Ratio-cum-product and dual to ratio-cum-product estimators

Ignas Bartkus[1]

[1] Vilnius Pedagogical University, Lithuania
e-mail: ignas.bartkus@gmail.com

Abstract

Ratio-cum-product and dual to ratio-cum-product estimators are compared by simulation for stratified sampling design. The definition of dual variable presented in Plikusas (2008) is used to define the dual variable for stratified sample design.

1 Introduction

The product ratio estimators using one dual auxiliary variable were introduced by Bandyopadhyay (1980) and Srivenkataramana (1980).

Ratio type estimators that uses two auxiliary variables are also considered in the literature. These estimators are defined for simple random sampling and can be effective when one auxiliary variable is positively correlated and the other variable is negatively correlated with the study variable. The ratio-cum-product estimator is presented in Singh (1969), and dual to ratio-cum-product estimator in Singh et all (2005). In the paper Singh et all (2005) the dual to ratio-cum-product estimator is considered for simple random sample without replacement.

The ratio estimators expressed as a weighted sum of some ratio and direct estimator are analyzed in Singh (2000).

2 Ratio-cum-product estimator

2.1 Simple random sample case

Consider a finite population $\mathcal{U} = (u_1, u_2, ..., u_N)$ of N units. Let a sample s of size n be drawn from this population by simple random sampling without replacements. Let y_k represents the value of a response variable y and two auxiliary variables x and z are available. The ratio and product estimators are

$$\hat{t}_R = \frac{\hat{t}_y}{\hat{t}_x}\, t_x, \quad \hat{t}_P = \frac{\hat{t}_y\, \hat{t}_z}{t_z},$$

where

$$\hat{t}_y = \frac{N}{n}\sum_{k\in s} y_k, \quad \hat{t}_x = \frac{N}{n}\sum_{k\in s} x_k, \quad \hat{t}_z = \frac{N}{n}\sum_{k\in s} z_k, \quad t_x = \sum_{k=1}^{N} x_k, \quad t_z = \sum_{k=1}^{N} z_k.$$

The ratio-cum-product estimator is defined as

$$\hat{t}_{RP} = \hat{t}_y \frac{\hat{t}_z}{\hat{t}_x}\, \frac{t_x}{t_z}.$$

The approximate variances of the estimators $\hat{t}_R$, $\hat{t}_P$, $\hat{t}_{RP}$ are

$$MSE(\hat{t}_R) = N^2 \frac{1-f}{n} \bar{Y}^2 \big(C_y^2 + C_x^2(1 - 2K_{yx})\big),$$

$$MSE(\hat{t}_P) = N^2 \frac{1-f}{n} \bar{Y}^2 \big(C_y^2 + C_z^2(1 + 2K_{yz})\big),$$

$$MSE(\hat{t}_{RP}) = N^2 \frac{1-f}{n} \bar{Y}^2 \big(C_y^2 + C_z^2(1 + 2K_{yz}) + C_x^2(1 - 2K_{yx} - 2K_{zx})\big),$$

where

$$f = \frac{n}{N}, \quad K_{yx} = \rho_{yx} C_y / C_x, \quad K_{zx} = \rho_{zx} C_z / C_x, \quad K_{yz} = \rho_{yz} C_y / C_z,$$

$$C_y = \frac{s_y}{\bar{Y}}, \quad s_y^2 = \frac{1}{N-1} \sum_{k=1}^{N} (y_k - \bar{Y})^2, \quad \bar{Y} = \frac{1}{N} \sum_{k=1}^{N} y_k,$$

$$\rho_{yx} = \frac{s_{yx}}{s_y s_x}, \quad s_{yx} = \frac{1}{N-1} \sum_{k=1}^{N} (y_k - \bar{Y})(x_k - \bar{X}),$$

C_x, C_z, ρ_{yz}, ρ_{zx}, $\bar{X}$, $\bar{Z}$ are defined analogously and respective to the subscripts used.

