Journal of Modern Applied Statistical Methods

Volume 7 | Issue 2 Article 17

11-1-2008

Bootstrap Confidence Intervals and Coverage Probabilities of Regression Parameter Estimates Using Trimmed Elemental Estimation

Matthew Hall
Child Health Corporation of America, matt.hall@chca.com

Matthew S. Mayo
University of Kansas Medical Center, mmayo@kumc.edu

Journal of Modern Applied Statistical Methods
May 2008, Vol. 7, No. 2, 514-525

1538 – 9472/08/$95.00

Bootstrap Confidence Intervals and Coverage Probabilities of Regression Parameter Estimates Using Trimmed Elemental Estimation

Matthew Hall
Child Health Corporation of America

Matthew S. Mayo
University of Kansas Medical Center

Mayo and Gray introduced the leverage residual-weighted elemental (LRWE) classification of regression estimators and a new method of estimation called trimmed elemental estimation (TEE), showing the efficiency and robustness of TEE point estimates. Using bootstrap methods, properties of various trimmed elemental estimator interval estimates to allow for inference are examined and estimates with ordinary least squares (OLS) and least sum of absolute values (LAV) are compared. Confidence intervals and coverage probabilities for the estimators using a variety of error distributions, sample sizes, and number of parameters are examined. To reduce computational intensity, randomly selecting elemental subsets to calculate the parameter estimates were investigated. For the distributions considered, randomly selecting 50% of the elemental regressions led to highly accurate estimates.

Key words: Elemental subsets, elemental regression, robust regression, coverage probabilities.

Introduction

A popular method of finding a solution to the multiple linear regression model

$$Y = X\beta + \varepsilon \qquad (1.1)$$

is to make use of the ordinary least squares (OLS) solution:

$$\hat{\beta}_{OLS} = (X^tX)^{-1} X^t Y.$$

In this nomenclature, Y is a $n \times 1$ vector of random observations, X is a $n \times p$ matrix of known constants, β is a $p \times 1$ vector of unknown parameters, and ε is a $n \times 1$ vector of random errors with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 I$. The OLS solution purposefully minimizes the sum of squared residuals

$$SSE(\hat{\beta}) = (Y - X\hat{\beta})^t (Y - X\hat{\beta}).$$

There are many reasons why this solution is desirable, such as ease of calculation and the well developed theory that supports it. However, the OLS solution is also known to be sensitive to outliers and/or violations of model assumptions.

Several attempts to develop solutions that are less sensitive to outliers have been developed. These include least absolute values (LAV) regression, which minimizes the sum of the absolute residuals, and L_p-norm regression, which minimizes the sum of the p^{th} powers of the absolute residuals. This article furthers the work of another method called the trimmed elemental estimator (TEE), first proposed by Mayo and Gray (1), that makes use of elemental subsets.

Elemental Subsets

In most cases when using model (1.1), n (the sample size) is much greater then p (the number of unknown parameters), and the system of equations becomes over-determined. However, in order to estimate

Matthew Hall is Senior Statistician, Department of Informatics. Email: matt.hall@chca.com. Matthew S. Mayo is Professor and Chair, Department of Biostatistics. Email him at MMAYO@kumc.edu.

$\beta_0, \beta_1, \beta_2, \ldots, \beta_p$, only $k = p+1$ observations are mathematically required. Thus, when solving the over-determined system, a choice must be made from infinitely many possible solutions in order to settle on a single regression line. One way to deal with this issue is to ignore the fact that only k observations are needed and to pool all n observations into a single system of k equations to solve: this is what OLS does. Alternatively, subsets of the data could be formed with exactly k observations, their corresponding fits found, and the best one taken: this is what LAV does. An even better method might be to take several of the fits in this scheme and use their combined information to settle on estimates. Mayo and Gray (1997) developed TEE for this purpose. Using either of these last two approaches makes use of elemental subsets and elemental regression.

