy_2} - b_2 b_1 S_{xy_{12}}}{b_1^2 S_{x_1}^2 - 2 b_1 b_2 S_{xy_{12}} + b_2^2 S_{x_2}^2} \quad \hat{\rho}_2^* = 1 - \hat{\rho}_1^*
\]  

The minimum MSE of \( \bar{y}_{Mreg} \) can be shown to be:

\[
\text{MSE}_{\text{min}}(\bar{y}_{Mreg}) = \frac{1-f}{n} \left( S_t^2 + \hat{\rho}_1^2 R_1^2 S_{x_1}^2 + \hat{\rho}_2^2 R_2^2 S_{x_2}^2 \right. \\
\left. + \hat{\rho}_1 \hat{\rho}_2 R_1 R_2 S_{xy_{12}} \right)
\]  

(13)

The suggested estimators

We propose the multivariate chain ratio estimator and regression estimator using linear combination of two auxiliary variables as follows:

\[
\bar{y}_{a} = \frac{\omega_1 \bar{X}_1 + \omega_2 \bar{X}_2}{\omega_1 \bar{x}_1 + \omega_2 \bar{x}_2}
\]  

(14)

\[
\bar{y}_{k} = \bar{y} + b(\bar{x}_k - \bar{x}_k)
\]  

(15)

where \( \omega_1, \omega_2 \) and \( k_1, k_2 \) are weights that satisfy the condition: \( \omega_1 + \omega_2 = 1 \) and \( k_1 + k_2 = 1 \).

The MSE of this new multivariate ratio estimator is given by

\[
\text{MSE}(\bar{y}_{a}) \leq \frac{1-f}{n} \left( S_t^2 + \hat{\rho}_1^2 R_1^2 S_{x_1}^2 + \hat{\rho}_2^2 R_2^2 S_{x_2}^2 \right. \\
\left. + \hat{\rho}_1 \hat{\rho}_2 R_1 R_2 S_{xy_{12}} \right)
\]  

(16)

where \( \omega_1 \bar{X}_1 + \omega_2 \bar{X}_2 = R_k \)
The optimum values of \( \omega_1 \) and \( \omega_2 \) are given by

\[
\omega_1 = \frac{a^2_1 \, \bar{y} \, \bar{x} - \bar{y} \, \bar{x} \, \bar{y}}{s_{xy}^2 \bar{y}^2 + s_{yx}^2 \bar{x}^2 + s_{yy}^2 + s_{xx}^2}
\]

where

\[
\omega_2 = 1 - \omega_1
\]

The minimum MSE of \( \bar{Y}_{\text{slc}} \) can be shown to be:

\[
MSE_{\text{min}}(\bar{Y}_{\text{slc}}) = \frac{1}{n} \left( \frac{S_y^2}{S_y^2} \right)^{2}(1 - \rho_{yx})^2
\]

where

\[
S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2
\]

\[
\rho_{yx}^2 = \frac{S_{yx}^2}{S_y^2 S_x^2} = \frac{(k_1 S_{x_1} + k_2 S_{x_2})^2}{S_y^2(k_1^2 S_{x_1}^2 + 2k_1k_2 S_{x_1 x_2} + k_2^2 S_{x_2}^2)}
\]

The optimum values of \( k_1 \) and \( k_2 \) are given by

\[
k_1^* = \frac{S_{x_1}^2 S_{x_2}^2 - S_{x_1 x_2} S_{x_1 x_2}}{S_y^2 S_{x_1}^2 + S_{x_1}^2 S_{x_2}^2 - S_{x_1 x_2} S_{x_1 x_2}} \quad \text{and} \quad k_2^* = 1 - k_1^*
\]

The minimum MSE of \( \bar{Y}_{\text{slc}} \) can be shown to be:

\[
MSE_{\text{min}}(\bar{Y}_{\text{slc}}) = \frac{1}{n} \left( \frac{S_y^2}{S_y^2} \right)^{2}(1 - \rho_{yx})^2
\]

where

\[
\rho_{yx}^2 = \frac{(k_1^* S_{x_1} + k_2^* S_{x_2})^2}{S_y^2(k_1^2 S_{x_1}^2 + 2k_1^*k_2^* S_{x_1 x_2} + k_2^2 S_{x_2}^2)}
\]

Efficiency comparison

We compare the MSE of the proposed multivariate ratio estimator using information of two auxiliary variables given in Eq. (17) with the MSE of traditional multivariate ratio estimator using information of two auxiliary variables given in Eq.(11) as follows:

\[
MSE(\bar{Y}_{\text{slc}}) < MSE(\bar{Y}_{\text{MR}})
\]

Table 1. Data Statistics.

| \( N \) | 180 | \( \bar{Y}_2 \) | 143.31 | \( S_y^2 \) | 19465.38 | \( \rho_{yx} \) | 0.862 |
| \( n \) | 70 | \( \bar{Y} \) | 1093.1 | \( S_y^2 \) | 11912.61 | \( \rho_{yx} \) | 0.842 |
| \( \bar{Y}_1 \) | 181.57 | \( S_y^2 \) | 694885.7 | \( \rho_{yx} \) | 0.973 |

We compare the MSE of the proposed regression estimators given in Eq. (19) with the MSE of the traditional multivariate regression estimator using information of two auxiliary variables given in Eq.(13) as follows:

\[
MSE(\bar{Y}_{\text{slc}}) < MSE(\bar{Y}_{\text{Mreg}})
\]

