

GENERALISED LYNCH MULTIVARIATE REGRESSION ESTIMATORS WITH k -PARAMETERS OF ORDER $\frac{1}{n}$

Adewara, A. A., Job, O., Abidoye, A. O. and Oyeyemi, G. M.

Department of Statistics
University of Ilorin
Ilorin, Kwara State
Nigeria

(e-mail : aaadewara@yahoo.com)

and

Gali, M. O.

FHA Homes Ltd.
P.M.B. 414, Garki
Abuja, FCT
Nigeria

Alabi, O. O.

Department of Mathematical Sciences
Olabisi Onabanjo University
Ago-Iwoye, Ogun State
Nigeria

ABSTRACT

In this paper, several authors have proposed ratio type estimators which utilized data from several auxiliary variates that involves the use of unknown weights which have to be estimated but Lynch proposed a multivariate regression estimator when there are two auxiliary variates (x -variates) which was found to be better and preferred, even to the conventional mean estimate.

In this paper, we proposed multivariate regression estimators with k -parameters of order $\frac{1}{n}$ as used by Lynch. Two data sets called Populations 1 and 2 were used to justify this research work. While Population 1 is based on the total monthly income and total monthly expenditure on food, rent, clothing, transportation and miscellaneous of 86 occupants of Kubwa Federal Housing Authority, Phase IV Estate, Abuja, Nigeria, Population 2 is based on the total monthly income and total monthly expenditure on food, rent, clothing, transportation and miscellaneous of 95 occupants of Gwarinpa II Estate, Abuja, Nigeria and it was observed that as k -parameters increases, our k multivariate regression estimate becomes smaller and better which makes the highest k -parameter multivariate regression estimator to be preferred.

Introduction

Several authors like Olkin (1958), Cochran (1977) and host of others have proposed ratio type estimators which utilized data from several auxiliary variates that involves the use of unknown weights which have to be estimated and assume knowledge of the population means of the auxiliary characteristics used Lynch (1978).

Also, a ratio-type multi-variate estimator which does not involve unknown weights and which, at most, assumes knowledge of the population means of the auxiliary characteristics least correlated

Keywords: Multivariate, Regression, Estimator, Parameter, Order, Mean, Variance.

with the characteristics of interest had being proposed earlier. Lynch (1978). Not quite long time ago, several authors, like Marsy (1999), Jianqing et. al. (1992), Volaufora et. al. (2007), Kang-Mo (2005) and host of others had contributed in no small measures to further researches in multivariate regression estimators.

In furtherance of researches in multivariate regression estimators, in this paper, we shall present several multivariate regression estimators with k -parameters of order $\frac{1}{n}$ that does not involves unknown weights. Their efficiency will be compared based on their approximate variances so that judgments can be made concerning the best choice of multivariate regression estimatoors in finite population sample surveys.

On Lynch Conventional Multivariate Regression Estimator

Since the most frequent applications for these methods are likely to be when there are two auxiliary variates (x -variates), Lynch (1978) proposed a multivariate regression estimator of the form,

$$\bar{Y}_{MR2} = \bar{y} + b_1(\bar{X}_1 - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2) \quad (1)$$

and

$$v(\bar{Y}_{MR2}) = \frac{(1-f)}{n} [s_y^2 - b_1^2 s_{x_1}^2 - b_2^2 s_{x_2}^2 - b_1 b_2 s_{x_1 x_2}] \quad (2)$$

where b_1 and b_2 are found by the usual least square methods.

Also,

$$v(\bar{y}) = \frac{(1-f)}{n} s_y^2 \quad (3)$$

called the conventional mean estimator.

On Our Porposed Generalized Lynch Multivariate Regression Estimator

Based on Lynch multivariate regression estimator that does not involves unknown weights, we then generalized our own multivariate regression estimators, thus:

$$(i) \quad \bar{Y}_{MR3} = \bar{y} + b_1(\bar{X}_1 - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2) + b_3(\bar{X}_3 - \bar{x}_3) \quad (4)$$

$$(ii) \quad \bar{Y}_{MR4} = \bar{y} + b_1(\bar{X}_1 - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2) + b_3(\bar{X}_3 - \bar{x}_3) + b_4(\bar{X}_4 - \bar{x}_4) \quad (5)$$

and

$$(iii) \quad \bar{Y}_{MR5} = \bar{y} + b_1(\bar{X}_1 - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2) + b_3(\bar{X}_3 - \bar{x}_3) + b_4(\bar{X}_4 - \bar{x}_4) + b_5(\bar{X}_5 - \bar{x}_5) \quad (6)$$

Generally,

$$(iv) \quad \bar{Y}_{MRk} = \bar{y} + b_1(\bar{X}_1 - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2) + b_3(\bar{X}_3 - \bar{x}_3) + \dots + b_k(\bar{X}_k - \bar{x}_k) \quad (7)$$

where $b_{i,s}$ are found by the usual least square methods.