2.2 Stratified simple random sample case

Assume the population $\mathcal{U}$ consists of H strata: $\mathcal{U} = \mathcal{U}_1 \cup \ldots \cup \mathcal{U}_H$. The size of stratum $\mathcal{U}_h$ is N_h, and the size of simple random sample s_h in stratum $\mathcal{U}_h$ is n_h, $h = 1, \ldots, H$. The ratio and product estimators are

$$\hat{t}_{Rst} = \frac{\hat{t}_{yxt}}{\hat{t}_{xst}} t_x, \quad \hat{t}_{Pst} = \frac{\hat{t}_{yst} \, \hat{t}_{zst}}{t_z},$$

where

$$\hat{t}_{yst} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} y_k, \quad \hat{t}_{xst} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} x_k, \quad \hat{t}_{zst} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} z_k,$$

The ratio-cum-product estimator is defined as

$$\hat{t}_{RPst} = \hat{t}_{yst} \frac{\hat{t}_{zst}}{\hat{t}_{xst}} \frac{t_x}{t_z}.$$

And the approximate variances of these estimators are

$$MSE(\hat{t}_{Rst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(C_{yh}^2 + C_{xh}^2(1 - 2K_{yxh})\big),$$

$$MSE(\hat{t}_{Pst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(C_{yh}^2 + C_{zh}^2(1 + 2K_{zxh})\big),$$

$$MSE(\hat{t}_{RPst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(C_{yh}^2 + C_{zh}^2(1 + 2K_{yzh}) + C_{xh}^2(1 - 2K_{yxh} - 2K_{zxh})\big),$$

where

$$f = \frac{n_h}{N_h}, \quad K_{yxh} = \rho_{yxh} C_{yh} / C_{xh}, \quad K_{zxh} = \rho_{zxh} C_{zh} / C_{xh}, \quad K_{yzh} = \rho_{yzh} C_{yh} / C_{zh},$$

$$C_{yh} = \frac{s_{yh}}{\bar{Y}}, \quad s_{yh}^2 = \frac{1}{N_h - 1} \sum_{k \in U_h} (y_k - \bar{Y}_h)^2, \quad \bar{Y}_h = \frac{1}{N_h} \sum_{k \in U_h} y_k, \quad \bar{Y} = \frac{1}{N} \sum_{k=1}^{N} y_k,$$

$$\rho_{xyh} = \frac{s_{xyh}}{s_{xh} s_{yh}}, \quad s_{xyh} = \frac{1}{N_h - 1} \sum_{k \in U_h} (y_k - \bar{Y}_h)(x_k - \bar{X}_h),$$

Estimators of the approximate variances are

$$\widehat{MSE}(\hat{t}_{Rst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(\hat{C}_{yh}^2 + \hat{C}_{xh}^2 (1 - 2\hat{K}_{yxh})\big),$$

$$\widehat{MSE}(\hat{t}_{Pst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(\hat{C}_{yh}^2 + \hat{C}_{zh}^2 (1 + 2\hat{K}_{zxh})\big),$$

$$\widehat{MSE}(\hat{t}_{RPst}) = \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \bar{Y}^2 \big(\hat{C}_{yh}^2 + \hat{C}_{zh}^2 (1 + 2\hat{K}_{yzh}) + \hat{C}_{xh}^2 (1 - 2\hat{K}_{yxh} - 2\hat{K}_{zxh})\big),$$

3 Dual to ratio-cum-product estimator

3.1 Simple random sample case

The dual variable for the improvement of ratio-type estimator was used by Bandyopadhyay (1980) and Srivenkataramana (1980). An estimator with two auxiliary variables were considered by Singh et al (2005).