An elemental subset of a data set is simply a subvector of the data. In the setting of model (1.1), a subvector $h = \{i_1, i_2, \ldots, i_p\}$ may be considered as a set of distinct indices from a set of n indices. X_h may be defined as the $p \times p$ submatrix of X containing the rows of X indexed by the subset h. Furthermore, Y_h can be defined as the corresponding $p \times 1$ subvector of Y. The solution to the elemental regression equation is given by:

$$\hat{\beta}_h = \left(\mathbf{X}_h^t \mathbf{X}_h\right)^{-1} \mathbf{X}_h^t \mathbf{Y}_h = \mathbf{X}_h^{-1} \mathbf{Y}_h .$$

With the advent of high speed computers, elemental regression has been revived from its forgotten past nearly 250 years ago. It was, in fact, a predecessor to least squares, introduced in 1755 by Boscovich. However, due to its computational intensity and the introduction of least squares, it fell out of favor with data analysts. The need for computational power is evident when considering even a small data set. For example, assume a sample size of 50 and the need to estimate three parameters. There are ${}_{50}C_3$ = 19,600 elemental subsets of the data that must be fit. This is clearly beyond human capability.

Renewed interest in elementals has occurred on many fronts. Going back to the early days of modern computers, Theil (1950) and Sen (1968) used elementals to develop simple linear regression estimators. On the diagnostics front, Rubin (1980), Hawkins (1993), and Welsch (1986) used elementals to detect outliers and perform other regression diagnostics. Rousseeuw and Bassett (1991) and Hawkins (1993) considered searching through the set of elemental regressions and selecting the optimal parameter estimates based on specified criteria. Hawkins further defined, for a specified fitting criterion, the best elemental estimator (BEE) as the optimal estimate over all elemental fits. Recently, Hawkins and Olive (2002) introduced the X-cluster algorithm as a form of elemental regression for large multiple regression datasets.

Mayo and Gray's (1997) contribution introduced regression estimators based on OLS in terms of elemental regression. Sheynin (1973) reported that Jacobi was the first to show that OLS can be viewed as a weighted average of elemental regressions:

$$\hat{\beta}_{OLS} = \frac{\sum_h \left|X_h^t X_h\right| \hat{\beta}_h}{\sum_h \left|X_h^t X_h\right|} = \sum_h \frac{\left|X_h^t X_h\right|}{\left|X^t X\right|} \hat{\beta}_h \qquad (1.2)$$

where h is the set of all possible elemental subsets and the single bars indicate determinates. Furthermore, the weights are defined as:

$$w_h = \frac{\left|\mathbf{X}_h^t \mathbf{X}_h\right|}{\left|\mathbf{X}^t \mathbf{X}\right|} .$$

Because these weights are between zero and one and must sum to one, OLS is a weighted average of the elemental regressions $\hat{\beta}_h$.

Mayo and Gray (1997) took this version of OLS and generalized it to a class of estimators which they called leverage-residual weighted elemental (LRWE) estimators of the form:

$$\hat{\beta}(\lambda, \rho) = \frac{\sum_h w[\lambda(h), \rho(h)] \hat{\beta}_h}{\sum_h w[\lambda(h), \rho(h)]} \qquad (1.3)$$

In this formulation, $\lambda(h)$ is a factor based on the leverage information for X_h , and $\rho(h)$ is a factor based on the degree of fit for the elemental regression h. The OLS version is observed (1.2) in this form where

$$\lambda(h) = \left|\mathbf{X}_h^t \mathbf{X}_h\right|, \ \rho(h) = 1 \text{ for all } h, \quad (1.4)$$

and

$$w[\lambda(h), \rho(h)] = \lambda(h)\rho(h).$$

OLS does not make use of the weight factor based on the degree of fit, $\rho(h)$. For this reason, in OLS, elemental regressions with extreme data points are weighted the same as those that behave normally. Thus, OLS can be easily influenced by the presence of outliers.

