Middle-East Journal of Scientific Research 11 (9): 1294-1298, 2012
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2012.11.09.1601

Best Linear Unbiased Estimator for the Linear Regression Model with Stable Disturbances using Raked Set Sample

[1]*Sahar Dorniani* and [2]*Nader Nematollahi*

[1]Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of Statistics, Allameh Tabataba'i University, Tehran, Iran

Abstract: In statistical settings where the cost of identifying experimental units and ranking them according to the attribute of interest is small compared to the cost of making formal measurements, a ranked set sampling provides improved efficiency over a simple random sample of the same size. In this paper, we consider the simple linear regression model with stable disturbances and with replicated observations. We derive the Best Linear Unbiased Estimator (BLUE) for parameter of this model, based on a ranked set sample. Furthermore, we compare numerically, mean absolute deviation of this estimator with the BLUE of this model, based on a simple random sampling.

Key words: Best linear unbiased estimator . linear regression model . ranked set sampling . stable distributions

INTRODUCTION

Consider the standard linear regression model

$$Y_{ij} = \beta x_i + U_{ij};\ i = 1,2,\cdots,n,\ \ j = 1,2,\cdots,m \qquad (1)$$

where Y_{ij} is dependent variable (or response variable) with m replication, x_i is independent variable (or predictor variable), β is an unknown parameter and U_{ij}'s are independent identically distributed random variables.

Usually, the parameter β in model (1) is estimated by Ordinary Least Square (OLS) method, which provides important results of Gauss-Markov theorem and central limit theorem. In these theorems, U_{ij}'s must have finite variance and usually have Normal distribution.

But, in several cases the variance of U_{ij} is infinite. For example, Fama [4] and Mittnik and Rachev [7]. In these cases suitable distribution for U_{ij} is a stable distribution.

A random variable X is said to have a stable distribution if there are parameters $0<\alpha<2$, $-1\le\beta\le1$, $\gamma<0$ and $-\infty<\delta<+\infty$ such that its characteristic function has the following form Nolan [11],

$$\phi_X(u) = E[\exp(iuX)]$$
$$= \begin{cases} \exp(-\gamma^{\alpha}|u|^{\alpha}[1 - i\beta(\tan\frac{\pi\alpha}{2})(\text{sign}(u))] + i\delta u) & ;\alpha \ne 1 \\ \exp(-\gamma|u|[1 + i\beta\frac{2}{\pi}(\text{sign}(u))\ln(u)] + i\delta u) & ;\alpha = 1 \end{cases}$$

where sign(.) is the sign function.

We will denote stable distribution by $S(\alpha,\beta,\gamma,\delta)$. In this notation the parameters α,β,γ and δ are called index of stability or characteristic exponent, skew parameter, scale parameter and location parameter, respectively.

We note that, if $X\sim S(\alpha,\beta,\gamma,\delta)$ then $E|X|^p = \infty$ for $p\ge\alpha$ and $E|X|^p < \infty$ for $0<p<\alpha$ [12]. So, for $0<\alpha<2$, the second moment of stable distribution does not exist and we cannot use the OLS method for the estimate of β.

In this paper, we consider the $S(\alpha,0,\gamma,0)$ for distribution of U_{ij}. Therefore, characteristic function of U_{ij} has the following simple form

$$\phi_U(u) = \exp(-\gamma^{\alpha}|u|^{\alpha}) \qquad (2)$$

Blattberg and Sargent [2] use another method (except OLS) for estimating β in model (1) with stable's errors as in (2). They show that the Best Linear Unbiased Estimator (BLUE) of β has the following form:

$$\hat{\beta}_{BLUE} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}|x_i|^{1/(\alpha-1)}\,\text{sign}(x_i)Y_{ij}}{\sum_{i=1}^{n}|x_i|^{\alpha/(\alpha-1)}},\ \ 1<\alpha<2 \qquad (3)$$

This estimator has two restrictions. It cannot be calculated for $\alpha\le1$ and its variance does not exist. To

Corresponding Author: Sahar Dorniani, Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran

overcome the restrictions of this estimator, we propose an estimator based on Ranked Set Sampling (RSS) method. To this end, in Section 2 we described RSS method. In Section 3 we derive BLUE for parameter β based on RSS and finally in Section 4 we compare our estimator of β with the estimator (3) by a simulation study.

