

PARAMETRIC AND NON-PARAMETRIC BOOTSTRAP: A SIMULATION STUDY FOR A LINEAR REGRESSION WITH RESIDUALS FROM A MIXTURE OF LAPLACE DISTRIBUTIONS

Melfi Alrasheedi

School of Business, King Faisal University, Saudi Arabia

Abstract

In this simulation study, we compared ordinary least squares (OLS), weighted least squares (WLS), and three bootstrap versions (resampling of data points, resampling residuals, generating new residuals from Laplace distributions) for a linear regression with independent residuals from a mixture of two Laplace distributions. Leverage points were removed from the data, more outliers were added, and knowledge about the two Laplace distributions was omitted. For the data set with more extreme outliers, all methods showed problems with the coverage probability of the confidence intervals for parameter estimation, but bootstrap method 1 was clearly more robust. For the base data set, there was no difference between bootstrap and WLS, similarly to the data set with some leverage points removed. Without knowledge of the two Laplace distributions, bootstrap method 2 performed best in that standard errors of the parameter estimates was lower and confidence intervals shorter. This result suggests that, depending on the sample kurtosis compared to distribution kurtosis, bootstrap method 2 (non-parametric) or 3 (parametric) is better.

Keywords: Parametric and non-parametric bootstrap, Laplace distribution, Weighted least squares, Kurtosis, Heteroscedasticity, Resampling

Introduction

Several methods can be used for linear regression analysis, including ordinary least squares (which is said to perform poorly in the case of heteroscedastic errors), weighted least squares (down-weights data points with a high residual variance in order to address heteroscedasticity), and bootstrap (which depends on fewer assumptions on the residuals and can be used with heteroscedasticity and outliers). Within the bootstrap methods, there are

variations in how to resample: the data points themselves, the residuals, or generating new residuals from a given distribution. This simulation study aimed to investigate the appropriateness of each method for a very specific situation: a linear regression line with additive independent residuals from a mixture of two Laplace distributions with mean zero and different variances (Rao, et al 1999, Aitken 1935) . We demonstrate that one method did not perform better than others, but that the optimal choice depends on the specific data.

Least squares estimation with non-constant error variance

For data analysis the relationship between the examined variables representing various aspects of the subjects of study and measuring physical values is often required knowledge. If we consider the case in which the goal is to define the linear dependence of two parameters X and Y , the difficulty lies in evaluating the unknown coefficients β_0, β_1 :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Equation (1) is often referred to as the linear regression model. The ordinary least squares (OLS) fitting procedure can be used to estimate unknown linear regression coefficients (Wolberg 2005). According to this approach, these parameters are obtained via function F :

$$F(b_0, b_1) = \sum_i (Y_i - b_1 X_i - b_0)^2 \quad (2)$$

The minimum of (2) corresponds to necessary linear regression parameters, which can be obtained using:

$$\frac{\partial F}{\partial b_1} = 0, \quad \frac{\partial F}{\partial b_0} = 0, \text{ resulting in}$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}, \text{ where } \bar{y} \text{ is the mean value of } Y \text{ and } \bar{x} \text{ is the mean value of } X$$

The geometric description of the OLS method is very simple and straightforward. The obtained fitted line $\hat{Y} = b_0 + b_1 X$ is known as the least squares regression line.

If the errors (ε) in the linear regression model (1) are expected to be zero, are uncorrelated, and have equal variances (σ^2), which is a constant that is independent of X , then the Gauss-Markov theorem states that the OLS estimator is the best linear unbiased estimator (BLUE). Here, 'best' indicates that which gives the lowest possible mean squared error of the estimate.

However, in a case for which errors (ε) have unequal variances (σ^2), the simple least squares method has many drawbacks. For example, it is not efficient. To account for such a situation, a linear regression model with unequal variances is introduced. This model has the same form as (1), but for estimating the coefficients of this model, the weighted least squares (WLS) scheme is used (Rao et al 1999) . According to this approach, the following method, which is a modification of (2), is used.

$$F(b_0, b_1) = \sum_i w_i (Y_i - b_1 X_i - b_0)^2$$

$$w_i = \frac{1}{\sigma_i^2}$$

The weighting coefficients are defined as reciprocals of the variances σ_i^2 for each data point. In this manner, the contribution of more noisy data to the overall estimation scheme is reduced. Aitken (1935) showed that using these weights, the estimator is again the best linear unbiased estimator (BLUE). Therefore, in a case for which the error variances are known, the WLS procedure is straightforward; otherwise the variances must be estimated first.