Trimmed Elemental Estimators

Instead of ignoring the goodness of fit of a regression to a set of elementals, $\rho(h)$ could be altered in the OLS formulation of (1.4). Mayo and Gray (1997) created what they called the trimmed elemental estimator (TEE) to trim out the elemental regressions that poorly fit the data or have extreme leverage. The benefit of such a strategy is to remove from consideration elemental regressions that are computed from outlying data, thus achieving a more robust regression. Using the same $\lambda(h)$ and $w[\lambda(h), \rho(h)]$ as in (1.4), they altered $\rho(h)$ to be the indicator function:

$$\rho(h) = I\left(\sum_{i=1}^{n}\left|e_{hi}\right| \le (1-\alpha_p)100^{th} \text{ percentile of the } {}_nC_p \sum_{i=1}^{n}\left|e_{hi}\right| \text{ values}\right)$$

Here, α_p is a trimming constant between zero and one and $\sum_{i=1}^{n}\left|e_{hi}\right|$ is the sum of absolute residuals (SAE) resulting from the elemental estimate $\hat{\beta}_h$. Depending on the proportion of regressions one would like to remove from consideration as determined by their goodness of fit, α_p can be adjusted accordingly. Thus, many trimmed elemental regression estimators can be found and denoted by TEE(α_p).

Mayo and Gray (2001) used simulation results to show the robustness and efficiency properties of TEE(α_p) point estimates to normal and symmetric non-normal error distributions, a feature which OLS does not enjoy. Results showed that the TEE(α_p) offers high efficiency under normality and is very robust to non-normality. This article furthers their work by examining some bootstrap confidence intervals of the trimmed elemental estimator and their properties and reducing computational intensity through random selection of elemental regressions.

Methodology

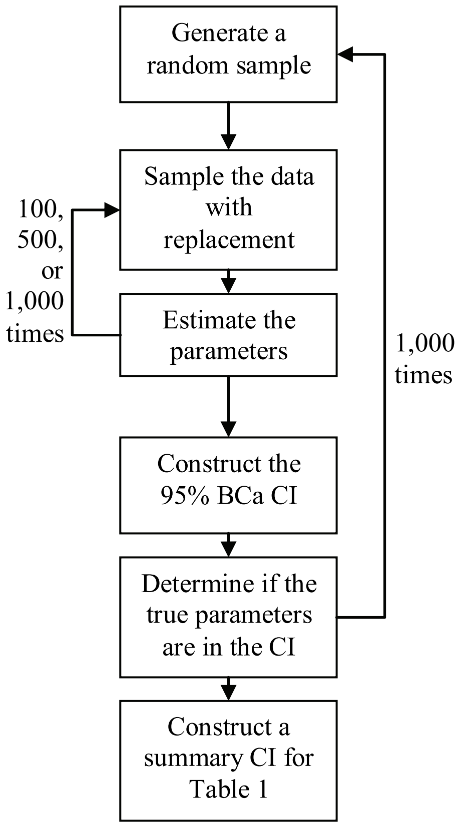
Simulation Design

Simulations were aimed at gaining a better understanding of the TEE(α_p) for inference by creating confidence intervals for the parameters and coverage probabilities under various scenarios. The objective was to compare these using the following methods: least absolute values (LAV), TEE(0.25), TEE(0.50), TEE(0.75), and OLS. Furthermore, a variety of error term distributions were assumed including: Normal, Laplace, Cauchy, 10% Contaminated Normal, and Student's t. These distributions were selected to provide a variety of weight in the tails of the distribution. In the simulations, Normal, Laplace, and t distribution parameter values had an error variance (σ^2) of 3.0. For the Normal distribution, standard normal variates were generated and multiplied by σ.

For Laplace, random variates from an exponential distribution were generated (mean = 1.0), randomly assigned a sign, and multiplied by $\sigma/2$. The Cauchy was the standard Cauchy distribution. For the 10% Contaminated Normal errors, standard normal variates were generated and-based on the value of a uniform random variate-were multiplied by either $\sqrt{5}\sigma$ (with probability 0.1) or σ (with probability 0.9). Finally, for the Student's t error distribution, three degrees of freedom were used in order for $\sigma^2 = 3$. The independent variable X was generated from a N(3,3) distribution.

