

Bootstrap Estimates of the Variances and Biases of Selected Robust Estimates of the Parameters of a Linear Model

DENNIS A. TAREPE

ORCID No. 0000-0001-7489-3356
da_tarepe@must.edu.ph

NESTOR C. RACHO

ORCID No. 0000-0002-1260-5627
ncracho@yahoo.com

RHEGIE M. CAGA-ANAN

ORCID No. 0000-0003-4323-9225
rhegiecagaanan@yahoo.com

Mindanao University of Science and Technology
Cagayan de Oro City, Philippines

ABSTRACT

When we want to apply the general linear model to a set of data, we have various methods to choose from to estimate the model parameters. The most popular one is the method of least-squares. This method, however, has weaknesses. Alternative regression methods are available which restrain the influence of outlying data points. The least-squares method of regression performs best if the population of errors is normally distributed. If there is a reason to believe that the distribution of errors may not be normal, then least-squares estimates and tests may be much less efficient than those provided by robust alternative methods such as the least-absolute-deviations (LAD), or M-estimators. Moreover, when the sample size is fixed, and no additional data can be obtained to support normal approximation, then the “bootstrap method” (Davison, 1997) may be used. The purpose of this paper is to explore the use of the bootstrap method in estimating the variances and biases of selected robust estimates of the parameters of a

linear model. Specifically, three robust estimators are considered. Generally, the simulation results reveal that the least-squares method still performs quite well for slight contamination while the robust methods perform better for moderate and high contaminations.

Keywords: Linear models, Robust alternatives, Bootstrap, Jackknife, Simulation

INTRODUCTION

Suppose we find ourselves in the following common data-analytic situation: a random sample x_1, x_2, \dots, x_n from an unknown probability distribution F is observed and we wish to estimate a parameter of interest θ on the basis of x_1, x_2, \dots, x_n . Furthermore, as a necessary pre-requisite, suppose we also wish to approximate the distribution of the statistic, call it T . This distribution can then be used in testing of a hypothesis about the parameter θ , estimating the variance and bias of T , or estimating the various moments of T .

However, suppose that the sample size n is fixed, and no additional data can be obtained to support the use of normal approximation. If the interest rests on estimating the variance, bias, moments, or functions of these for the statistic T , then the “bootstrap method” (Davison, 1997) may be used. In this method, the empirical distribution of the sample x_1, x_2, \dots, x_n is constructed, say F_n , and variate from this distribution is generated, then estimates of the variance or moments of T can be obtained from the newly generated variates from F_n .

The bootstrap is a recently developed technique for making certain kinds of statistical inferences. The bootstrap is a data-based simulation method for statistical inference, which can be used to produce inferences. It was introduced in 1979 as a computer-based method for estimating the standard error of T . It enjoys the advantage of being completely automatic. The bootstrap estimate of the standard error requires no theoretical calculations and is available no matter how mathematically complicated the estimator T may be.

The bootstrap methods had been used in different situations, e.g. estimation of location and scale parameters, estimation of the parameters of a linear model, and others, to approximate the variance and reduce the biases of the estimators in each case. An extensive review of the state-of-the-art in bootstrap methods can be found in Davison *et al.* (1997), and others.

OBJECTIVES OF THE STUDY

The purpose of this paper is to explore the use of bootstrap statistics in estimating the variances and biases of selected robust estimates of the parameters of a linear model, specifically, the variance and bias of $\hat{\beta}$ obtained from the following robust alternatives: 1) median approach (Padua, 1987); 2) the generalized bootstrap estimate of β ; and 3) and the minimum absolute deviation method (Huber, 2009) are estimated by the bootstrap methods.

FRAMEWORK

Basic Concepts

The Linear Model

The linear model is given by:

$$Y = X\beta + \varepsilon, \tag{1}$$

where Y is an $n \times 1$ vector of random observable quantities, X is an $n \times p$ matrix of constants assumed to be of full rank p , β is a $p \times 1$ vector of unknown parameters, ε is an $n \times 1$ vector of unobservable random errors. We assume that $E(\varepsilon) = 0$ and: (a) $\text{var}(\varepsilon) = 0$ and: (a) $\text{var}(\varepsilon) = \sigma^2 I$ or (b) $\text{var}(\varepsilon) = \Sigma$, a positive-definite matrix.

