Open Journal of Statistics, 2012, 2, 300-304
http://dx.doi.org/10.4236/ojs.2012.23036 Published Online July 2012 (http://www.SciRP.org/journal/ojs)

Scientific Research

Multivariate Ratio Estimator of the Population Total under Stratified Random Sampling

Oscar O. Ngesa[1], **George O. Orwa**[2], **Romanus O. Otieno**[2], **Henry M. Murray**[2]
[1]Ministry of State for Planning, National Development and V2030, Nairobi, Kenya
[2]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
Email: oscanges@yahoo.com

Received March 7, 2012; revised April 10, 2012; accepted April 30, 2012

ABSTRACT

Olkin [1] proposed a ratio estimator considering *p* auxiliary variables under simple random sampling. As is expected, Simple Random Sampling comes with relatively low levels of precision especially with regard to the fact that its variance is greatest amongst all the sampling schemes. We extend this to stratified random sampling and we consider a case where the strata have varying weights. We have proposed a Multivariate Ratio Estimator for the population mean in the presence of two auxiliary variables under Stratified Random Sampling with L strata. Based on an empirical study with simulations in *R* statistical software, the proposed estimator was found to have a smaller bias as compared to Olkin's estimator.

Keywords: Ratio Estimator; Stratification; Auxiliary Variables; Lagrange's Multiplier

1. Introduction

Auxiliary variables have been used to increase precision of estimators especially in regression and ratio estimators [2]. This is particularly so in cases of complex surveys, more so in situations where some information on the survey variable might be missing [3].

These classical methods of estimation are based on direct estimators, *i.e.*, those which use the response variable, y and information provided by an auxiliary variable, *x*, highly correlated with the main variable [4].

2. Review of Multivariate Ratio Estimators

Olkin [1] proposed a multivariate generalization of the ratio estimator. Olkin proposed an estimator for the population total, denoted by $\hat{Y}_{MR}$, and defined as

$$\hat{Y}_{MR} = W_1 \frac{\bar{y}}{\bar{x}_1} X_1 + W_2 \frac{\bar{y}}{\bar{x}_2} X_2 + \cdots + W_p \frac{\bar{y}}{\bar{x}_p} X_p \qquad (2.1)$$

which in other context can also be written as;

$$\hat{Y}_{MR} = W_1 \hat{Y}_{R_1} + W_2 \hat{Y}_{R_2} + \cdots + W_p \hat{Y}_{R_p} \qquad (2.2)$$

where $\hat{Y}_{R_i} = \frac{\bar{y}}{\bar{x}_i} X_i$ is the component of the population total ratio estimate affiliated to the i^{th} auxiliary variable W_i are the weights which maximize the precision of $\hat{Y}_{MR}$, subject to a linear constraint $W_1 + W_2 + \cdots + W_p = 1$. This estimate of population total also will be accurate if the regression line of Y on $X_1, X_2, \cdots, X_p$ is a straight line going through the origin. The population totals for the auxiliary variables X_i must be explicitly known.

3. The Proposed Estimator

Consider a population which has been divided into L strata, with the strata being disjoint, the sample elements from each stratum are sampled and when the measurement y_{hi} is done, measurement for the i^{th} unit in the h^{th} stratum, two auxiliary variables, say, x_{hi1} and x_{hi2} are also measured for that i^{th} unit. Let $\hat{Y}_{MRE}$ denote the proposed multivariable estimator under the stratified random sampling scheme for the population total. $\hat{Y}_{MRE}$ is therefore defined as;

$$\hat{Y}_{MRE} = \sum_{i=1}^{L} \hat{Y}_{MRi} \qquad (2.3)$$

where the individual components are defined as follows:

$\hat{Y}_{MR1} = W_{11}\hat{Y}_{R_{11}} + W_{12}\hat{Y}_{R_{12}}$ ··· for the 1st stratum.