RANKED SET SAMPLING

Ranked set sampling is known to be a statistical method for data collection that generally leads to more efficient estimators than competitors based on Simple Random Sampling (SRS). In real life sampling situation where the measurement of the variable of interest from the experimental units is costly or time-consuming but the ranking of sample items related to the variable can be easily done by judgment without actual measurement, the RSS method can be used and is highly beneficial and much superior to the standard SRS. The concept of RSS was used first time by McIntyre [6], to estimate the population mean of pasture yields in agricultural experimentation. Dell and Clutter [3] provided mathematical foundation for RSS. The properties of RSS in a variety of statistical procedures have been investigated in the literature.

The concept behind a RSS in a simple linear regression model with replication can be briefly described as follows: suppose for each x_i in model (1), we have m^2 values of Y such as $Y_{i1},\ldots,Y_{im^2}$. We divide these values into m groups, $V_{i1},\ldots,V_{im}$, each group contains m random variables. Let $Y_{i,j(j)}$ be the j^{th} ordered element in group V_{ij}. Using RSS method, for each x_i $(i = 1,\ldots,n)$ we only measure m values of dependent variable, $Y_{i,j(j)}$ $(j = 1,\ldots,m)$. Table 1 shows the ordered statistics and ranked set samples obtained from them for each x_i.

For simplicity, we denoted $Y_{i,j(j)}$ by $Y_{i(j)}$. In fact, $Y_{i(j)}$ is the j^{th} order statistics in the group V_{ij} for each x_i. It is important to note that the only measured variable in RSS is $Y_{i(j)}$ for $i = 1,\ldots,n$, $j = 1,\ldots,m$.

BEST LINEAR UNBIASED ESTIMATOR OF b

Usually for finding BLUE of an arbitrary parameter, a linear combination of a random sample is considered and its coefficients are determined such that it has minimum variance under unbiased condition. But since the moments of a random variable with stable distribution, here Y_{ij}, does not exist, we cannot use a random sample of Y_{ij} for constructing BLUE for β in model (1). In this paper, we use linear combination of ranked set sample, i.e., $Y_{i(j)}$, for estimating β.

Table 1: Display of m random sample for constructing RSS for each x_i

V_{i1}:	$Y_{i,1(1)}$	$Y_{i,1(2)}$	$\cdots$	$Y_{i,1(m)}$
V_{i2}:	$Y_{i,2(1)}$	$Y_{i,2(2)}$	$\cdots$	$Y_{i,2(m)}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
V_{im}:	$Y_{i,m(1)}$	$Y_{i,m(2)}$	$\cdots$	$Y_{i,m(m)}$

It should be noted that all of the moments of order statistics of stable distributions, which produce a ranked set sample $Y_{i(j)}$, does not exist. Therefore, we are limited ourselves to use the order statistics for which their moments are exist. Number of these order statistics can be determined by the following lemma.

Lemma 3.1: Let m be a real number, $X_1,\ldots,X_m$ be a random sample from $S(\alpha,\beta,\gamma,\delta)$-distribution with $0<\alpha<2$ and $X_{(1)},\ldots,X_{(m)}$ be its corresponding order statistics.

(I) Suppose $-1<\beta<1$. In order that $E\left(X_{(j)}^k\right)$ exists it is necessary and sufficient that $\alpha^{-1}k < j < m+1-\alpha^{-1}k$.

(II) Suppose $\alpha\geq 1$ and $\beta = 1$ (or $\beta = -1$). In order that $E\left(X_{(j)}^k\right)$ exists it is sufficient that $\alpha^{-1}k < j < m+1-\alpha^{-1}k$.

(III) Suppose $\alpha<1$ and $\beta = 1$ (or $\beta = -1$). In order that $E\left(X_{(j)}^k\right)$ exists it is necessary and sufficient that $j < m+1-\alpha^{-1}k$ (or $\alpha^{-1}k < j$).

For a proof of Lemma 3.1 see Mohammadi and Mohammadpour [8].