An intuitive approach to estimating β is to find the vector which minimizes the sum of squares of errors (SSE):

$$SSE = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta) \tag{2}$$

where ε' denotes the transpose of ε . The least-squares estimate of β is:

$$\hat{\beta} = (X'X)^{-1} X'Y \tag{3}$$

$$RSS = (Y - \hat{Y})'(Y - \hat{Y}) \tag{4}$$

where RSS means “residual sum of squares.” The expected value and variance of $\hat{\beta}$ are given by:

$$E(\hat{\beta}) = \beta \tag{5}$$

$$\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

so that $\hat{\beta}$ is an unbiased estimator of β . Note that $\hat{\beta}$ is a linear function of Y . The theorem of Gauss and Markov (Graybill, 1976) states that among all linear unbiased estimates of β , the least-squares estimates have the smallest variance. The variance of $\hat{\beta}$ from (5) contains σ^2 which may be estimated from:

$$\sigma^2 = \frac{RSS}{n - p} \tag{6}$$

It can be shown that if ε is multivariate normal, then δ^2 has a chi-squared distribution with $E(\delta^2) = \sigma^2$.

When $\text{var}(\varepsilon) = \Sigma$ is not diagonal, then we can find an orthogonal matrix P such that $P'\Sigma P = D$ is diagonal. The transformation $Z = PY$ yields:

$$Z = PX\beta + P\varepsilon = X'\beta + \varepsilon' \quad (7)$$

where $\text{var}(\varepsilon) = P'\Sigma P = D$ is diagonal. The resulting estimate for β is:

$$\hat{\beta} = (X'X)^{-1}X'Z = (X'X)^{-1}X'Y \quad (8)$$

as before. However, the fact that $\text{var}(\varepsilon^*) = D \neq \sigma^2 I$ can be exploited to obtain better estimates of $Z = D^{1/2}Y$ as follows: Let β^* , where $D = D^{1/2} \cdot D^{1/2}$. Then,

$$\hat{\beta}^* = (X'D^{-1}X)^{-1}X'D^{-1}Y \quad (9)$$

is called the weighted least-squares estimate of β . The effect of weighting by D is to downplay the effects of points, which correspond to these with larger variances.

Robust Alternatives

The recognition that the structure of the covariance matrix for the error term may not be of the form led statisticians to the weighted least-squares approach. The idea was to down weight the effects of those points whose errors have larger variances. In general, outliers in data sets can grossly affect the least-squares estimates of $\sigma^2 I$, hence, the need to develop robust alternatives to the least-squares procedure.

Since its discovery almost 200 years ago, least squares have been the most popular method of regression analysis. However, over the last two or three decades, interests in other methods have increased significantly. This is due to newly known deficiencies in the least-squares method and the significant increase in computational efficiency of modern machines. A number research articles on alternative approaches to regression analysis have been published since then.

Recently, development of these approaches continues, and further research and experience may lead to modifications and improvements. However, enough knowledge and experience have already been gained to be able to say that currently proposed alternative methods give sound results, have worthwhile advantages over the least-squares methods and can be recommended for practical use.

The alternative methods are chosen because they represent different approaches to regression analysis, and they have received considerable attention in the statistics field.

M-Estimators

Huber (2009) considered the estimate of β obtained by minimizing:

$$Q = \sum_{i=1}^n |y_i - x_i' \beta| \tag{10}$$

which he calls the Least-Absolute-Deviation (LAD) estimate of β , $\hat{\beta}_{LAD}$. Although the estimate cannot be expressed in closed form, this can be computed using standard linear programming techniques. The variance-covariance matrix of the estimate naturally is not expressible in closed form. This makes for an ideal situation where the bootstrap approach can be applied.

More generally, consider minimizing a convex function of the residuals given by:

$$Q = \sum_{i=1}^n \rho(y_i - x_i' \beta) \tag{11}$$

which Huber (2009) called the M-Estimators of β .

Estimates Based on Quantiles

There are other robust methods for selecting the parameters of a linear model. Padua (1987), for example, considered the following procedure:

Consider the augmented data matrix $[Y : X]$ and consider all possible $\binom{n}{p}$

submatrices $[Y_I : X_I]$, $I = 1, 2, \dots, \binom{n}{p}$ least-squares estimates of β , $\{\hat{\beta}_I\}$, may be

computed. The proposed estimate of β is given by:

$$\tilde{\beta}_{PAD} = \text{median}\{\hat{\beta}_I\} \text{ componentwise} \tag{12}$$

This estimate was found to have a high breakdown point and is robust. The variance of this estimator cannot be expressed in closed form, although asymptotic results revealed that under suitable conditions, the estimate converges in distribution to a multivariate normal distribution.

L-Estimates

The estimation of the variances and biases of $\hat{\beta}_{LAD}$ and $\tilde{\beta}_{PAD}$ for a sample of size n with p parameters to estimate is of interest from the point of view of practicability. Bootstrap estimates of these variances are provided in this paper. Moreover, bootstrap estimates of the biases are similarly computed in order to assess the relative sizes of the mean-squared errors (MSE) for fixed n .