$\hat{Y}_{MR2} = W_{21}\hat{Y}_{R_{21}} + W_{22}\hat{Y}_{R_{22}}$ ··· for the 2nd stratum.

$\hat{Y}_{MRL} = W_{L1}\hat{Y}_{R_{L1}} + W_{L2}\hat{Y}_{R_{L2}}$ ··· for L^{th} the stratum.

This can further be represented in a single equation as follows;

$$\hat{Y}_{MRh} = W_{h1}\hat{Y}_{R_{h1}} + W_{h2}\hat{Y}_{R_{h2}} \qquad (2.4)$$

where $h = 1, 2, \cdots, L$ are the various strata.

4. Variance of the Proposed Estimator

To compute the values of the weights, the general Equation (2.4) is used and this will cater for each stratum by just changing the value of *h* in respective strata. Subtracting Y_h to the right hand side and left hand side of equation (2.4) yields

$$\hat{Y}_{MRh} - Y_h = W_{h1}\hat{Y}_{R_{h1}} + W_{h2}\hat{Y}_{R_{h2}} - Y_h \qquad (2.5)$$

But it is known that the sum of the weights in each stratum is 1, so $W_{h1} + W_{h2} = 1$. This implies that

$$Y_h = \left(W_{h1} + W_{h2}\right)Y_h \qquad (2.6)$$

Replacing Equation (2.6) to the right hand side of Equation (2.5), yields

$$\hat{Y}_{MRh} - Y_h = W_{h1}\hat{Y}_{R_{h1}} + W_{h2}\hat{Y}_{R_{h2}} - \left(W_{h1} + W_{h2}\right)Y_h$$

$$\hat{Y}_{MRh} - Y_h = W_{h1}\hat{Y}_{R_{h1}} + W_{h2}\hat{Y}_{R_{h2}} - W_{h1}Y_h - W_{h2}Y_h$$

Collecting the like terms with respect to weights yields

$$\hat{Y}_{MRh} - Y_h = W_{h1}\left(\hat{Y}_{R_{h1}} - Y_h\right) + W_{h2}\left(\hat{Y}_{R_{h2}} - Y_h\right) \qquad (2.7)$$

Squaring each side and taking Expectation on either side, assuming negligible bias, Equation (2.7) leads to

$$V\left(\hat{Y}_{MRh}\right) = W^2{}_{h1}V\left(\hat{Y}_{R_{h1}}\right) + 2W_{h1}W_{h2}Cov\left(\hat{Y}_{R_{h1}}, \hat{Y}_{R_{h2}}\right) + W^2{}_{h2}V\left(\hat{Y}_{R_{h2}}\right) \qquad (2.8)$$

Equation (2.8) can be written in notation as follows,

$$V\left(\hat{Y}_{MRh}\right) = W^2{}_{h1}V_{h11} + 2W_{h1}W_{h2}V_{h12} + W^2{}_{h2}V_{h22} \qquad (2.9)$$

where

$V_{h11} = \text{Variance}\left(\hat{Y}_{R_{h1}}\right)$,

$V_{h22} = \text{Variance}\left(\hat{Y}_{R_{h2}}\right)$

and

$V_{h12} = \text{Covariance}\left(\hat{Y}_{R_{h1}}, \hat{Y}_{R_{h2}}\right)$

We then proceed to find the values of the weights W_{h1} and W_{h2} that minimize the variance $V\left(\hat{Y}_{MRh}\right)$ subject to the linear constraint $W_{h1} + W_{h2} = 1$.

To achieve this, we form a function Φ which has the variance and the linear constraint mentioned above.

$$\Phi = V\left(\hat{Y}_{MRh}\right) + \lambda\left(1 - W_{h1} - W_{h2}\right) \qquad (2.10)$$

with λ being the Lagrange's Multiplier.