One method for estimating regression coefficients and their confidence intervals is to apply the well-known bootstrap technique (Efron and Tibshirani 1993, Amiri et al 2008, Zhu and Jing 2010, Efron 1987). This approach is based on the general idea of resampling from the given data set to generate additional samples for estimating desired quantities. Depending on how these additional samples are obtained, the bootstrap technique is divided into two types: parametric and nonparametric (Benton and Krishnamoorthy 2002). We will compare three different bootstrap methods:

1. If n is the number of pairs (X, Y) , then draw n samples from the pairs with replacement, perform the estimation procedure for the coefficients, and repeat this B times.
2. Perform an initial estimation of the parameters and obtain estimated errors e by subtracting the initial model fit from the data $e = Y - b_0 - b_1 X$. Resample using replacement from these n estimated errors e and add them again to the initial model fit. Perform the parameter estimation for these pairs and repeat the entire procedure B times. This and method 1 are non-parametric bootstraps since they do not involve any distributions for the errors.
3. Estimate the parameters for error distribution and then bootstrap by generating n residuals e from this distributions to give $Y = b_0 + b_1 X + e$. Perform the estimation

procedure for the coefficients and repeat this B times. This is a parametric bootstrap as we are assuming/using a form for the density of the errors.

Data and Methodologies

To reveal the strengths and weaknesses of the three bootstrap and OLS and WLS methods, we simulated data sets. We chose the regression line

$$Y = 2.7 + 7.375X + \varepsilon \quad (3)$$

for X uniformly distributed in the range of 10 to 15. For the errors ε , we used a mixture of two Laplace distributions (also referred to as double-exponential distribution) (Alrasheedi 2012). The density f of the Laplace function with expectation 0 is

$$f(x, \lambda) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}.$$

Laplace $(0, \lambda)$ distributed variables can be generated as the difference of two independent identically distributed Exponential $(1/\lambda)$ random variables. We generated 4000 Laplace $(0, 0.243)$ together with 4000 Uniform $(10,15)$ (group A) and 6000 Laplace $(0, 0.101)$ together with 6000 Uniform $(10,15)$ (group B) distributed variables for ε and X and obtained Y using model (3). We referred to this as our base data set. This type of data can be obtained, for example, by measuring physical values with two different devices with different error variance. The error distribution for the entire data set is a mixture of two Laplace variables with two different variances; therefore, it is heteroscedastic. The Laplace distribution is more heavily tailed than the normal distribution, so we expect a greater number of 'outliers' in the data set. We modified our base data set in different manners to make the differences between the three bootstrap methods more pronounced. We created a data set with a greater number of 'outliers' by changing the values of three points of group A and B each to result in strong leverage points (referred to as the data set with a greater number of outliers) and another by removing three points (two from group A, one from group B), which have the strongest leverage (referred to as the data set with fewer outliers) by replacing its values by the average of the neighbors.

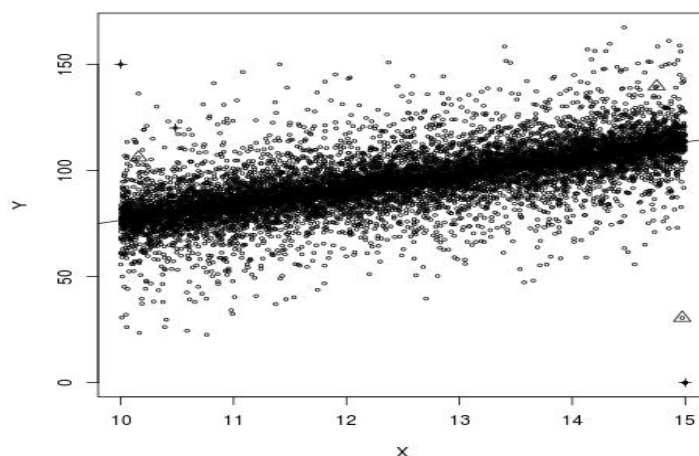


Fig. 1: Base data set with added outliers (crosses) and removed outliers (triangles with a circle inside) and regression lines.