Statistics which are a linear combination of order statistics have similarly been tried out in the past. The L-estimates of θ in a location parameter problem can be expressed as:

$$L = \sum_{i=1}^n w_i x_{(i)} \quad (13)$$

where w_i are weights and $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ (Stigler, 1973). In the case of the linear model, Rousseeuw (1987) and others considered the following procedure:

Delete one observation and compute the least-squares estimate of β from the remaining $(n - 1)$ observations. Using this estimate of β estimate $Y_{(i)}$, call it $\hat{Y}_{(i)}$, the fitted value of Y for the deleted case. Do this for all the other observations. Discard all the Y_i 's whose fitted residuals exceed a pre-defined value. Fit a least-squares line on the points that remain.

A serious objection to this procedure arises when $m > 1$ outliers exist. The least-squares estimate fitted on the $(n-1)$ remaining observations will be grossly affected by the outliers present even if one of them is removed. Consequently, the results may still turn out to be poor estimates of the true β .

Consequently, Racho (1999) considered a better test for outliers in which $k = [n\alpha]$ observations are discarded as outliers. His procedure may be described as follows:

Let β^* be the median estimate of Maritz (1979) (other robust estimates of β may also be considered). Compute the n fitted residuals:

$$r_i = y_i - x_i \beta^* \quad i = 1, 2, \dots, n \quad (14)$$

The ordered values $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(n)}$ are considered. Delete the $[n\alpha]$, $0 < \alpha < 1$, lowest and highest observations corresponding to these residuals. Define a new set of fitted residuals:

$$r_i^* = \begin{cases} r_i, & \text{if } r_{(a)} \leq r_i \leq r_{(1-a)} \\ 0, & \text{else.} \end{cases} \quad (15)$$

On the remaining $n(1-\alpha)$ observations compute the least-squares estimate $\tilde{\beta}$. The estimator is called the trimmed least-squares $n(1-2\alpha)$ estimate.

As in other robust procedure, the fixed sample variance of the estimator cannot be expressed in closed form.

METHODOLOGY

The Jackknife Method

Quenouille (1949) discovered a nonparametric estimate of bias which he later called the jackknife. The idea is as follows: Let x_1, x_2, \dots, x_n be iid random quantities coming from a distribution $F(x; \theta)$ and these estimate $\hat{\theta}(x_1, x_2, \dots, x_n)$ is obtained. If $E(\hat{\theta}) \neq \theta$, then the quantity $B = E(\hat{\theta}) - \theta$ is called its bias. Quenouille's (1949) method is based on sequentially deleting points x_i and recomputing $\hat{\theta}$. The removing point x_i from the data set provides a mass of $\frac{1}{n-1}$ to each of the remaining points. Let $\hat{\theta}_{(i)}$ be the estimator $\hat{\theta}$ computed with the i th point deleted. Quenouille's estimate of the bias is:

$$\hat{B}ias = (n-1)(\hat{\theta}_{(i)} - \hat{\theta}) \tag{16}$$

where $\hat{\theta}_{(i)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}$. The "bias corrected" estimate of θ is then:

$$\tilde{\theta} = \hat{\theta} - \hat{B}ias = n\hat{\theta} - (n-1)\hat{\theta}_{(i)} \tag{17}$$

The estimator $\hat{\theta} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$ of the variance of the distribution can be "bias corrected" in this manner. Following Quenouille's rule, we find that:

$$\hat{B}ias = \frac{1}{n(n-1)} \sum_{i=1}^n (x_{(i)} - \bar{x})^2 \tag{18}$$

yielding;

$$\tilde{\theta} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \tag{19}$$

which we know is an unbiased estimate of σ^2 .

On the other hand, Tukey (1958) recommended that the variance should be computed as follows:

$$\hat{v}ar = \frac{n-1}{n} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(i)})^2 \tag{20}$$

where $\hat{\theta}_{(i)}$ is given by (16).

It is interesting to apply the bootstrap method for estimating the variances and biases of the comprehensive estimator of β , $\tilde{\beta}_{PAD}$, found in (3). The algorithm is as follows:

Algorithm 1

1. Compute $\tilde{\beta}_{PAD}$ using the full data set.
2. Sample with replacement n rows from the original data set and compute

the least-squares estimate of β , $\hat{\beta}_{(i)}$, $i=1, 2, \dots, n$. Get the median component wise.

