Research Journal of Applied Sciences 2 (9): 935-938, 2007
ISSN: 1815-932X

Monte Carlo Method of Estimation in Double Sampling under a Particular Linear Regression Model

A.A. Adewara
Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria

Abstract: In this study, we made use of a particular linear regression model for three regression and mean per unit estimators and from the estimated mean square errors obtained on these estimators through simulation, we observed that one of the newly proposed regression estimators performed better and hence, preferred.

Key words: Mean square error, estimator, estimate, particular linear regression model, double sampling

INTRODUCTION

Monte carlo technique according to Kendall and Buckland (1982) is a way of obtaining the solution of mathematical problems in a stochastic context through sampling experiments. It is the solution of any mathematical problem by sample methods. The procedure is to conduct an artificial stochastic model of the mathematical process and then to perform sampling experiments upon it.

On the model used: The linear regression model used in this study is of the form:

$$y_i = \beta x_i + \mu_i, i = 1, 2, ..., N$$

Where x_i follows a gamma distribution, μ_1 is normally distributed with mean 0 and variance $(\mu/x) = \sigma^2 x_i^t$ and t is said to be the variance function usually between 0 and 2.

The gamma distribution with parameter α, β and probability density function,

$$\frac{\beta^\alpha}{\Gamma(\alpha)} X^{\alpha-1} e^{-\beta x}$$

was used,

$$\alpha > 0, \beta > 0, E(X) = \frac{\alpha}{\beta}, E(Y) = \alpha \text{ and } Var(X) = \frac{\alpha}{\beta^2}$$

Where, α and β are defined as a "scale" and "shape" parameters, respectively while E(X) and E(Y) are the means of X and Y population.

In sample survey, gamma distribution is a distribution that has skewed population in which a small proportion of the sampling units in the population may account for a high proportion of an aggregate or average being measure. Examples of these skewed populations are sample of farms by size, distribution of sales of retail trade in establishment and family income distribution of the population.

Here, both the sigma (σ), alpha (α), beta (β) and t values are specified, where t controls the dependence of the variance of y on the value of x. It is always regarded as a non-negative constant.

Also, a statistical transformation was done from a normal distribution to uniform distribution which made x_i to be distributed as gamma since the normal distribution is the most important distribution in statistics which is likely to provide a good model for a variate when:-

- There is a strong tendency for the variate to take a central value.
- Positive and negative deviations from the central value are equally likely.
- The frequency of deviations falls off rapidly as the deviations becomes larger Cooke *et al*. (1982).

The following estimators are considered for comparisons:-

- $\bar{y}$, mean per unit
- $\bar{y}_{dl0} = \bar{y} - \beta(\bar{x} - \bar{x}')$
- $\bar{y}_{dl1} = \bar{y}^* - \beta(\bar{x}^* - \bar{x}')$ and
- $\bar{y}_{dl2} = \bar{y}^{**} - \beta(\bar{x}^{**} - \bar{x}')$

Mean square error of the conventional estimator, $\bar{y}_{dl0}$:
Let N be the population size, n' and n be the first sample and subsample drawn, $\bar{x}'$ and $\bar{x}$ be the sample mean of the first sample and subsample drawn, respectively.

$\bar{x} = \bar{X}(1 + \Delta_{\bar{x}})$ and $\bar{y} = \bar{Y}(1 + \Delta_{\bar{y}})$. Then,

$$Mse(\bar{y}) = (\frac{1}{n} - \frac{1}{N})s^2_y \approx (\frac{1}{n})s^2_y, \text{ as } N \to \infty, 1/N \to 0$$

and $\bar{y}_{dl0} = \bar{y} - \beta(\bar{x} - \bar{x}')$. Then,

$$\begin{aligned}Mse(\bar{y}_{dl0}) &\approx v(\bar{y}_{dl0}) = v[\bar{y} - \beta(\bar{x} - \bar{x}').] \\ &= v(\bar{y}) + \beta^2(\bar{x} - \bar{x}') - 2\beta\,cov(\bar{y}, \bar{x} - \bar{x}') \\ &= v(\bar{y}) + \beta^2[v(\bar{x}) + v(\bar{x}') - 2\,cov(\bar{x} - \bar{x}')] \\ &- 2\beta[cov(\bar{x}, \bar{y}) - cov(\bar{x}', \bar{y})].\end{aligned}$$