Simulation and Data Analysis

All simulations and analyses were conducted using the statistical software package R. For all tables, given bootstrap estimates are the median over the 100 simulations and 'range' indicates the extremes of all 100 simulations (there is no 'range' for the OLS and WLS methods since they give one estimate). For the confidence intervals, 'yes' and 'no' indicate whether the true parameters were covered and single numbers, for example, 100, indicates how many of the 100th percentile confidence intervals covered the true parameters.

First, we examined the entire base data set while ignoring the two groups (Table 1). With the R function **lm**, we first performed linear regression (OLS) and obtained estimates for the coefficients 3.54 and 7.32, which are not similar to the true coefficients 2.7 and 7.375, particularly considering the sample size of 10,000. Nevertheless, the regression line in the plot appears identical to the true regression line. Next, we obtained confidence intervals for the coefficients using the command **confint**. The diagnostic plots show some problems with the fit; several outliers were observed and the Q-Q plot against the normal distribution shows significant deviation, which is expected, but it indicates that some of the results, particularly the confidence intervals, are inaccurate since they depend on the assumption of normality, and some of the points are strong leverage points and significantly influence parameter estimation (Figure 2).

Table 1: Base data set

Estimates	OLS	WLS	Bootstrap 1	Bootstrap 2	Bootstrap 3
Intercept	3.54	3.49	3.490	3.485	3.489
Range			3.44 - 3.53	3.42 - 3.55	3.44 - 3.55
Slope	7.32	7.3251	7.3248	7.3251	7.3249
Range			7.3215 7.3293	7.3205 7.3309	7.3202 7.3291
Standard error intercept	0.955	0.6948	0.7	0.6954	0.6957
Range			0.664-0.749	0.654-0.724	0.652-0.736
Standard error slope	0.076	0.05534	0.05557	0.05536	0.05527
Range			0.0527 0.0595	0.052 0.0579	0.0523 0.0587
95% CI cov intercept	(1.67, 5.41) yes	(2.12, 4.85) yes	100	100	100
Mean length	3.74	2.7241	2.7244	2.7022	2.712
99% CI cov intercept	(1.08, 6) yes	(1.7, 5.28) yes	100	100	100
Mean length	4.92	3.58	3.539	3.544	3.523
95% CI cov slope	(7.17, 7.47) yes	(7.22, 7.43) yes	100	100	100
Mean length	0.3	0.2169	0.2167	0.2153	0.2163
99% CI cov slope	(7.12, 7.52) yes	(7.18, 7.47) yes	100	100	100
Mean length	0.39	0.285	0.2803	0.2821	0.2804

In the next step, we used the group to which a data point belongs to estimate the error variances. We performed linear regressions (OLS) similarly to as described above separately for the groups A and B, obtained two sets of residuals, and estimated their variances. The maximum likelihood estimator for λ in the case of the Laplace distribution is:

$$\lambda = \frac{1}{N} \sum (|Y - \tilde{Y}|),$$

Where, Y is the sample median. The variance of the Laplace distribution is $2\lambda^2$. The reciprocals of the variances were then used as weights for the WLS method. Here, the estimates for the coefficients were slightly better; the standard errors of the estimates were smaller, and the confidence intervals were shorter, but the diagnostic plots show the same problems (Figure 2).