Let $k = 2$ and $\beta = 0$, then from part (I) of Lemma 3.1 and the fact that the bounds of j must be an integer, for each x_i we can find indices j for which the second moment (or variance) of $Y_{i(j)}$ exists. Thus, for $0<\alpha<2$,

$$E\,|\,Y_{i(j)}\,|^2<\infty \text{ iff } j\in J=\{s: \frac{2}{\alpha}<s<m+1-\frac{2}{\alpha},\ s=1,\ldots,m\} \quad (4)$$

Therefore, we propose to find BLUE of β based on $Y_{i(j)}$'s with finite second moment. This idea was used in Ni Chuiv *et al*. [9] to find BLUE for location parameter of Cauchy distribution, which is a stable distribution with $\alpha = 1$ and $\beta = 0$.

Barreto and Barnett [1] use all elements of RSS to find BLUE of β in (1) where the error terms have Normal distribution. We use their procedure with stable errors. Let $Y_{i(j)}$, $j\in J$ be a RSS drown from stable distribution.

We recall that if $U \sim S(\alpha,0,\gamma,0)$ with known parameter α and γ, then

$$Z = \frac{U}{\gamma} \sim S(\alpha,0,1,0)$$

(Nolan, [11]). Therefore, from model (1),

$$Z_{ij} = (Y_{ij} - \beta x_i)/\gamma = U_{ij}/\gamma \sim S(\alpha,0,1,0)$$

Now, let

$$Z_{i(j)} = \frac{Y_{i(j)} - \beta x_i}{\gamma} \qquad (5)$$

where $Z_{i(1)}, Z_{i(2)},\ldots, Z_{i(m)}$ is a RSS from Z corresponding to x_i, i = 1,2,…,n. We denoted the mean and variance of $Z_{i(j)}$ by $\eta_{i(j)}$ and $\tau_{i(j)}$ respectively, which can be calculated as follows:

$$\tau_{i(j)} = \zeta_{i(j)} - \eta_{i(j)}^2 \qquad (6)$$

and

$$\begin{aligned}\eta_{i(j)} &= E(Z_{i(j)})\\ &= \frac{m!}{(j-1)!(m-j)!}\int_{-\infty}^{\infty} z f_Z(z)(F_Z(z))^{j-1}(1-F_Z(z))^{m-j}dz\end{aligned}$$

$$\begin{aligned}\zeta_{i(j)} &= E(Z_{i(j)}^2)\\ &= \frac{n!}{(j-1)!(m-j)!}\int_{-\infty}^{\infty} z^2 f_Z(z)(F_Z(z))^{j-1}(1-F_Z(z))^{m-j}dz\end{aligned} \qquad (7)$$

where $f_z(.)$ and $F_z(.)$ are pdf and cdf of Z, respectively. With attention to (7), it can be seen that for each i = 1,…,n, $\eta_{i(j)}$ and $\tau_{i(j)}$ do not depend on i. So, we denote them by $\eta_{(j)}$ and $\tau_{(j)}$, respectively.

We note that, because there is no explicit formula for the pdf and cdf of stable distributions, for computing the values of $f_z(z)$ and $F_z(z)$ we use STABLE program of Nolan [10]. So, integrals in (7) can be calculated numerically through a short code with a reasonable precision.

Now, by using (5) we can write

$$E(Y_{i(j)} \mid x_i) = \beta x_i + \gamma\eta_{i(j)}$$

and

$$Var(Y_{i(j)} \mid x_i) = \gamma^2 \tau_{i(j)}$$

for $i = 1,2,\ldots,n,\ j = 1,2,\ldots,m$.

A matrix form for this model is

$$Y = XB + U$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_{1(1)}\\ \vdots\\ Y_{1(m)}\\ Y_{2(1)}\\ \vdots\\ Y_{2(m)}\\ \vdots\\ Y_{n(1)}\\ \vdots\\ Y_{n(m)}\end{pmatrix},\ \mathbf{X} = \begin{pmatrix} x_1 & \eta_{1(1)}\\ \vdots & \vdots\\ x_1 & \eta_{1(m)}\\ x_2 & \eta_{2(1)}\\ \vdots & \vdots\\ x_2 & \eta_{2(m)}\\ \vdots & \vdots\\ x_n & \eta_{n(1)}\\ \vdots & \vdots\\ x_n & \eta_{n(m)}\end{pmatrix},\mathbf{B} = \begin{pmatrix}\beta\\ \gamma\end{pmatrix},\ \mathbf{U} = \begin{pmatrix} U_{1(1)}\\ \vdots\\ U_{1(m)}\\ U_{2(1)}\\ \vdots\\ U_{2(m)}\\ \vdots\\ U_{n(1)}\\ \vdots\\ U_{n(m)}\end{pmatrix}$$