Using the unconditional expectation E_1 and conditional expectation E_2 over the first sample and the second subsample given the first sample, unconditional variance V_1 and conditional variance V_2 of the first sample and the second subsample given the first sample, then,

$$v(\bar{y}) = \frac{1}{n}s^2{}_y.$$

$$v(\bar{x}) = \frac{1}{n}s^2{}_x.$$

$$v(\bar{x}') = \frac{1}{n'}s^2{}_x.$$

$$cov(\bar{x}, \bar{x}') = \frac{1}{n'}s^2{}_x.$$

$$cov(\bar{x}, \bar{y}) = \frac{1}{n}s_{xy}.$$

$$cov(\bar{x}', \bar{y}) = \frac{1}{n'}s_{xy}.$$

Then,

$$Mse(\bar{y}_{dl0}) = (\frac{1}{n'} - \frac{1}{N})s^2{}_y + (\frac{1}{n} - \frac{1}{n'})(s^2{}_y - 2\beta s_{xy} + \beta^2 s^2{}_x).$$

Okafor (2002) as $N \to \infty$, $1/N \to 0$ Then,

$$Mse(\bar{y}_{dl0}) \approx (\frac{1}{n'})s^2{}_y + (\frac{n' - n}{n'n})(s^2{}_y - 2\beta s_{xy} + \beta^2 s^2{}_x).$$

Mean square error of $\bar{y}_{dl1}$: Let ,

$$\bar{X} = \theta\bar{x} + (1-\theta)\bar{x}^*, \bar{Y} = \theta\bar{y} + (1-\theta)\bar{y}^*, \theta = \frac{1}{n'} - \frac{1}{N}.$$

As $N \to \infty$, $1/N \to 0$. $\theta = \frac{1}{n'}$, Then,

$$\bar{x}^* = \bar{X}(1 - (\frac{1}{n'-1})\Delta_{\bar{x}^*})$$

and

$$\bar{y}^* = \bar{Y}(1 - (\frac{1}{n'-1})\Delta_{\bar{y}^*})$$

$\bar{y}_{dl1} = \bar{y}^* - \beta(\bar{x}^* - \bar{x}')$. Then,

$$\begin{aligned}Mse(\bar{y}_{dl1}) &\approx v(\bar{y}_{dl1}) = v[\bar{y}^* - \beta(\bar{x}^* - \bar{x}').] \\ &= v(\bar{y}^*) + \beta^2(\bar{x}^* - \bar{x}') - 2\beta\,cov(\bar{y}^*, \bar{x}^* - \bar{x}') \\ &= v(\bar{y}^*) + \beta^2[v(\bar{x}^*) + v(\bar{x}') - 2cov(\bar{x}^* - \bar{x}')] \\ &- 2\beta[cov(\bar{x}^*, \bar{y}^*) - cov(\bar{x}', \bar{y}^*)].\end{aligned}$$

Where,

$$v(\bar{y}^*) = \frac{1}{n}(\frac{1}{n'-1})^2 s^2{}_y.$$

$$v(\bar{x}^*) = \frac{1}{n}(\frac{1}{n'-1})^2 s^2{}_x.$$

$$v(\bar{x}') = \frac{1}{n'}s^2{}_x.$$

$$cov(\bar{x}^*, \bar{x}') = \frac{1}{n'}(\frac{1}{n'-1})s^2{}_x.$$

$$cov(\bar{x}^*, \bar{y}^*) = \frac{1}{n}(\frac{1}{n'-1})^2 s_{xy}.$$

$$cov(\bar{x}', \bar{y}^*) = \frac{1}{n'}(\frac{1}{n'-1})s_{xy}.$$

Then,

$$Mse(\bar{y}_{dl1}) \approx (\frac{1}{n'})(\frac{1}{n'-1})^2 s^2{}_y + (\frac{n'-n}{n'n})(\frac{1}{n'-1})^2 (s^2{}_y - 2\beta s_{xy} + \beta^2 s^2{}_x).$$