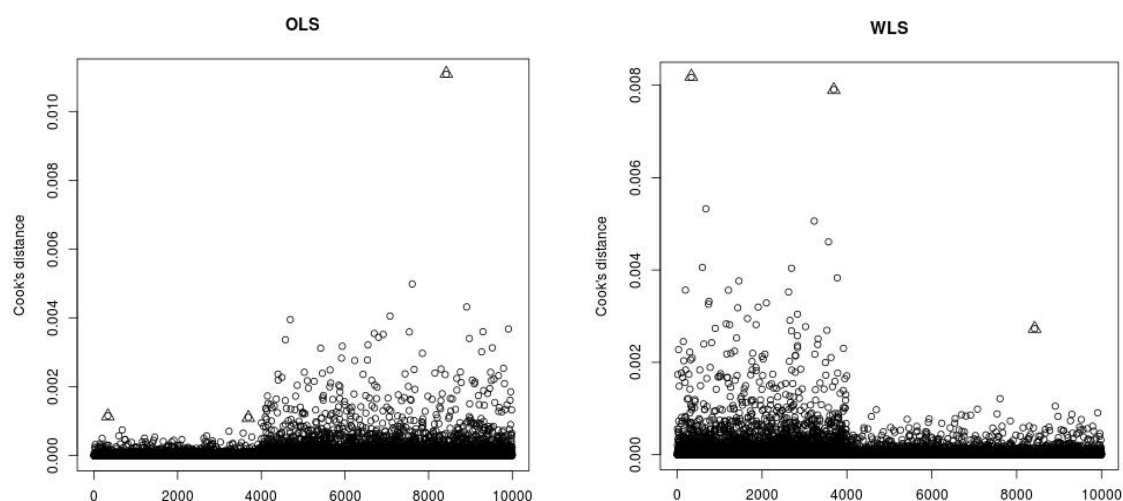


Fig. 2: Cook's distance for all data points. Points 1 to 4000 belong to group A, and 4001 to 10,000 to group B. Triangles indicate outliers which are removed from the base data set with fewer outliers. The left-hand picture shows Cook's distances for OLS and the right-hand picture for WLS.

We then continued to apply each of the three bootstrap methods to the base dataset. For bootstrap methods 2 and 3, we began with a WLS estimate similarly to above, obtained the residuals from the initial fit for method 2 and the parameters for the two Laplace distributions for method 3, resampled from the residuals (method 2) or generated new residuals from the distribution using the estimated parameters (method 3), and added these values to the initial fit. From these data, the WLS estimates were obtained. For each bootstrap run, we generated $B = 1000$ samples and repeated the runs 100 times. A short test using $B = 2000$ indicated that 1000 bootstrap runs provide sufficient coverage of the data. For each bootstrap method, we obtained 100 estimates of the coefficients, 100 estimates for the variances of the coefficients, and 100 percentile confidence intervals. Percentile confidence intervals were obtained by sorting the B estimates for the coefficients and discarding the 5% most extreme values to obtain a 95% confidence interval, for example. The three bootstrap methods showed very similar estimates to the WLS, with only the average length of the confidence intervals slightly shorter. There was no clear difference between the bootstrap methods.

Next, we applied all methods to the dataset with fewer outliers (Table 2). The parameter estimates were closer to the true values for all methods. In fact, the OLS estimate performed slightly better than other methods. The standard errors for the estimates and the length of the confidence intervals were worse using OLS than using other methods. WLS and

bootstrap methods showed similar results. Among the bootstrap methods, bootstrap 2 showed slightly lower standard errors of the estimates and shorter confidence intervals, but the parameter estimates were a bit worse compared to the other methods. Compared to the base data set, all methods benefited from the removal of three leverage points out of 10,000 data points, particularly the parameter estimates.

Table 2: Data set without three outliers

Estimates	OLS	WLS	Bootstrap 1	Bootstrap 2	Bootstrap 3
Intercept	3.54	3.49	3.490	3.485	3.489
Range			3.44 - 3.53	3.42 - 3.55	3.44 - 3.55
Slope	7.32	7.3251	7.3248	7.3251	7.3249
Range			7.3215 7.3293	7.3205 7.3309	7.3202 7.3291
Standard error intercept	0.955	0.6948	0.7	0.6954	0.6957
Range			0.664-0.749	0.654-0.724	0.652-0.736
Standard error slope	0.076	0.05534	0.05557	0.05536	0.05527
Range			0.0527 0.0595	0.052 0.0579	0.0523 0.0587
95% CI cov intercept	(1.67, 5.41) yes	(2.12, 4.85) yes	100	100	100
Mean length	3.74	2.7241	2.7244	2.7022	2.712
99% CI cov intercept	(1.08, 6) yes	(1.7, 5.28) yes	100	100	100
Mean length	4.92	3.58	3.539	3.544	3.523
95% CI cov slope	(7.17, 7.47) yes	(7.22, 7.43) yes	100	100	100
Mean length	0.3	0.2169	0.2167	0.2153	0.2163
99% CI cov slope	(7.12, 7.52) yes	(7.18, 7.47) yes	100	100	100
Mean length	0.39	0.285	0.2803	0.2821	0.2804

For the data set with a greater number of outliers (Table 3), OLS showed the best parameter estimates, followed by bootstrap method 1. Some of the confidence intervals no longer covered the true parameters. For WLS, the 95% confidence intervals did not cover the parameters, but the 99% intervals did. For the bootstrap methods, only method 3 covered the true intercept for the 99% confidence interval in 93 of 100 cases. For the 95% interval, the slope method 1 covered the true parameter in 91 of 100 cases, method 2 for 17 of 100 cases, and method 3 for 2 of 100 cases. For the 95% interval, the intercept showed worse results, for method 1 in 24 of 100 cases and in methods 2 and 3 in 1 of 100 cases.