Approximate Variances Of $\bar{Y}_{MRi}$

In line with Lynch approximate variance, we then have:

$$(i) \quad v(\bar{Y}_{MR3}) = \frac{(1-f)}{n} [s_y^2 - b_1^2 s_{x_1}^2 - b_2^2 s_{x_2}^2 - b_3^2 s_{x_3}^2 - b_1 b_2 s_{x_1 x_2} - b_1 b_3 s_{x_1 x_3} - b_2 b_3 s_{x_2 x_3}] \quad (8)$$

$$(ii) \quad v(\bar{Y}_{MR4}) = \frac{(1-f)}{n} [s_y^2 - b_1^2 s_{x_1}^2 - b_2^2 s_{x_2}^2 - b_3^2 s_{x_3}^2 - b_4^2 s_{x_4}^2 - b_1 b_2 s_{x_1 x_2} - b_1 b_3 s_{x_1 x_3} - b_1 b_4 s_{x_1 x_4} - b_2 b_3 s_{x_2 x_3} - b_2 b_4 s_{x_2 x_4} - b_3 b_4 s_{x_3 x_4}] \quad (9)$$

and

$$(iii) \quad v(\bar{Y}_{MR5}) = \frac{(1-f)}{n} [s_y^2 - b_1^2 s_{x_1}^2 - b_2^2 s_{x_2}^2 - b_3^2 s_{x_3}^2 - b_4^2 s_{x_4}^2 - b_5^2 s_{x_5}^2 - b_1 b_2 s_{x_1 x_2} - b_1 b_3 s_{x_1 x_3} - b_1 b_4 s_{x_1 x_4} - b_1 b_5 s_{x_1 x_5} - b_2 b_3 s_{x_2 x_3} - b_2 b_4 s_{x_2 x_4} - b_2 b_5 s_{x_2 x_5} - b_3 b_4 s_{x_3 x_4} - b_3 b_5 s_{x_3 x_5} - b_4 b_5 s_{x_4 x_5}] \quad (10)$$

Generally,

$$(iv) \quad v(\bar{Y}_{MRk}) = \frac{(1-f)}{n} [s_y^2 - b_1^2 s_{x_1}^2 - \dots - b_k^2 s_{x_k}^2 - b_1 b_2 s_{x_1 x_2} - \dots - b_{k-1} b_k s_{x_{k-1} x_k}] \quad (11)$$

Data Used**Population 1**

The data used for population 1 is based on the total monthly income (y) and total monthly expenditure distributions on food (x_1), rent (x_2), clothing (x_3), transportation (x_4) and miscellaneous (x_5) of 86 occupants of Kubwa Federal Housing Authority, Phase IV Estate, Abuja, Nigeria. Then samples of size 28 are drawn using simple random sampling method.

Summary Of Estimates

$N = 86$, $n = 28$, $\bar{y} = 670088.4532$, $\bar{x}_1 = 13642.86$, $\bar{x}_2 = 11571.43$, $\bar{x}_3 = 5107.14$, $\bar{x}_4 = 8214.29$, $\bar{x}_5 = 17405.09$, $s_y^2 = 1441982834$, $s_{x_1}^2 = 84238095.24$, $s_{x_2}^2 = 50846560.85$, $s_{x_3}^2 = 9654761.905$, $s_{x_4}^2 = 34544973.54$, $s_{x_5}^2 = 228159079.2$, $s_{x_1 x_2} = 56346938.78$, $s_{x_1 x_3} = 17645408.16$, $s_{x_1 x_4} = 44826530.61$, $s_{x_1 x_5} = 85477901.12$, $s_{x_2 x_3} = 13653061.22$, $s_{x_2 x_4} = 35341836.73$, $s_{x_2 x_5} = 63452991.63$, $s_{x_3 x_4} = 11762755.1$, $s_{x_3 x_5} = 26666850.66$, $s_{x_4 x_5} = 51901959.18$, $b_1 = 0.673$, $b_2 = 1.047$, $b_3 = 0.578$, $b_4 = 1.673$ and $b_5 = 1.158$

Population 2

The data used for population 2 is based on total monthly income and total monthly expenditure distribution on food (x_1), rent (x_2), clothing (x_3), transportation (x_4) and miscellaneous (x_5) of 95 occupants of Gwarinpa II Estate, Abuja, Nigeria. Then samples of size 24 are drawn using simple random sampling method.

Summary Of Estimates

$N = 95$, $n = 24$, $\bar{y} = 44758.91542$, $\bar{x}_1 = 14803.75$, $\bar{x}_2 = 13416.67$, $\bar{x}_3 = 4416.67$, $\bar{x}_4 = 8750.0$, $\bar{x}_5 = 12333.33$, $s_y^2 = 491351392.6$, $s_{x_1}^2 = 100613033.2$, $s_{x_2}^2 = 84862318.84$, $s_{x_3}^2 = 11057971.01$, $s_{x_4}^2 = 52956521.74$, $s_{x_5}^2 = 143753623.2$, $s_{x_1x_2} = 68045520.83$, $s_{x_1x_3} = 25995104.17$, $s_{x_1x_4} = 61598854.17$, $s_{x_1x_5} = 59923750.0$, $s_{x_2x_3} = 27118055.56$, $s_{x_2x_4} = 58062500$, $s_{x_2x_5} = 65881944.44$, $s_{x_3x_4} = 21843750$, $s_{x_3x_5} = 27788194.44$, $s_{x_4x_5} = 58666666.67$, $b_1 = 0.815$, $b_2 = -0.626$, $b_3 = -3.231$, $b_4 = 2.33$ and $b_5 = 0.460$.