Consider again simple random sampling of size n and introduce the linear transformation of the variables x and z:

$$x_k^* = (1+g)\bar{X} - g x_k, \quad z_k^* = (1+g)\bar{Z} - g z_k, \quad \text{for} \quad k = 1, \ldots, N,$$

where $g = n/(N-n)$. Then

$$\bar{x}^* = (1+g)\bar{X} - g\bar{x}, \quad \bar{z}^* = (1+g)\bar{Z} - g\bar{z}$$

are unbiased estimators for $\bar{X}$ and $\bar{Z}$. It is easy to see that $t_x^* = \sum_{k=1}^{N} x_k^* = t_x$ and $t_z^* = \sum_{k=1}^{N} z_k^* = t_z$. Correlation coefficient between variables y and x^*, $Corr(y, x^*) = -Corr(y, x) = -\rho_{yx}$ and $Corr(y, z^*) = -\rho_{yz}$. The dual to ratio-cum-product estimator is

$$\hat{t}_{RP}^* = N\,\bar{y}\,\frac{\bar{x}^*}{\bar{X}}\,\frac{\bar{Z}}{\bar{z}^*}\,.$$

The approximate variance of $\hat{t}_{RP}^*$ is

$$MSE(\hat{t}_{RP}^*) = N^2 \frac{1-f}{n} \bar{Y}^2 \big(C_y^2 + g C_z^2 (g + 2K_{yz}) + g C_x^2 (g - 2gK_{zx} - 2K_{yx})\big),$$

It is shown in Singh et al (2005) that for $1 - g > 0$ ($N > 2n$), $\hat{t}_{RP}^*$ is more efficient than Horvitz-Thompson estimator $\hat{t}_y$ and ratio-cum-product estimator $\hat{t}_{RP}$ when

$$\frac{1}{2}\, g < \frac{K_{yx} c_x^2 - K_{yz} c_z^2}{c_x^2 + c_z^2 - 2K_{zx} c_x^2} < \frac{1}{2}\,(1+g).$$

3.2 Stratified simple random sample case

Consider stratified simple random sampling of size n_h in stratum $\mathcal{U}_h$, denote $g_h = n_h/(N_h - n_h)$ for $h = 1, \ldots, H$. The dual transformation for stratified and arbitrary sampling design is defined in Plikusas (2008).

Here we use the direct generalization of dual transformation, and define transformation of the auxiliary variable x:

$$x_k^* = (1 + g_h)\bar{X}_h - g_h x_k, \quad \text{for} \quad k \in \mathcal{U}_h,$$

where $\bar{X}_h = \frac{1}{N_h}\sum_{k\in\mathcal{U}_h} x_k$. The transformation for the variable z are defined analogously. Note that $\sum_{k=1}^N x_k^* = \sum_{k=1}^N x_k = t_x$. The relation $Corr(y, x^*) = -\rho_{xy}$ is not valid in the case of stratified sample.

The dual to ratio-cum-product estimator is defined as

$$\hat{t}^*_{RPst} = \hat{t}_{yst}\frac{\hat{t}^*_{xst}}{t_x}\frac{t_z}{\hat{t}^*_{zst}},$$

where

$$\hat{t}_{yst} = \sum_{h=1}^H \frac{N_h}{n_h}\sum_{k\in s_h} y_k, \text{ and } \hat{t}^*_{xst} = \sum_{h=1}^H \frac{N_h}{n_h}\sum_{k\in s_h} x_k^*.$$

The approximate variance of dual to ratio-cum-product estimator is

$$MSE(\hat{t}^*_{RPst}) = \sum_{h=1}^H N_h^2 \frac{1-f_h}{n_h}\bar{Y}^2\big(C^2_{yh} + g_h C^2_{zh}(g_h + 2K_{yzh}) + g_h C^2_{xh}(g_h - 2g_h K_{zxh} - 2K_{yxh})\big),$$

And the estimator of the approximate variance

$$\widehat{MSE}(\hat{t}^*_{RPst}) = \sum_{h=1}^H N_h^2 \frac{1-f_h}{n_h}\bar{Y}^2\big(\hat{C}^2_{yh} + g_h \hat{C}^2_{zh}(g_h + 2\hat{K}_{yzh}) + g_h \hat{C}^2_{xh}(g_h - 2g_h \hat{K}_{zxh} - 2\hat{K}_{yxh})\big).$$

4 Simulation study

In this section some empirical study is presented to observe the behavior of the estimators in the case of stratified simple random sample design. A real populations from some Lithuanian Enterprise survey were used for the simulation. During the simulation study several populations were examined. It was observed that for skewed populations, and big sampling fractions from strata, the product estimators are not efficient. They are beaten by simple ratio estimator, despite the high correlation with both variables.