Finally, we applied the bootstrap methods to the base data set without using any previous knowledge of the groups A and B (Table 4). For method 2, we resampled all 10,000 residuals, and for method 3 we assumed that the residuals showed Laplace distribution and estimated one Laplace parameter for all residuals. A Q-Q plot showed a well-fit linear relationship between simulated Laplace variables and the residuals. All obtained estimates were very close to OLS, but bootstrap method 3 gave lower standard errors for the estimates and shorter confidence intervals.

Table 3: Data set with more outliers

Estimates	OLS	WLS	Bootstrap 1	Bootstrap 2	Bootstrap 3
Intercept	4.319	4.341	4.3397	4.3421	4.3446
Range			4.2750 4.4119	4.2544 4.3965	4.2918 4.3921
Slope	7.2547	7.2532	7.2533	7.25297	7.25296
Range			7.2474 7.2586	7.2487 7.26	7.2491 7.2574
Standard error intercept	0.977	0.7293	0.8305	0.7464	0.7111
Range			0.776 0.877	0.692 0.782	0.673 0.744
Standard error slope	0.0777	0.058	0.0674	0.0595	0.0565
Range			0.063 0.071	0.055 0.062	0.053 0.059
95% CI cov intercept	(2.4, 6.23) yes	(2.91, 5.77) no	24	1	1
Mean length	3.83	2.86	3.23	2.93	2.76
99% CI cov intercept	(1.8, 6.84) yes	(2.46, 6.22) yes	100	100	93
Mean length	5.03	3.76	4.24	3.8	3.6
95% CI cov slope	(7.1, 7.41) yes	(7.14, 7.367) no	91	17	2
Mean length	0.3	0.228	0.2625	0.233	0.22
99% CI cov slope	(7.05, 7.45) yes	(7.1, 7.4) yes	100	100	100
Mean length	0.4	0.2992	0.3443	0.3038	0.2865

Table 4: Base data set without knowledge about the groups A and B

Estimates	OLS	Bootstrap 1	Bootstrap 2	Bootstrap 3
Intercept	3.539	3.537	3.539	3.544
Range		3.466 3.616	3.471 3.631	3.474 3.599
Slope	7.32	7.3204	7.32	7.3197
Range		7.3136 7.3259	7.3123 7.3258	7.315 7.3255
Standard error intercept	0.955	0.976	0.957	0.896

Range		0.918 1.03	0.913 1.014	0.846 0.955
Standard error slope	0.076	0.0778	0.0762	0.0715
Range		0.0731 0.082	0.0726 0.0801	0.0672 0.0754
95% CI cov intercept	(1.67, 5.41) yes	100	100	100
Mean length	3.74	3.81	3.74	3.5
99% CI cov intercept	(1.08, 6) yes	100	100	100
Mean length	4.92	4.95	4.87	4.58
95% CI cov slope	(7.17, 7.47) yes	100	100	100
Mean length	0.3	0.303	0.296	0.278
99% CI cov slope	(7.12, 7.52) yes	100	100	100
Mean length	0.39	0.395	0.386	0.363

Discussion

The simulation results show that for a data set with heavy leverage points (data set with a larger number of outliers), bootstrap method 1 is the most robust. The bootstrap resampling procedure often does not choose some of the leverage points, so that the estimates are less influenced by special points. Methods 2 and 3 begin with an initial fit which is under the influence of leverage points, and cannot easily overcome this. Using method 1, the confidence intervals did not always show the desired coverage probability. The comparison between the base data set and that with fewer outliers showed that the bootstrap methods benefit strongly from elimination of leverage points. The OLS method did not perform poorly, even without data normalization. However, the OLS estimator is the BLUE and estimates of the standard errors are valid. To be sure of the quality of the confidence intervals, the data should be normalized (at least approximately through the central limit theorem). However, the OLS then behaved in a conservative manner, resulting in confidence intervals which were a bit longer. Although the bootstrap depends on fewer assumptions and is therefore more robust, it should not be used in all situations, but can be used for linear regression of standard diagnostics and potentially some treatment of the data.