From Equation (2.9),

$$V\left(\hat{Y}_{MRh}\right) = W_{h1}^2V_{h11} + 2W_{h1}W_{h2}V_{h12} + W_{h2}^2V_{h22}$$

replacing this into Equation (2.10) yields;

$$\Phi = W_{h1}^2V_{h11} + 2W_{h1}W_{h2}V_{h12} + W_{h2}^2V_{h22} + \lambda\left(1 - W_{h1} - W_{h2}\right)$$

To minimize this function with respect to the weights W_{h1} and W_{h2}, we differentiate partially the function Φ with respect to these weights each at a time.

$$\frac{\partial\Phi}{\partial W_{h1}} = 2W_{h1}V_{h11} + 2W_{h2}V_{h12} - \lambda \qquad (2.11)$$

$$\frac{\partial\Phi}{\partial W_{h2}} = 2W_{h1}V_{h12} + 2W_{h2}V_{h22} - \lambda \qquad (2.12)$$

For optimization, we equate the partial derivative Equations (2.11) and (2.12), each to zero. These yields;

$$\lambda = 2W_{h1}V_{h11} + 2W_{h2}V_{h12} \qquad (2.13)$$

$$\lambda = 2W_{h1}V_{h12} + 2W_{h2}V_{h22} \qquad (2.14)$$

It follows that Equations (2.13) and (2.14) are equal, then

$$2W_{h1}V_{h11} + 2W_{h2}V_{h12} = 2W_{h1}V_{h12} + 2W_{h2}V_{h22}$$

The 2 is common and can be cancelled out. We proceed to collect like terms with respect to the weights and this yield

$$W_{h1}\left(V_{h11} - V_{h12}\right) = W_{h2}\left(V_{h22} - V_{h12}\right) \qquad (2.15)$$

It is known that $W_{h1} + W_{h2} = 1$, hence $W_{h2} = 1 - W_{h1}$. From this Equation (2.15) will reduce to

$$W_{h1}\left(V_{h11} - V_{h12}\right) = \left(1 - W_{h1}\right)\left(V_{h22} - V_{h12}\right)$$

and

$$W_{h1}\left\{\left(V_{h11} - V_{h12}\right) + \left(V_{h22} - V_{h12}\right)\right\} = \left(V_{h22} - V_{h12}\right)$$

Then it follows, by making W_{h1} the subject of the formula,

$$W_{h1} = \frac{\left(V_{h22} - V_{h12}\right)}{\left\{\left(V_{h11} - V_{h12}\right) + \left(V_{h22} - V_{h12}\right)\right\}}$$

Opening the brackets in the denominator yields

$$W_{h1} = \frac{\left(V_{h22} - V_{h12}\right)}{\left(V_{h11} - 2V_{h12} + V_{h22}\right)} \qquad (2.16)$$

To get the value of weight W_{h2}, we use the linear constraint $W_{h2} = 1 - W_{h1}$

$$W_{h2} = 1 - \frac{\left(V_{h22} - V_{h12}\right)}{\left(V_{h11} - 2V_{h12} + V_{h22}\right)}$$

which may be written as,

$$W_{h2} = \frac{\left(V_{h11} - 2V_{h12} + V_{h22}\right)}{\left(V_{h11} - 2V_{h12} + V_{h22}\right)} - \frac{\left(V_{h22} - V_{h12}\right)}{\left(V_{h11} - 2V_{h12} + V_{h22}\right)}$$

$$W_{h2} = \frac{\left(V_{h11} - V_{h12}\right)}{\left(V_{h11} - 2V_{h12} + V_{h22}\right)} \qquad (2.17)$$

Equations (2.16) and (2.17) give the weights that minimize the variance $V\left(\hat{Y}_{MRh}\right)$ for stratum *h*.

These weights can now be substituted in the proposed model to get the population total.

5. Empirical Study

An empirical study was carried out to estimate the population total of a simulated population and compare the performance of the proposed model to that of Olkin [1].