Results

From the estimates obtained on the two data sets (populations 1 and 2) above, we have this summary:

Estimator	Population 1	Population 2
$v(\bar{y})$	3.473×10^7	1.530×10^7
$v(\bar{Y}_{MR2})$	3.1360×10^7	1.365×10^7
$v(\bar{Y}_{MR3})$	3.0692×10^7	1.0863×10^7
$v(\bar{Y}_{MR4})$	2.4707×10^7	0.6311×10^7
$v(\bar{Y}_{MR5})$	0.6690×10^7	0.3884×10^7

Here,

For population 1, we observed that (i). $v(\bar{Y}_{MR2}) < v(\bar{y})$, meaning that Lynch multivariate regression estimator ($\bar{Y}_{MR2}$) is better than the mean estimator ($\bar{y}$) because it has the least approximated variance as claimed by the author in the literature and

(ii) $v(\bar{Y}_{MR5}) < v(\bar{Y}_{MR4}) < v(\bar{Y}_{MR3}) < v(\bar{Y}_{MR2}) < v(\bar{y})$, meaning that the fifth multivariate regression estimator, ($\bar{Y}_{MR5}$), has the least approximated variance being the highest parameter considered in this research work.

Also, for population 2, we observed that (i). $v(\bar{Y}_{MR2}) < v(\bar{y})$, meaning that Lynch multivariate regression estimator ($\bar{Y}_{MR2}$) is better than the mean estimator ($\bar{y}$) because it has the least approximated variance as claimed by the author in the literature and

(ii) $v(\bar{Y}_{MR5}) < v(\bar{Y}_{MR4}) < v(\bar{Y}_{MR3}) < v(\bar{Y}_{MR2}) < v(\bar{y})$, meaning that the fifth multivariate regression estimator ($\bar{Y}_{MR5}$) has the least approximated variance being the highest parameter considered in this research work.

Conclusion

Considering multivariate regression estimators of k -parameters with only terms of order $\frac{1}{n}$, we can conclude that:

(i) Lynch multivariate regression estimator is better than ordinary mean estimator. Than is Lynch multivariate regression estimator $((\bar{Y}_{MR2}))$ is better than the mean estimator $((\bar{y}))$ because it has the least approximated variance as claimed by the author in the literature.

(ii) As the multivariate regression estimators of k -parameter increases, the corresponding approximated variances decreases indicating that the highest multivariate regression estimator of k -parameter, $(\bar{Y}_{MRk})$, will be preferred since it has the least approximated variance.

Therefore, the generalized multivariate regression estimators of k -parameters, $(\bar{Y}_{MRk})$, with only terms of the order $\frac{1}{n}$ are realistic. Further research could be embarked on when we have multivariate regression estimators of k -parameters, $(\bar{Y}_{MRk})$, with terms of the order $> \frac{1}{n}$ which is not part of this research work.

REFERENCES

1. Cochran, W. G. : Sampling Techniques, New York, John Wiley and Sons, 3rd Edition. (ISBN. 0-471-16240-X. QA-276.6.C6, 1977, 001.4'222.77-728).
2. Jianqing, Fan. and Masry, E. : Multivariate Regression Estimation With Errors-In Variables: A Symptotic Normality For Mixing Processes. Journal Of Multivariate Analysis, (ISSN: 0047-259X, CODEN: JMVAAL). 43(2), (1992), 237-271.
3. Kang-Mo Jung : Multivariate Least-Trimmed Squares Regression Estimator. Computational Statistics And Data Analysis, (DOI: 10. 1016/j.cida.2004.01.008), 48(2), (2005), 307-316.
4. Lynch, G. W. : The Choice Of Auxiliary Variables In Multivariate Ratio And Regression Estimators. (1978), www.amstat.org/Sections/Srms/Proceedings/Papers/1978_137.
5. Masry, E. : Multivariate Regression Estimation Of Continuous-Time Processes From Sampled Data. Local Polynomial Fitting Approach. Information Theory, IEEE Transactions, (DOI: 10.1109/18.7821116). 45(6), (1999), 1939-1953.
6. Olkin, I. : Multivariate Ratio Estimation For Finite Populations. Biometrika, 45, (1958), 154-165.
7. Volaufora, Julia and Komornik. Jozef : Weighted Multivariate Regression Estimates Solved By Random Effects Approach. Biometrical Journal, (DOI: 10.1002/bimj. 4710360304). 36(3). (2007), 275-283.