It should be noted that both auxiliary variables initially are positively correlated with the study variable. So, first of all we transform the variable z to dual, and consider the transformed variable as given negatively correlated auxiliary variable. Below some results when product estimators performs efficiently are presented.

Population **I**

y - An income of enterprise, x - Number of employees, z - Number of employees from another source (dual variable).

$N = 150$, $n = 50$, $t_y = 30803297$, $t_x = 11875$, $t_z = 20433$, $C_y^2 = 0.7265$, $C_x^2 = 0.8539$, $C_z^2 = 0.0592$, $\rho_{yx} = 0.9106$, $\rho_{yz} = -0.8999$, $\rho_{zx} = -0.9437$.

Population **II**

y - An income of enterprise, x - Number of employees, z - Number of employees from another source (dual variable).

$N = 150$, $n = 50$, $t_y = 24268559$, $t_x = 8659$, $t_z = 13581$, $C_y^2 = 1.0071$, $C_x^2 = 0.8025$, $C_z^2 = 0.0489$, $\rho_{yx} = 0.9616$, $\rho_{yz} = -0.9243$, $\rho_{zx} = -0.9492$.

These populations are stratified into three strata by the size of the variable x. The sample size in each strata satisfies inequality $2n_h < N_h$ and 100 samples were drawn.

Table 1. **Simulation results for the Population I**

Estimator	Average estimate	Estimated bias	Average estimate of variance $\times 10^{11}$	Approximate variance $\times 10^{11}$	$\widehat{MSE}$ $\times 10^{11}$	CV
$\hat{t}_{yst}$	30816540.3	-13243.3	19.482	19.3805	19.382	0.0452
$\hat{t}_{Rst}$	31492033.1	688736.1	10.771	11.3235	15.515	0.0338
$\hat{t}_{RPst}$	30802062.5	-1234.5	15.908	11.5718	15.908	0.0349
$\hat{t}^*_{RPst}$	30829099.6	25802.5	15.634	10.6642	15.637	0.0335

Table 2. **Simulation results for the Population II**

Estimator	Average estimate	Estimated bias	Average estimate of variance $\times 10^{11}$	Approximate variance $\times 10^{11}$	$\widehat{MSE}$ $\times 10^{11}$	CV
$\hat{t}_{yst}$	24248625.6	19933.4	15.704	15.958	15.961	0.0521
$\hat{t}_{Rst}$	24208723.7	-59835.3	11.668	11.683	11.704	0.0446
$\hat{t}_{RPst}$	24208723.7	-59835.3	6.3473	5.6632	6.3831	0.0311
$\hat{t}^*_{RPst}$	24234630.1	-33928.9	5.7149	6.1563	5.7264	0.0324

Tables 1 and 2 show that for stratified simple random sampling design dual ratio-cum-product estimator can be more efficient than other estimators considered. For Population II (Table 2), dual ratio-cum-product estimator has a little bit bigger coefficient of variation then ratio-cum-product estimator.

References

Bandyopadhyay, S. (1980) Improved ratio and product estimators. *Sankhyā Series C*, **42**(2), 45-49.

Plikusas, A. (2008) Some overview of the ratio type estimators In: *Workshop on survey sampling theory and methodology*, Statistics Estonia.

Singh, M. P. (1969) Comparison of some ratio-cum-product estimators. *S*ankhyā Series B, **31**, 375-382.

Singh, H. P., Singh, R., Espejo, M. R., Pineda, M. D., Nadarajan, S. (2005) On the efficiency of the dual to ratio-cum-product estimator. *Mathematical Proceedings of the Royal Irish Academy*, **105A**(2), 51-56.

Srivenkataramana, T. (1980) A dual to ratio estimator in sample surveys. *Biometrica*, **67**, 199-204.

Upadhyaya, L. N., Singh, H. P. (1999) Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, **41**(5), 627-636.

Upadhyaya, L. N., Singh, H. P., Vos, J. V. E. (1985) On the estimation of population means and ratios using supplementary information. *Statistica Neerlandica*, **39**(3), 309-318.