3. Repeat the process B times (the number of bootstrap samples).

$$4. \text{ Let } \tilde{\beta}_{(c)} = \frac{1}{n} \sum_{i=1}^n \tilde{\beta}_{(i)} .$$

$$5. \hat{\text{Bias}} = (n-1)[\tilde{\beta}_{(c)} - \tilde{\beta}].$$

$$6. \hat{\text{var}} = \frac{n-1}{n} \sum_{i=1}^n (\tilde{\beta}_{(i)} - \tilde{\beta}_{(c)})(\tilde{\beta}_{(i)} - \tilde{\beta}_{(c)})' .$$

The “bias corrected” estimator of β is then given by:

$$\hat{\beta}_{\text{med}} = n\tilde{\beta} - (n-1)\tilde{\beta}_{(c)} . \quad (21)$$

A similar algorithm is developed for estimating the bias and variance of the LAD estimator β , $\hat{\beta}_{\text{LAD}}$, by the bootstrap.

A Generalized Median Estimate Using the Bootstrap Method

A generalization of the method of Padua (1987) is developed using the idea of bootstrapping. From the original set of n observations (x_i, y_i) sample with replacement n observations and compute $\hat{\beta}_{(i)}$, the least-squares estimate of β . Do this B times, where B is the number of bootstrap samples. Take $\tilde{\beta}_{\text{GME}} = \text{median}_{1 \leq i \leq B} \{\hat{\beta}_{(i)}\}$

The basic purpose of the estimator $\tilde{\beta}_{\text{GME}}$ is to guard against the effects of gross outliers to which the least-squares estimate, $\hat{\beta}_{\text{LSE}}$, is very sensitive to.

RESULTS AND DISCUSSION

The Experimental Set-up

We consider the linear model (1) with. For each p , sample sizes are used. The matrix X and the vector Y are generated as iid random variables from and will remain fixed throughout the experiment.

The errors are generated from the Tukey's (1972) contaminated normal model given by:

$$\varepsilon_i \text{ iid } (1-\lambda)N(0,1) + \lambda N(0,16), \quad (22)$$

where, respectively. The first two values represent “slight contamination”, the third represents “moderate contamination” while the last value of λ represents “high contamination”. Values of λ greater than 20% indicate “massive contamination”. In this case, even the use of robust regression may be dubious.

The models used are:

$$\text{Model 1: } y_i = 1.52 + 0.98x_1 + \varepsilon_i \tag{23}$$

$$\text{Model 2: } y_i = 1.52 + 0.98x_1 - 1.12x_2 + \varepsilon_i$$

$$\text{Model 3: } y_i = 1.52 + 0.98x_1 - 1.12x_2 + 2.05x_3 + \varepsilon_i$$

$$\text{Model 4: } y_i = 1.52 + 0.98x_1 - 1.12x_2 + 2.05x_3 - 0.75x_4 + \varepsilon_i$$

The following estimators are evaluated from the data sets (1) usual least-squares estimator, $\hat{\beta}_{LSE}$, (2) median estimator (Padua, 1987) for $h = p$, $\tilde{\beta}_{PAD}$, (3) median estimator for $h = n - 1$ or the median bootstrap estimator $\tilde{\beta}_{GME}$, $i = 1, 2, \dots, n$, and (4) LAD estimator, $\hat{\beta}_{LAD}$.

Bootstrap estimates of the biases and variances of $\tilde{\beta}_{PAD}$ and $\hat{\beta}_{LAD}$ are computed using

Algorithm 1. Bootstrap estimates of the biases and variances of $\tilde{\beta}_{GME}$ are computed using

Algorithm 2. For each run, a bootstrap sample is used.

Algorithm 2

1. Compute the n least-squares estimates $\hat{\beta}_{(i)}$, $i = 1, 2, \dots, n$ from the data $j \in \{1, 2, \dots, i-1, i+1, \dots, n\}$ set by sampling with replacement.
2. Sample with replacement $\hat{\beta}_{(i)}^j$, $j = 1, 2, \dots, n$ and compute $\tilde{\beta}^j = \text{median}\{\hat{\beta}_{(i)}^j\}$ where or the median of the remaining $\hat{\beta}_{(i)}$.
3. Let $\tilde{\beta}_{(i)}^j = \frac{1}{n} \sum_{j=1}^n \tilde{\beta}^j$.
4. Compute: $\text{Bias} = (n-1)[\tilde{\beta}_{(i)}^j - \tilde{\beta}^j]$, and $\text{var} = \frac{n-1}{n} \sum (\tilde{\beta}_{(i)}^j - \tilde{\beta}^j)(\tilde{\beta}_{(i)}^j - \tilde{\beta}^j)'$.