and U is the random error vector with

$$\begin{aligned}Var(\mathbf{Y}) = Var(\mathbf{U}) = \mathbf{V} = \gamma^2 diag\{&\tau_{1(1)},\ldots,\tau_{1(m)},\\ &\tau_{2(1)},\ldots,\tau_{2(m)},\ldots,\tau_{n(1)},\ldots,\tau_{n(m)}\}\end{aligned}$$

Lloyd [5] shows that, the best linear unbiased order statistics estimator of the parameter vector B in above system, takes the following form

$$\mathbf{B}^* = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \qquad (8)$$

with variance matrix

$$Var(\mathbf{B}^*) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\gamma^2$$

We recall that, members of a ranked set sample are independent. Then our estimator vector B^* will be simpler than Lloyd's estimator, which is shown in (8). Similar to the method used by Barreto and Barnett [1], we find the BLUE of parameter vector $B' = (\beta,\gamma)$ in the following theorem.

Theorem 3.2: Consider model (1) and $Y_i(j)$, i = 1,…,n, j = 1,…,m be a RSS from $S(\alpha,0,\gamma,0)$-distribution as indicated in Table 1. Then the BLUE of the vector parameter $B' = (\beta,\gamma)$ is given by

$$\tilde{\beta}_{BLUE} = \sum_{i=1}^{n}\frac{c^* + e^* x_i}{\Delta^*}\sum_{j\in J} Y_{i(j)}/\tau_{(j)} \qquad (9)$$

and

$$\tilde{\gamma}_{BLUE} = \sum_{i=1}^{n}\frac{g^*}{\Delta^*}\sum_{j\in J}\eta_{(j)}Y_{i(j)}/\tau_{(j)} \qquad (10)$$

where

$$c^* = -n(\sum_{i=1}^{n} x_i)(\sum_{j\in J}\frac{1}{\tau_{(j)}})(\sum_{j\in J}\frac{\eta_{(j)}^2}{\tau_{(j)}})$$

Table 2: Sample MAD of $\hat{\beta}_{BLUE}$ and $\tilde{\beta}_{BLUE}$ for different values of $\alpha \le 1$

α	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$MAD(\hat{\beta}_{BLUE})$	*	*	*	*	*	*	*	*	*
$MAD(\tilde{\beta}_{BLUE})$	0.947	0.872	0.794	0.662	0.654	0.638	0.601	0.504	0.446

Table 3: Sample MAD of $\hat{\beta}_{BLUE}$ and $\tilde{\beta}_{BLUE}$ for different values of $\alpha > 1$

α	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$MAD(\hat{\beta}_{BLUE})$	4.335	3.034	2.089	1.690	1.331	1.079	0.993	0.928
$MAD(\tilde{\beta}_{BLUE})$	2.376	1.777	1.453	1.342	1.033	0.711	0.701	0.505

$$e^* = n^2 \Big(\sum_{j \in J} \frac{1}{\tau_{(j)}}\Big)\Big(\sum_{j \in J} \frac{\eta_{(j)}^2}{\tau_{(j)}}\Big)$$

$$g^* = \Big(\sum_{j \in J} \frac{1}{\tau_{(j)}}\Big)^2 \Big[n \sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2\Big]$$

$$\Delta^* = n\Big(\sum_{j \in J} \frac{1}{\tau_{(j)}}\Big)^2 \Big(\sum_{j \in J} \frac{\eta_{(j)}^2}{\tau_{(j)}}\Big)\Big[n \sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2\Big]$$

the set J is defined in (4) and $\eta_{(j)}$ and $\tau_{(j)}$ are defined in (6) and (7), respectively.