In order to achieve the research goals, various quantities of 95% bias-corrected and accelerated (BC_a) bootstrap confidence intervals for OLS, LAV were calculated, and various trimmed elemental estimators and determined the number of times the true value of the parameter was in the intervals. Figure 1 shows the flowchart for the simulations.

Figure 1: Simulations flowchart.



The bootstrap is a well-developed approach to calculating approximate confidence intervals for parameter estimates when exact confidence intervals do not exist by repeatedly resampling the data with replacement. The BC_a method was introduced by Efron (1987) as an improvement to the bias-corrected (BC) method of Efron (1982) in order to provide confidence intervals for a wider class of problems. It constitutes a method for setting approximate confidence intervals for a parameter based on the percentiles of the bootstrap histogram, a bias correction, and an acceleration constant which measures how rapidly the standard error is changing on the normalized scale. For a complete review of various bootstrap confidence intervals including BC_a, see DiCiccio and Efron (1996). As a way of summarizing the BC_a confidence intervals, an overall 95% interval was calculated for each parameter. For this interval, the lower limit represents the value for which 2.5% of the lower boundaries of the BC_a confidence intervals are less than this value. Similarly, the upper limit represents the value for which 2.5% of the upper boundaries of the BC_a confidence intervals are greater than this value. All simulations were performed on a Dell 1.6GHz Pentium 4 computer with 1.0 GB of RAM using Digital FORTRAN 90.

In order to verify that the program was performing properly, the performance was tested using the two extreme methods under consideration: LAV, which takes only a single elemental regression, and OLS, which uses all of the elemental regressions. Comparing the parameter estimates (p = 2, n = 25) provided by the program for the three error distributions to the estimates provided by SAS© version 8e, agreement to five significant digits was obtained.

Results

In order to understand how the TEE(α_p) estimators would act under different situations, the following simulation scenarios were chosen:

a) a small sample size of 10 with two parameters;
b) a moderate sample size of 25 with three parameters;
c) a moderate sample size of 25 with two parameters; and
d) a large sample size of 100 with five parameters.

Sample sizes and number of parameters were chosen to limit computing time while allowing properties of the confidence intervals across a variety of scenarios to be ascertained. The results of simulations (c) and (d) are not presented here, they were performed to verify that the results did not change dramatically when the sample size and number of parameters was altered. The results of these simulations were very similar to the results discussed in greater detail below. Any exceptions are noted.

For these simulations, there were ${}_{10}C_2$ = 45, ${}_{25}C_3$ = 2,300, ${}_{25}C_2$ = 300, and, ${}_{100}C_5$ = 75,287,520 elemental subsets that had to be fit for each bootstrap sample, respectfully. For simulation (a), Table 1 shows the summary 95%

intervals for the BC_a confidence intervals for β_1 using the method previously described. The smallest confidence interval in each scenario is highlighted. Figure 2 shows the coverage probabilities for the 1,000 BC_a confidence interval created by the bootstrap (100, 500, or 1,000 samples) for β_1 from simulation (a). Similarly, Figure 3 shows the coverage probabilities for β_1 and β_2 from simulation (b).

From Table 1, it is evident that the summary intervals tend to tighten around the true values of the parameters as the number of bootstrap samples increase. As long as the error term is Normal or 10% Contaminated Normal, OLS does quite well. Furthermore, regarding the 1,000 bootstraps, it is apparent that OLS is difficult to distinguish from TEE(0.25) when the error is Normal, 10% Contaminated Normal, or Student's t. However, as expected, when the error term is either Cauchy or Laplace, OLS is clearly not the best choice. With a Cauchy error term, it appears that TEE(0.75) performs best for the slope regardless of sample size or the number of parameters (simulations (b) , (c), and (d) also showed TEE(0.75) to be superior). When the error follows the Laplace distribution, TEE(0.50) or TEE(0.25) seem to be the best (simulations (b), (c), and (d) showed TEE(0.50) to be slightly better than TEE(0.25)). In sum, it appears that TEE(0.50) performs very well for all of the error distributions considered. Although not shown, the results were very similar for the intercept in all four simulations with only slightly wider intervals. The parameter β_2 in simulation (b) had very similar results to those discussed above for β_1.