6. Description of the Study Population

In this section we simulated a population (y_i, x_{1i} and x_{2i}), which has 10 strata $i = 1, 2, \cdots, 10$ in which each stratum differs from others. This difference was achieved by using different error terms e_i while generating the y_i using $y_i = a_i x_{1i} + b_i x_{2i} + e_i$. The coefficients a_i and b_i are randomly generated from a uniform distribution while y_i, x_{1i} and x_{2i} are randomly gene-rated from normal distribution with different parameters.

7. Computational Procedure

A sample of size 300 was selected randomly from the simulated population index-wise, that is if index *i* is selected then the sample elements will have y_i, x_{1i} and x_{2i}. This was repeated for all the ten strata, the selected sample was used in the proposed model to estimate the population total. The ten strata were again joined together to form one huge stratum, index-wise sample of size 1000, was selected and then using Olkin's model, the population total was estimated. The procedure above was repeated for 1000 samples and the population totals using each model was recorded.

8. Simulation Results

The population total estimates of the two methods were compared to that of the true population (simulated) total. The True population total is 28,235,645. **Table 1** summarizes the statistics corresponding to each estimator. **Figures 1** and **2** show the plotted values of the population total estimates of proposed model and Olkin's model, respectively, repeated for 1000 simulations each.

In order to show the difference in variability between the two methods, the two plots above are now combined into one graph using a common scale in the **Figure 3**.

9. Conclusions

From the summary table above, it can be seen that the proposed estimator gives a total with a very small bias as compared to the Olkin's. Also, the proposed model can be seen to have a small Root Mean Square Error (RMSE)

Table 1. Summary statistics for each method.

	Min.	Median	3rd Qrt	Max	Mean	Bias	RMSE
Proposed Method	2,821,006	2823185	2,823,565	2823987	2,825,123	2,823,579	144.53
Olkin's Method	2,746,765	2805085	2,822,892	2840866	2,903,358	2,822,799	7659.34

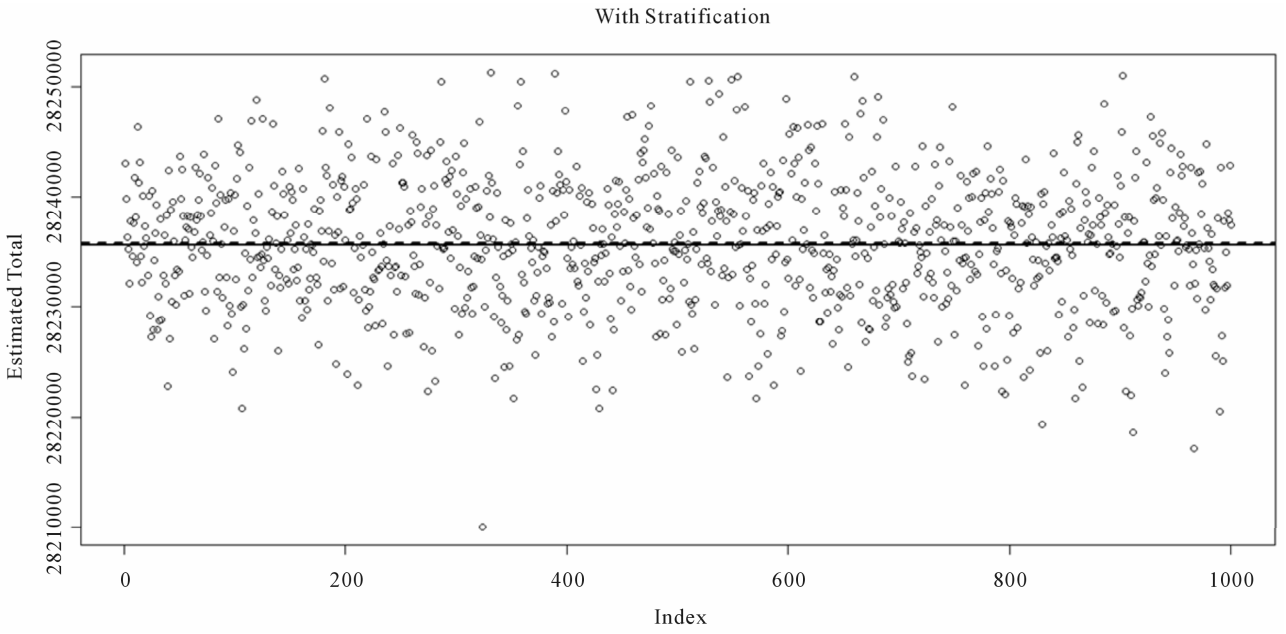

Figure 1. Plot of the population totals with proposed model for the 1000 samples.