Apart from the confidence intervals for the data set with a larger number of outliers, WLS performed quite well. Parameter estimates were better than for OLS, but this is because WLS used additional information regarding the two groups A and B. This reduces the influence of specific data points, but in the case of the extreme leverage points, this method was not sufficient.

When applied to the base data set, without knowledge regarding the groups A and B, method 3 performed better since it gave smaller standard errors and shorter confidence intervals than the other two methods. Amiri et al. (2008) compared parametric and non-parametric bootstrap methods and concluded that when bootstrapping, variance in the behavior of the bootstrap methods depends on kurtosis. If the sample kurtosis is larger than the kurtosis of the distribution used in method 3, the obtained standard errors were smaller for method 3. This suggests that a similar rule would hold here: if the kurtosis of the residuals obtained after initial fitting is larger than the kurtosis of the distribution, method 3 gives lower standard errors and shorter confidence intervals. The kurtosis of the residuals, defined as

$$K_s = \frac{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^4}{\left(\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2 \right)^2} - 3,$$

after initial fit when neglecting groups A and B, is 4.4 (for a general discussion regarding kurtosis, see Gill and Joanes 1998, Decarlo 1997). Kurtosis of the Laplace distribution is 3. For groups A and B separately, kurtoses are 2.45 and 2.46. In these situations (Table 1-3), method 3 did not perform better than the other methods. High kurtosis indicates a heavily tailed distribution with a higher likelihood of extreme values. If sample kurtosis is high, extreme values are already present may influence the estimation process. If new residuals are generated from a distribution with lower kurtosis, the chances for retaining extreme values are less, resulting in smaller standard errors and shorter confidence intervals.

Conclusion

None of the tested methods were consistently better than the others; rather, performance depended on the available data. For the base data set (Laplace errors with different variances), only OLS was slightly worse in all aspects, as it did not take into account the information regarding the two groups with different error variances. Removing leverage points from the data set helped to improve all methods and only OLS was slightly worse with respect to the standard variations of parameter estimates and lengths of the confidence intervals. With leverage points added to the data set, some methods showed problems with the coverage probability of the confidence intervals. OLS and bootstrap 1 performed the best in this situation. Applying all methods (without WLS) to the base data set, but without using knowledge regarding the two groups with different error variances, bootstrap 3 showed smaller standard errors for the estimates and shorter confidence intervals. These results

indicate that OLS still performs reasonably well even without data normalization. The performance of WLS was comparable to that of the three bootstrap methods. Bootstrap 1 (resampling the data pairs) was more robust towards outliers/leverage points and bootstrap 3 (resampling from a known distribution) showed lower variance if the kurtosis of the sample residuals was larger than the kurtosis of the distribution from which resampling was conducted.

References:

Aitken, A., On Least Squares and Linear Combinations of Observations, Proceedings of the Royal Society of Edinburgh, Vol. 55, pp. 42–48, 1935.

Alrasheedi M., Confidence Intervals for Double Exponential Distribution: A Simulation Approach, International Journal of Computational and Mathematical Sciences; 6(1) pp108-112, 2012.

Amiri S., Rosen V., and S. Zwanzig, On the comparison of parametric and nonparametric bootstrap. Department of Mathematics, Uppsala University, 2008.

Benton D., and Krishnamoorthy K., Performance of the parametric bootstrap method in small sample interval estimates. Advances and Applications in Statistics, 2, pp 269–285, 2002.

DeCarlo L., On the meaning and use of kurtosis. Psychological Methods, 2, pp. 292-307, 1997.

Efron B., Better Bootstrap Confidence Intervals, Journal of the American Statistical Association Vol. 82, No. 397 pp171–185, 1987.

Efron B., Tibshirani R, Introduction to the Bootstrap. Chapman & Hall, 1993.

Gill D., and Joanes C., Comparing measures of sample skewness and kurtosis. The Statistician, Vol. 47, part 1, pp 183-189, 1998.

Rao C., Toutenburg H., Fieger A., Heumann C., Nittner T., and S. Scheid, Linear Models: Least Squares and Alternatives. Springer Series in Statistics, 1999.

Wolberg J., Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments. Springer, 2005.

Zhu, J, P. Jing, The Analysis of Bootstrap Method in Linear Regression Effect, Journal of Mathematics Research Vol. 2, No. 4, pp 64-69, 2010.