Simulation Results

For the simulation process, the following notations were used:

1. $\hat{\beta}_{LSE}$, the least-squares estimator
2. $\tilde{\beta}_{PAD}$, the median estimator of Padua (1987) for
3. $\tilde{\beta}_{GME}$, the median bootstrap estimator
4. $\hat{\beta}_{LAD}$, the least-absolute-deviations estimator (Huber, 1980)

The tables below show the estimates of β (Tables 1-4), biases, variances, and mean-squared errors of the four estimators of β (Tables 5-8 for $p = 2$, and $\lambda = 0.01$, $\lambda = 0.05$, (slight contamination) $\lambda = 0.01$ (moderate contamination), and $\lambda = 0.02$ (high contamination). Simulation results for $p = 3, 4, 5$ are also discussed below.

Table 1. Estimates of β for $p=2$, $\lambda=0.01$

| Sample Size | $\lambda = 0.01$ | | | |
|-------------|---------------------|-----------------------|-----------------------|---------------------|
| | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | 1.7524 | 1.3075 | 1.7860 | 1.4240 |
| | 1.2115 | 1.0246 | 1.2172 | 1.0409 |
| n = 20 | 1.6886 | 1.8538 | 1.6852 | 1.7138 |
| | 0.8186 | 0.9180 | 0.8148 | 1.2173 |
| n = 30 | 1.4123 | 1.3442 | 1.4145 | 1.4289 |
| | 0.9850 | 0.9228 | 0.9858 | 0.9798 |

Table 2. Estimates of β for $p = 2$, $\lambda = 0.05$

| Sample Size | $\lambda = 0.05$ | | | |
|-------------|---------------------|-----------------------|-----------------------|---------------------|
| | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | 0.7958 | 0.8350 | 0.8018 | 0.7349 |
| | 2.1224 | 1.9612 | 2.1361 | 2.0446 |
| n = 20 | 1.5635 | 1.7063 | 1.5705 | 1.5540 |
| | 0.4522 | 0.4338 | 0.4505 | 0.4473 |
| n = 30 | 1.2943 | 1.2449 | 1.3002 | 1.2863 |
| | 0.7422 | 0.6599 | 0.7427 | 0.7643 |

Table 3. Estimates of β for $p=2$, $\lambda= 0.10$

| Sample Size | $\lambda = 0.10$ | | | |
|-------------|---------------------|-----------------------|-----------------------|---------------------|
| | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | 1.5655 | 2.0653 | 1.6104 | 1.5438 |
| | 1.2868 | 1.3559 | 1.3454 | 1.2495 |
| n = 20 | 1.3830 | 1.9826 | 1.3784 | 1.3922 |
| | 1.1647 | 1.1141 | 1.1661 | 1.1680 |
| n = 30 | 1.5704 | 1.1230 | 1.5545 | 1.6413 |
| | 1.0197 | 1.2439 | 1.0057 | 1.0918 |

Table 4. Estimates of β for $p=2$, $\lambda = 0.20$

| Sample Size | $\lambda = 0.20$ | | | |
|-------------|---------------------|-----------------------|-----------------------|---------------------|
| | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | -0.4326 | -1.1384 | -0.3365 | -0.5578 |
| | -2.9451 | -1.6671 | -2.8682 | -2.7618 |
| n = 20 | 0.8466 | 1.5762 | 0.8224 | 0.9130 |
| | -0.3340 | -0.4584 | -0.3315 | -0.4147 |
| n = 30 | 1.4682 | 1.7646 | 1.4552 | 1.4457 |
| | 0.4229 | 0.3639 | 0.4259 | 0.4173 |

Table 5. Variances, Biases, and MSE's for $p=2, \lambda=0.01$

| Sample Size | | $\lambda = 0.01$ | | | |
|-------------|--------|---------------------|-----------------------|-----------------------|---------------------|
| | | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | Bias | | -0.3043 | 0.0000 | 1.2196 |
| | | | | -0.4596 | 0.0000 |
| | var | 0.0975 | 0.0639 | 0.0001 | 0.1524 |
| | | 0.1328 | 0.1211 | 0.0006 | 0.1752 |
| MSE | 0.1152 | 0.2444 | 0.0004 | 1.2759 | |
| n = 20 | Bias | | -0.6444 | 0.0000 | 0.4667 |
| | | | | -1.7693 | 0.0000 |
| | var | 0.0443 | 0.0209 | 0.0000 | 0.0248 |
| | | 0.0413 | 0.1183 | 0.0000 | 0.1141 |
| MSE | 0.0428 | 1.8424 | 0.0000 | 0.4321 | |
| n = 30 | Bias | | -2.6923 | 0.0000 | 0.0302 |
| | | | | -0.0100 | 0.0000 |
| | var | 0.0377 | 0.0310 | 0.0000 | 0.0328 |
| | | 0.0194 | 0.0465 | 0.0000 | 0.0309 |
| MSE | 0.0286 | 0.3663 | 0.0000 | 0.0328 | |