Numerical illustration

The comparison among these estimators is given by using a data set whose statistics are given in Table 1[14]. We apply the traditional multivariate ratio estimator and regression estimator using information of two auxiliary variables, given in Eqs.(8) and (9) and proposed chain ratio estimator and regression estimator of a finite population mean using linear combination of two auxiliary

Table 2. MSE Values of Estimators.

| Estimators | MSE |
|------------|-----|
| \( \bar{Y}_{\text{MR}} \) | 0.1576 |
| \( \bar{Y}_{\text{slc}} \) | 0.1574 (\( \chi = 0.96 \)) |
| \( \bar{Y}_{\text{slc}} \) | 0.1766 |
| \( \bar{Y}_{\text{Mreg}} \) | 0.1574 |

DOI:10.1371/journal.pone.0081085.t001

DOI:10.1371/journal.pone.0081085.t002
variables, given in Eqs. (14) and (15), to data whose statistics are given in Table 1. We assume to take the sample size \( n = 70 \), from \( N = 100 \) using SRSWOR. The MSE of these estimators are computed as given in Eqs. (11), (13), (17) and (19).

**Results and Discussion**

MSE values of the traditional multivariate ratio estimator and regression estimator using information of two auxiliary variables and proposed chain ratio estimator and regression estimator using linear combination of two auxiliary variables can be seen in Table 2.

From Table 2, we notice that our proposed chain ratio estimator using linear combination of two auxiliary variables \( \hat{\tau}_{ch} (\alpha = 0.96) \) is more efficient than traditional multivariate ratio estimator using information of two auxiliary variables and our proposed regression estimator using linear combination of two auxiliary variables \( \hat{\tau}_{reg} \) is more efficient than traditional multivariate regression estimator using information of two auxiliary variables. We examine the conditions for this data set,

\[
S_{11}^{2}(\omega_{1}^{2}x_{1}^{2}R_{k1}^{2} - \varepsilon_{1}^{2}R_{11}^{2}) + S_{12}^{2}(\omega_{2}^{2}x_{1}^{2}R_{k2}^{2} - \varepsilon_{2}^{2}R_{12}^{2})
- 2S_{12}(\omega_{1}^{2}x_{1}^{2}R_{k1}^{2} - \varepsilon_{1}^{2}R_{12}^{2})R_{22} + 2S_{12}^{2}(\omega_{1}^{2}x_{1}^{2}R_{k1}^{2} - \varepsilon_{1}^{2}R_{12}^{2}) = -0.02563 < 0
\]

\[
\frac{(k_{1}^{1}S_{11} + k_{2}^{1}S_{12})^{2}}{(k_{1}^{2}S_{11}^{2} + 2k_{1}^{1}k_{2}^{2}S_{11}S_{12} + k_{2}^{2}S_{12}^{2})} = 16.1849 >
2w_{1}^{1}b_{1}S_{11} + 2w_{2}^{1}b_{2}S_{12} - w_{1}^{1}b_{1}^{2}S_{11}^{2}
- w_{2}^{1}b_{2}^{2}S_{12}^{2} = 13.9842
\]

The result shows that the condition (20) and condition (21) are satisfied. Therefore, we suggest that we should apply the proposed estimators to this data set.

**Conclusions**

We develop a new chain ratio estimator and a new regression estimator of a finite population mean using two auxiliary variables and theoretically show that the proposed estimators are more efficient than the traditional ratio estimator and traditional regression estimator using two auxiliary variables in certain condition.

**Author Contributions**

Conceived and designed the experiments: JL. Performed the experiments: JL. Analyzed the data: JL. Contributed reagents/materials/analysis tools: JL. Wrote the paper: JL.

**References**

1. Kadilar C, Cingi H (2005) A new estimator using two auxiliary variables. Appl Math Comput 162: 901–908.
2. Kadilar C, Candan M, Cingi H (2007) Ratio estimators using robust regression. Hacet J Math Stat 36: 181–188.
3. Kadilar C, Cingi H (2006) Improvement in estimating the population mean in simple random sampling. Appl Math Lett 19: 75–79.
4. Bacardi S, Kadilar C (2008) Ratio estimators with unequal probability designs. Pak J Statist 24: 167–172.
5. Gupta S, Shabbir J (2007) On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. J Stat Plan Infer 137: 1606–1611.
6. Al-Omari AI, Jemain AA, Ibrahim K (2009) New ratio estimators of the mean using simple random sampling and ranked set sampling methods. Revista Investigacion Operacional 30: 97–108.
7. Kadilar C, Unyazici Y, Cingi H (2009) Ratio estimator for the population mean using ranked set sampling. Stat Papers 50:301–309.
8. Upadhyaya LN, Singh HP (1999) Use of transformed auxiliary variable in estimating the finite population mean. Biometrical J 41:627–636.
9. Tailor R, Parmar R, Kim JM, Tailor R (2011) Ratio-cum-Product estimators of population mean using known population parameters of auxiliary variable. Commu Korean Stat Soc 18: 155–164.
10. Khare BB, Srivastava U, Kumar K (2013) A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey. J Sci Res 57: 147–153.
11. Cochran WG (1977) Sampling Techniques. New-York: John Wiley and Sons.
12. Kadilar C, Cingi H (2003) A study on the chain ratio-type estimator. Hacet J Math Stat 32: 105–108.
13. Feng SY, Ni JX, Zou GH (1998) The Theory and Methods of Sampling Survey. Beijing: China Statistics Press, 143–150p. (in Chinese)
14. Feng SY, Shi XQ (1996) The Sampling Survey—Theory, Method and Practice. Shanghai: Shanghai Scientific and Technical Publishers, 147–151p. (in Chinese)