Mean square error of $\bar{y}_{dl2}$: Let,

$$\bar{X} = (1+\theta)\bar{x}^{**} - \theta\bar{x}, \bar{Y} = (1+\theta)\bar{y}^{**} - \theta\bar{y}, \theta = \frac{1}{n'} - \frac{1}{N}.$$

As $N \to \infty$, $1/N \to 0$. Then,

$$\theta = \frac{1}{n'}, \bar{x}^{**} = \bar{X}(1 + (\frac{1}{n'+1})\Delta_{\bar{x}^{**}})$$

and

$$\bar{y}^{**} = \bar{Y}(1 + (\frac{1}{n'+1})\Delta_{\bar{y}^{**}}) \cdot$$

$\bar{y}_{dl2} = \bar{y}^{**} - \beta(\bar{x}^{**} - \bar{x}')$. Then,

$$\begin{aligned} Mse(\bar{y}_{dl2}) &\approx v(\bar{y}_{dl2}) = v[\bar{y}^{**} - \beta(\bar{x}^{**} - \bar{x}').] \\ &= v(\bar{y}^{**}) + \beta^2(\bar{x}^{**} - \bar{x}') - 2\beta cov(\bar{y}^{**}, \bar{x}^{**} - \bar{x}') \\ &= v(\bar{y}^{**}) + \beta^2[v(\bar{x}^{**}) + v(\bar{x}') - 2cov(\bar{x}^{**} - \bar{x}')] \\ &- 2\beta[cov(\bar{x}^{**}, \bar{y}^{**}) - cov(\bar{x}', \bar{y}^{**})]. \end{aligned}$$

Where,

$$v(\bar{y}^{**}) = \frac{1}{n}(\frac{1}{n'+1})^2 s^2_y.$$

$$v(\bar{x}^{**}) = \frac{1}{n}(\frac{1}{n'+1})^2 s^2_x.$$

$$v(\bar{x}') = \frac{1}{n'} s^2_x.$$

$$cov(\bar{x}^{**}, \bar{x}') = \frac{1}{n'}(\frac{1}{n'+1}) s^2_x.$$

$$cov(\bar{x}^{**}, \bar{y}^{**}) = \frac{1}{n}(\frac{1}{n'+1})^2 s_{xy}.$$

$$cov(\bar{x}', \bar{y}^{**}) = \frac{1}{n'}(\frac{1}{n'+1}) s_{xy}.$$

Then,

$$Mse(\bar{y}_{dl2}) \approx (\frac{1}{n'})(\frac{1}{n'+1})^2 s^2_y + (\frac{n'-n}{n'n})(\frac{1}{n'+1})^2 (s^2_y - 2\beta s_{xy} + \beta^2 s^2_x).$$

$$Bias(\bar{y}_{dl0}) = Bias(\bar{y}_{dl1}) = Bias(\bar{y}_{dl2}) = 0$$

Therefore,

$$Mse(\bar{y}_{dl0}) \approx (\frac{1}{n'}) s^2_y + (\frac{n'-n}{n'n})(s^2_y - 2\beta s_{xy} + \beta^2 s^2_x).$$

$$Mse(\bar{y}_{dl1}) \approx (\frac{1}{n'})(\frac{1}{n'-1})^2 s^2_y + (\frac{n'-n}{n'n})(\frac{1}{n'-1})^2 (s^2_y - 2\beta s_{xy} + \beta^2 s^2_x)$$

and

$$Mse(\bar{y}_{dl2}) \approx (\frac{1}{n'})(\frac{1}{n'+1})^2 s^2_y + (\frac{n'-n}{n'n})(\frac{1}{n'+1})^2 (s^2_y - 2\beta s_{xy} + \beta^2 s^2_x).$$

Here, the coefficient of $(\frac{1}{n'}) s^2_y$ and $(\frac{n'-n}{n'n})(s^2_y - 2\beta s_{xy} + \beta^2 s^2_x)$:-

- +1 and +1 for $\bar{y}_{dl0}$
- $(\frac{1}{n'-1})^2$ and $(\frac{1}{n'-1})^2$ for $\bar{y}_{dl1}$
- $(\frac{1}{n'+1})^2$ and $(\frac{1}{n'+1})^2$ for $\bar{y}_{dl2}$

One could see that their differences lies on these coefficients.