Table 6. Variances, Biases, and MSE's for $p=2, \lambda=0.05$

| Sample Size | | $\lambda = 0.05$ | | | |
|-------------|--------|---------------------|-----------------------|-----------------------|---------------------|
| | | $\hat{\beta}_{LSE}$ | $\tilde{\beta}_{PAD}$ | $\tilde{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | Bias | | -2.1909 | 0.0000 | 0.0294 |
| | | | | -3.2429 | 0.0000 |
| | var | 0.1994 | 0.3398 | 0.0000 | 0.1516 |
| | | 0.2928 | 0.3396 | 0.0010 | 0.2149 |
| MSE | 0.2461 | 0.7998 | 0.0005 | 0.2237 | |
| n = 20 | Bias | | -1.4719 | 0.0000 | -0.0020 |
| | | | | -2.7188 | 0.0000 |
| | var | 0.0496 | 0.1423 | 0.0000 | 0.0826 |
| | | 0.0544 | 0.0889 | 0.0000 | 0.0978 |
| MSE | 0.0520 | 0.4895 | 0.0000 | 0.0904 | |
| n = 30 | Bias | | -0.9887 | 0.0000 | 0.0140 |
| | | | | -2.4785 | 0.0000 |
| | var | 0.0495 | 0.0798 | 0.0000 | 0.0605 |
| | | 0.0376 | 0.0829 | 0.0000 | 0.0728 |
| MSE | 0.0436 | 0.3642 | 0.0000 | 0.0669 | |

Table 7. Variances, Biases, and MSE's for $p=2$, $\lambda=0.10$

| Sample Size | | $\lambda = 0.10$ | | | |
|-------------|---------------|---------------------|---------------------|---------------------|---------------------|
| | | $\hat{\beta}_{LSE}$ | $\hat{\beta}_{PAD}$ | $\hat{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | B \hat{a} s | | -5.2961 | 0.0000 | 1.5622 |
| | | | -5.8097 | 0.0000 | -3.3213 |
| | v \hat{a} r | 0.2822 | 1.2007 | 0.0086 | 1.7931 |
| | | 0.3961 | 1.2962 | 0.0001 | 8.1472 |
| MSE | 0.3392 | 3.2149 | 0.0044 | 1.1706 | |
| n = 20 | B \hat{a} s | | 0.4902 | 0.0000 | 0.0597 |
| | | | -4.1297 | 0.0000 | -0.0617 |
| | v \hat{a} r | 0.1239 | 0.4332 | 0.0023 | 0.1383 |
| | | 0.1265 | 0.1892 | 0.0006 | 0.1493 |
| MSE | 0.1252 | 0.8959 | 0.0015 | 0.1475 | |
| n = 30 | B \hat{a} s | | -4.0052 | 0.0000 | -0.2622 |
| | | | 1.8527 | 0.0000 | -0.1859 |
| | v \hat{a} r | 0.1003 | 0.3704 | 0.0017 | 0.1076 |
| | | 0.1115 | 0.1452 | 0.0001 | 0.1411 |
| MSE | 0.1059 | 0.9995 | 0.0009 | 0.1760 | |

Table 8. Variances, Biases, and MSE's for $p=2$, $\lambda= 0.20$

| Sample Size | | $\lambda = 0.20$ | | | |
|-------------|---------------|---------------------|---------------------|---------------------|---------------------|
| | | $\hat{\beta}_{LSE}$ | $\hat{\beta}_{PAD}$ | $\hat{\beta}_{GME}$ | $\hat{\beta}_{LAD}$ |
| n = 10 | B \hat{a} s | | 4.0969 | 0.0000 | -1.0083 |
| | | | -0.8274 | 0.0000 | 1.0417 |
| | v \hat{a} r | 0.7950 | 1.0107 | 0.0191 | 1.2765 |
| | | 2.1746 | 2.1459 | 0.0698 | 2.8415 |
| MSE | 1.4848 | 10.3129 | 0.0445 | 3.1099 | |
| n = 20 | B \hat{a} s | | -4.4062 | 0.0000 | 0.0131 |
| | | | 5.5462 | 0.0000 | -0.0746 |
| | v \hat{a} r | 0.4365 | 0.7302 | 0.0044 | 0.4271 |
| | | 0.4682 | 0.4910 | 0.0001 | 0.7433 |
| MSE | 0.4524 | 2.5698 | 0.0023 | 0.5881 | |
| n = 30 | B \hat{a} s | | -3.4811 | 0.0000 | 0.0077 |
| | | | -1.5685 | 0.0000 | 0.0126 |
| | v \hat{a} r | 0.3347 | 0.6696 | 0.0000 | 0.3568 |
| | | 0.4581 | 0.4411 | 0.0197 | 0.3660 |
| MSE | 0.3964 | 0.7844 | 0.0099 | 0.3615 | |