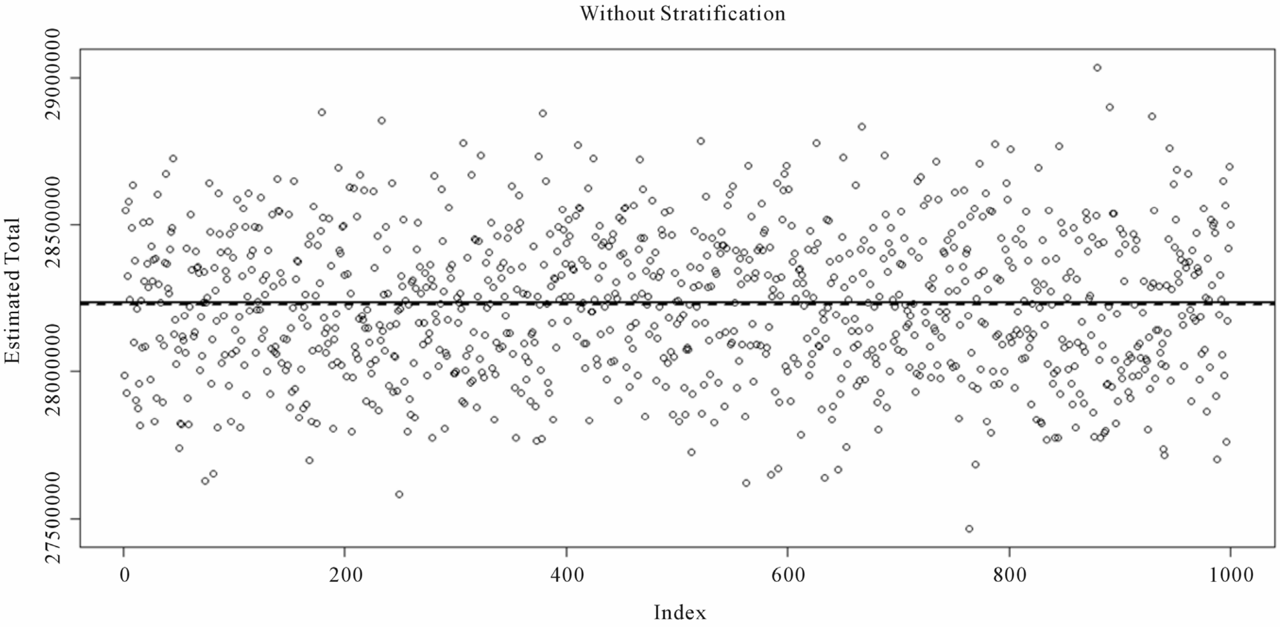

Figure 2. Plot of the population totals without stratification for the 1000 samples.