In order to know the effect of the various values of β used, this research would focus on the formulae above instead of equivalent sample formulae stated below:-

$$Mse(\bar{y}_{dl0}) \approx (\frac{1}{n'}) s^2_y + (\frac{n'-n}{n'n})(\frac{n-1}{n-2}) s^2_y (1 - \hat{\rho}^2).$$

$$Mse(\bar{y}_{dl1}) \approx (\frac{1}{n'})(\frac{1}{n'-1})^2 s^2_y + (\frac{n'-n}{n'n})(\frac{n-1}{n-2})(\frac{1}{n'-1})^2 s^2_y (1 - \hat{\rho}^2)$$

and

$$Mse(\bar{y}_{dl2}) \approx (\frac{1}{n'})(\frac{1}{n'+1})^2 s^2_y + (\frac{n'-n}{n'n})(\frac{n-1}{n-2})(\frac{1}{n'+1})^2 s^2_y (1 - \hat{\rho}^2).$$

Under what condition would an estimator of this kind be preferred?: An estimator of this kind among others being considered here would be preferred if

- It has the least estimated bias.
- It has the least estimated mean square error.
- It has the least estimated mean square error ratio over mean per unit estimator which must be less than or equals 1 and.
- It has the highest percentage relative efficiency.

RESULTS AND DISCUSSION

Using the gamma distribution described above, where n'= 140, n = 2,20,40,80 and 100, α = 1,2,3 and β = 1,2 and 3. Then, the estimates obtained using these information are shown in the Table 1-3.

Table 1: Estimated mean square error of $\bar{y}, \bar{y}_{dl0}, \bar{y}_{dl1}$ and $\bar{y}_{dl2}$ when n' = 140, n = 2, 20, 40, 80 and 100, $\alpha = 1$ and $\beta = 1, 2$ and 3

n'	n	α	β	mse($\bar{y}$)	mse($\bar{y}_{dl0}$)	mse($\bar{y}_{dl1}$)	mse($\bar{y}_{dl2}$)
140	2	1	1	1.8242	1.8552	9.6×10^{-5}	9.3×10^{-5}
			2	27.5348	27.8381	0.00144	0.00140
			3	123.5017	123.8050	6.41×10^{-3}	6.23×10^{-3}
140	20	1	1	0.0333	0.0419	2.2×10^{-6}	2.1×10^{-6}
			2	0.5649	0.5743	2.97×10^{-5}	2.89×10^{-5}
			3	2.9739	2.9833	1.54×10^{-4}	1.5×10^{-4}
140	40	1	1	0.0165	0.0220	1.14×10^{-6}	1.11×10^{-6}
			2	0.2804	0.2869	1.49×10^{-5}	1.44×10^{-5}
			3	1.4759	1.4824	7.7×10^{-5}	7.5×10^{-5}
140	80	1	1	0.0108	0.0118	6.1×10^{-6}	5.9×10^{-6}
			2	0.1460	0.1472	7.6×10^{-6}	7.4×10^{-6}
			3	0.7181	0.7193	3.7×10^{-5}	3.6×10^{-5}
140	100	1	1	31.3132	34.1866	1.77×10^{-3}	1.72×10^{-3}
			2	0.1203	0.1209	6.26×10^{-6}	6.08×10^{-6}
			3	0.6009	0.6015	3.11×10^{-5}	3.03×10^{-5}