For the least-squares estimates $\hat{\beta}_{LSE}$, it can be shown that this estimate is unbiased for β that is $E(\hat{\beta}) = \beta$. In the simulation results above, the bias of $\hat{\beta}$ were not reflected since the algorithm allows only the generation of one data set. Also, if ε is multivariate normal with mean 0 and variance $\sigma^2 I$ then $\hat{\beta}_{LSE}$ is the uniformly minimum variance unbiased estimator (UMVUE) for β (Graybill, 1976). It is evident from the tables above (and consequently, the rest of the simulation results) that are generally speaking when the contamination is slight, i.e. when $\lambda = 0.01$ and $\lambda = 0.05$, the variances and MSE's of $\hat{\beta}_{LSE}$ are comparatively lower.

However, when the contamination is moderate ($\lambda = 0.10$) or high ($\lambda = 0.20$), $\hat{\beta}_{LSE}$ increases its biases and variances, which in turn increases its MSE's. The least-squares estimates of the parameters of a linear model are known to be sensitive to the effects of extreme or outlying observations. Outliers can grossly influence the estimated values of the parameters as well as their variances to the extent that a single outlier can distort these values away from their true values significantly (Huber, 1984; Stigler, 1973).

The three robust estimates of β are estimates which are relatively insensitive to the effects of gross outliers. They are "resistant" to the influence of outliers. The use of robust estimators may mitigate the effects of outlying observations. However, their properties are not generally well-understood, especially for a fixed sample size n . It is for this reason that robust alternatives to the usual least-squares method have not been properly used in practice (Efron, 1982).

In Padua's (1987) procedure, $\hat{\beta}_{PAD}$ is computed as the component-wise median of the least-squares estimates of the $\binom{n}{p}$ blocks. Larger block sizes allowed for greater flexibility in estimating the variance of the residuals and also provided greater protection against outliers. Thus, the empirical results showed that as we increase the block size p , the performance of the $\hat{\beta}_{PAD}$ is greatly affected.

The median bootstrap estimator, $\hat{\beta}_{GME}$, has the best performance of all the estimators considered in this paper. This can be attributed to the fact that the median bootstrap estimator has breakdown roughly 50%. The breakdown of an estimator is defined as the smallest proportion of the data that can have an arbitrarily large effect on its value. The high breakdown is good, with 50% being the largest value that makes sense.

The least-absolute-deviations estimator, $\hat{\beta}_{LAD}$, generally produces larger MSE's than the least-squares. However, it is important to note that the M-estimators have been found to be unaffected by departures from normality (Efron and Tibshirani, 1993). These results are illustrated in the bar graphs below.

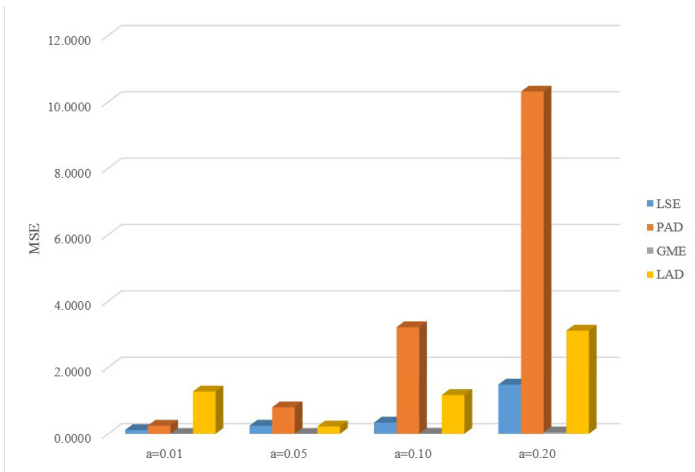


Figure 1. MSE's for p=2, n=10.

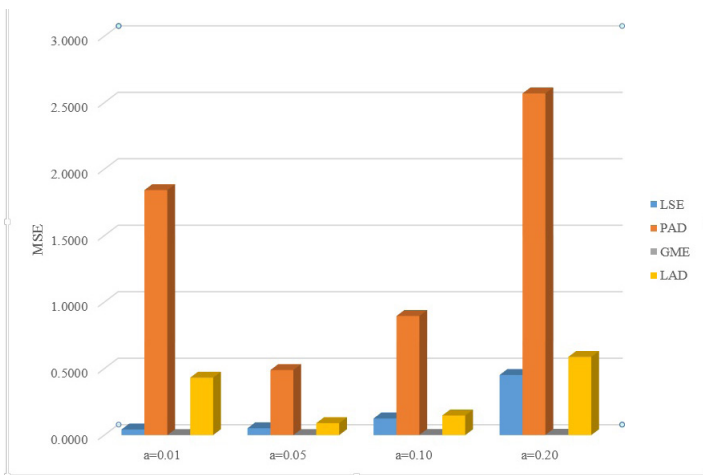


Figure 2. MSE's for p=2, n=20.