Proof: The optimal L-estimator of $B' = (\beta,\gamma)$ takes the form (β^*, γ^*), where

$$\beta^* = \sum_{i=1}^{n} \sum_{j \in J} \nu_{ij} Y_{i(j)} \quad , \quad \gamma^* = \sum_{i=1}^{n} \sum_{j \in J} \theta_{ij} Y_{i(j)}$$

It can be show that, the values of v_{ij} and θ_{ij} take the similar forms as given in Barreto and Barnett [1], by replacing indices j = 1,…,m with $j \in J$. Therefore we can obtain $\tilde{\beta}_{BLUE}$ and $\tilde{\gamma}_{BLUE}$ as in (9) and (10), respectively.

We note that, $\tilde{\beta}_{BLUE}$ is defined for every $0<\alpha<2$ and its variance is finite.

COMPARISON OF ESTIMATORS

We compare $\hat{\beta}_{BLUE}$ in (3) and $\tilde{\beta}_{BLUE}$ in (9) through a simulation study. In this simulation study, we set $\beta = 3$, x = 0.1(0.1)1 and for each x we generate a set of Y_{ij}'s from (1) to obtain 3 ranked set sample that has finite second moment based on (4). From these values we calculate values of absolute deviation of $\hat{\beta}_{BLUE}$ and $\tilde{\beta}_{BLUE}$ for some values of $\alpha \le 1$ and $\alpha > 1$, i.e., $|\hat{\beta}_i - \beta|$ where $\hat{\beta}_i$ is one of $\hat{\beta}_{BLUE}$ or $\tilde{\beta}_{BLUE}$. Then we replicate this procedure 10000 times and calculate Mean Absolute Deviation (MAD) of $\hat{\beta}_{BLUE}$ and $\tilde{\beta}_{BLUE}$ from these iterations, i.e.,

$$MAD(\hat{\beta}_i) = \frac{1}{10000} \sum_i |\hat{\beta}_i - \beta|$$

which are shown in Tables 2 and 3. The S-PLUS functions for calculating MADs can be obtained from the first author.

Note that $MAD(\hat{\beta}_{BLUE})$ cannot be calculated for $0<\alpha\le 1$. From Table 3 we see that $MAD(\tilde{\beta}_{BLUE}) < MAD(\hat{\beta}_{BLUE})$ which shows the better performance of $\tilde{\beta}_{BLUE}$ with respect to $\hat{\beta}_{BLUE}$ for $\alpha>1$.

REFERENCES

1. Barreto, M.C.M. and V. Barnett, 1999. Best linear unbiased estimators for the simple linear regression model using ranked set sampling. Environmental and Ecological Statistics, 6: 119-133.
2. Blattberg, R. and T. Sargent, 1971. Regression with non-Gaussian stable disturbances: some sampling results. Econometrica, 39: 501-510.
3. Dell, T.R. and J.I. Clutter, 1972. Ranked set sampling theory with order statistics background. Biometrics, 28: 545-553.
4. Fama, E.F., 1965. The Behavior of Stock-Market Prices. Journal of Business, 38 (1): 34-105.
5. Lloyd, E.H., 1952. Generalized least-squares theorem. In Contributions to order statistics. Sarhan, A.E. and B.G. Greenberg (Eds.). John Wiley.

6. McIntyre, G.A., 1952. A method for unbiased selective sampling using ranked-set sampling. Australian Journal of Agricultural Research, 3: 385-390.
7. Mittnik, S. and S. Rachev, 1993. Modeling asset returns with alternative stable distributions. Economics Reviews, 12: 261-330.
8. Mohammadi, M. and A. Mohammadpour, 2011. Existence of order statistics moments of α-stable distributions. Preprint.
9. Ni Chuiv, N., B.K. Sinha and Z. Wu, 1998. Estimation of the location parameter of a Cauchy distribution using a ranked set sample. Journal of Applied Statistical Science, 3: 297-308.
10. Nolan, J.P., 2006. User manual for STABLE 4.0 S-Plus/R version. Robust Analysis, Inc. www.RobustAnalysis.com
11. Nolan, J.P., 2010. Stable distributions-models for heavy tailed data. Birkhauser. Boston.
12. Samorodnitsky, G. and M. Taqqu, 1994. Stable Non-Gaussian Random Processes. Chapman and Hall.