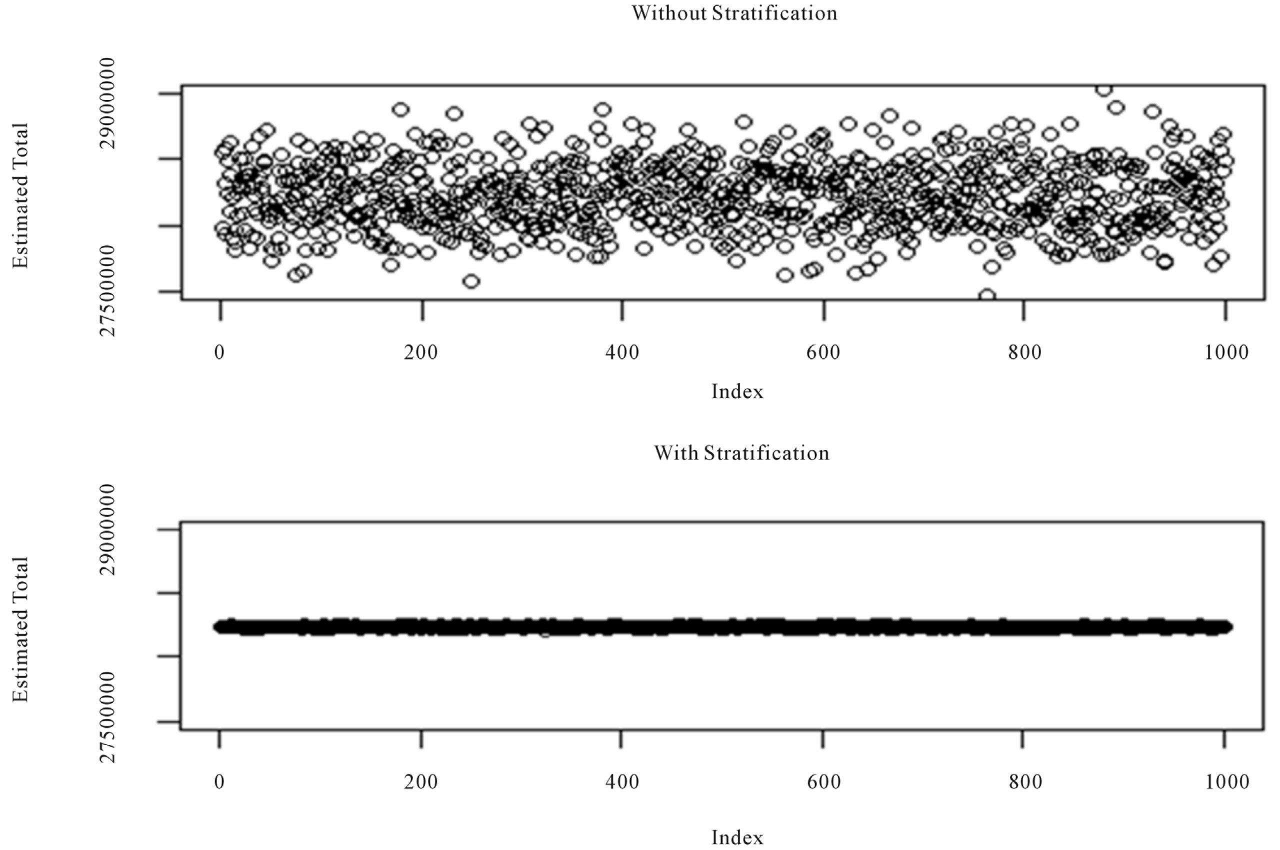

Figure 3. Figures 1 and 2 plotted on a common scale.

as compared to Olkin's estimator.

The combined graph also shows that the population total estimate is more variable in Olkin's as compared to the proposed model.

The limiting condition to allow the use of this estimator is the requirement of existence of linear relationship through the origin between the variable of interest, *y*, and the auxiliary variables.

REFERENCES

[1] I. Olkin, “Multivariate Ratio Estimation for Finite Populations,” *Biometrika*, Vol. 45, No. 1-2, 1956, pp. 154-165.

[2] W. G. Cochran, “Sampling Techniques,” 3rd Edition, Wiley, New York, 1977.

[3] L. Y. Deng and R. S. Chikura, “On the Ratio and Regression Estimation in Finite Population Sampling,” *American Statistician*, Vol. 44, No. 4, 1990, pp. 282-284.

[4] P. V. Sukhatme and B. V. Sukhatme, “Sampling Theories of Survey with Applications,” Iowa State University Press, Ames, 1970.