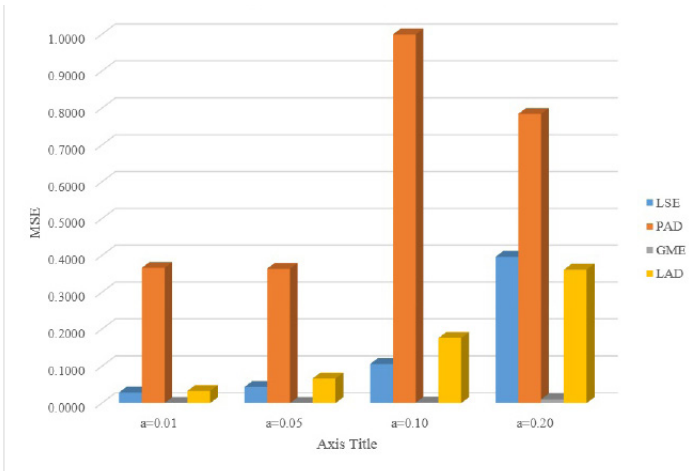


Figure 3. MSE's for p=2, n=30.

CONCLUSIONS

The supporting theories, concepts, and empirical results of this study led to the following conclusions:

1. The least-squares estimator, $\hat{\beta}_{LSE}$, performs best if the population error has a normal distribution. It can likewise perform equally well when contamination of the data is only slight;
2. Among the three robust alternatives considered in this paper, the median bootstrap estimator, $\hat{\beta}_{MB}$, has the best performance, in general;
3. The median estimator of Padua (1987), $\hat{\beta}_{MP}$, has the largest computational complexity, which is of order $n \times \binom{n}{p}$;
4. The bootstrap method could be used to compute the bias and variance of an estimator to make it less biased;
5. In general, the bootstrap method can provide approximations of almost any statistics, without taking into account the form of their underlying distribution.

RECOMMENDATIONS

It may be worthwhile to consider the following recommendations for further research work related to the existing study.

1. Consider good outlier detection techniques before applying the bootstrap method.
2. For the case of a fixed sample size n , use robust alternative that performs equally well with others but which require lesser computing time.

ACKNOWLEDGMENT

The authors would like to thank the Mindanao University of Science and Technology (MUST) administration for the financial assistance provided in the conduct of this research project.

LITERATURE CITED

- Davison, A.C. & Hinkley, D.V.
1997 Bootstrap Methods and their Application. Cambridge University Press. United Kingdom.
- Efron, B.
1982 The Jackknife, the Bootstrap and Other Resampling Plans. SIAM. [2.4, 2.6, 3.5, 3.7].
- Efron, B. & Tibshirani, R. J.
1993 An Introduction to the Bootstrap. Chapman & Hall. New York.
- Graybill, F. A.
1976 Theory and Application of the Linear Model. Wadsworth Publishing Company, Inc. California.
- Huber, P. J. & Ronchetti, E.M.
2009 Robust Statistics. 2nd ed. John Wiley & Sons, Inc. New Jersey.

Maritz, J.S. & Jarrett, R.

1978 A Note on Estimating the Variance of the Sample Median. *J. Ann. Stat. Assoc.*, Vol. 73, 194-196.

Padua, R.N.

1987 Median Estimates of Regression Parameters. *Proceedings of National Research Council of the Philippines*, Vol. 7, 212-218.

Quenouille, M. H.

1949 Approximate Tests of Correlation in Time-Series 3. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 45, No. 03, pp. 483-484). Cambridge University Press.

Rousseeuw, P. J.

1987 Least Median of Squares Regression. *J. Ann. Stat. Assoc.* 79, 871-880 [7.6, 7.8].

Rousseeuw, P. J. & Leroy, A.M.

2003 *Robust Regression and Outlier Detection*. Wiley, New York

Stigler, S.M.

1973 The Asymptotic Distribution of the Trimmed Mean. *The Annals of Statistics*, 472-477.

Tukey, J.W.

1960 A Survey of Sampling from Contaminated Distributions (in *Contributions to Probability and Statistics*, I. Olkin, et al.). 448-485 [1.2, 2.3, 4.4].