Table 2: Estimated mean square error of $\bar{y}, \bar{y}_{dl0}, \bar{y}_{dl1}$ and $\bar{y}_{dl2}$ when n' = 140,n = 2, 20, 40, 80 and 100, $\alpha = 2$ and $\beta = 1, 2$ and 3

n'	n	α	β	mse($\bar{y}$)	mse($\bar{y}_{dl0}$)	mse($\bar{y}_{dl1}$)	mse($\bar{y}_{dl2}$)
140	2	2	1	1.0562	1.2387	6.4×10^{-5}	6.23×10^{-5}
			2	7.9533	8.1357	4.2×10^{-4}	4.09×10^{-4}
			3	33.7278	33.9102	1.76×10^{-3}	1.71×10^{-3}
140	20	2	1	0.0703	0.0849	4.39×10^{-6}	4.23×10^{-6}
			2	1.3694	1.3854	7.17×10^{-5}	6.97×10^{-5}
			3	7.1739	7.1894	3.72×10^{-4}	3.6210^{-4}
140	40	2	1	1.7629	2.2500	1.2×10^{-6}	1.1×10^{-6}
			2	0.2495	0.2543	1.32×10^{-5}	1.28×10^{-5}
			3	1.2757	1.2807	6.63×10^{-5}	6.44×10^{-5}
140	80	2	1	0.0248	0.0268	1.38×10^{-6}	1.34×10^{-6}
			2	0.3354	0.3373	1.75×10^{-5}	1.70×10^{-5}
			3	1.6836	1.6856	8.7×10^{-5}	8.5×10^{-5}
140	100	2	1	0.0241	0.0247	1.28×10^{-6}	1.24×10^{-6}
			2	0.3590	0.3597	1.86×10^{-5}	1.81×10^{-5}
			3	1.8133	1.8140	9.4×10^{-5}	9.1×10^{-5}

Table 3: Estimated mean square error of $\bar{y}, \bar{y}_{dl0}, \bar{y}_{dl1}$ and $\bar{y}_{dl2}$ when n' = 140, n = 2, 20, 40, 80 and 100, $\alpha = 3$ and $\beta = 1, 2$ and 3

n'	n	α	β	mse($\bar{y}$)	mse($\bar{y}_{dl0}$)	mse($\bar{y}_{dl1}$)	mse($\bar{y}_{dl2}$)
140	2	3	1	0.0673	0.1045	5.4×10^{-6}	5.3×10^{-6}
			2	2.6259	2.6631	1.38×10^{-6}	1.34×10^{-6}
			3	15.1234	15.1605	7.8×10^{-4}	7.60×10^{-4}
140	20	3	1	0.2157	0.2313	1.7×10^{-5}	1.16×10^{-5}
			2	2.7557	2.7724	1.43×10^{-4}	1.39×10^{-4}
			3	13.3728	13.3895	0.00069	0.00067
140	40	3	1	0.0570	0.0618	3.20×10^{-6}	3.11×10^{-6}
			2	0.8651	0.8700	4.5×10^{-5}	4.38×10^{-5}
			3	4.4034	4.4082	2.28×10^{-4}	2.22×10^{-4}
140	80	3	1	0.0514	0.0526	2.72×10^{-6}	2.65×10^{-6}
			2	0.7721	0.7733	4.0×10^{-5}	3.89×10^{-5}
			3	3.8821	3.8833	2.01×10^{-4}	1.95×10^{-4}
140	100	3	1	0.0245	0.0252	1.3×10^{-6}	1.27×10^{-6}
			2	0.3563	0.3570	1.85×10^{-5}	1.80×10^{-5}
			3	1.7941	1.7947	9.29×10^{-5}	9.03×10^{-5}

From Table 1-3, irrespective of the values of α and β, the unique findings are that double sampling regression estimator, $\bar{y}_{dl2}$ has:

- The least estimated mean square error.
- The least estimated mean square error ratio over mean per unit estimator which is less than 1.
- The highest percentage relative efficiency
- $mse(\bar{y}_{dl2}) < mse(\bar{y}_{dl1}) < mse(\bar{y}) < mse(\bar{y}_{dl0})$. Hence, double sampling regression estimator, $\bar{y}_{dl2}$ is preferred.

REFERENCES

Cooke, D., A.H. Craven and G.M. Clarke, 1982. Basic Statistical Computing, Edward Arnoid (Publishers) Ltd.

Kendall, M.G. and W.R. Buckland, 1982. A dictionary of Statistical terms. 4th Edn. Longman Group Ltd, London and New York.

Okafor, C.O., 2002. Sample Survey Theory with Applications. Afro-orbis Publications ltd, University of Nsukka, Nigeria.