Figures 2 and 3 show how the different methods performed at covering the true values of the parameters with their 95% BC_a confidence intervals for simulations (a) and (b), respectively. Although not shown in either figure, the confidence intervals for the intercept fail to include the true parameter more frequently than the slope confidence intervals. Nonetheless, the coverage probabilities for the intercept ranged from 0.90 to 0.97 for all simulations. Considering the 1,000 bootstrap samples (dashed lines) in the figures, the coverage probabilities for the error distributions studied ranges from 0.90 to 0.98. Thus, all of the methods captured the true values of the parameters quite well. However, regardless of the error distribution considered, TEE(0.50) appears to perform very consistently.

Furthermore, it is observed that either LAV or TEE(0.75) has the highest coverage probabilities, while OLS has the lowest for the error distributions under consideration. In fact, since the coverage probabilities were expected to be at 0.95, it is generally the case that LAV and TEE(0.75) performed above this level, TEE(0.50) and TEE(0.25) performed at this level, and OLS performed below this level. Hence, the coverage probability decreases as the trimming constant (α_p) decreases. The data from the other simulations were very similar and are not presented here. Once again, the coverage probabilities for β_2 in simulation (b) were similar to the probabilities for β_1 described above.

An objective in this article was to reduce the amount of necessary computations to achieve an acceptable estimate for the parameters using TEE(α_p). How this might be accomplished through random selection of elemental subsets as suggested by Hawkins (1993) for the BEE was investigated.

For simulation purposes, all of the elementals were first used to construct all of the elemental regressions $\hat{\beta}_h$. Specified proportions (30%, 50%, 70% or 90%) of these were then randomly selected in order to calculate parameter estimates through equation (1.3). This was performed with 10,000 data sets, and the median estimate was calculated for each error distribution at each percentage. The median was selected since it is a more robust measure of central tendency when compared to the mean. For β_1 when n=10 and p=2, the medians are displayed in Figure 4.

Using 50%, 70%, or 90% of the elemental regressions seems to provide accurate estimates for β_1 as long as the error distribution is one of those under consideration here. By selecting only 30% of the elemental regressions, the median estimates diverged further from the true value when compared to the other proportions, especially for the Normal, 10% Contaminated Normal, and the Student's.

Table 1: Summary intervals of 1,000 BC_a confidence intervals for β_1 when N=10, p=2. The true value is one.

	100 Bootstraps	500 Bootstraps	1000 Bootstraps
Normal			
LAV	(-2.645, 3.880)	(-2.023, 3.900)	(-1.930, 4.029)
TEE (0.75)	(-2.385, 3.537)	(-1.948, 3.706)	(-1.913, 3.952)
TEE (0.50)	(-2.032, 3.232)	(-1.620, 3.338)	(-1.525, 3.433)
TEE (0.25)	(-1.692, 2.950)	(-1.254, 3.065)	(-1.192, 3.111)
OLS	(-1.530, 2.838)	(-1.200, 2.994)	(-1.117, 3.106)
Cauchy			
LAV	(-29.597, 18.777)	(-31.469, 26.944)	(-25.958, 18.943)
TEE (0.75)	(-29.036, 16.883)	(-30.733, 26.192)	(-24.885, 18.313)
TEE (0.50)	(-31.722, 16.812)	(-28.499, 31.962)	(-29.893, 19.275)
TEE (0.25)	(-40.439, 27.037)	(-30.955, 31.913)	(-40.622, 24.040)
OLS	(-39.576, 31.148)	(-38.294, 38.800)	(-42.077, 22.391)
Laplace			
LAV	(-8.493, 7.962)	(-7.793, 8.521)	(-5.495, 7.960)
TEE (0.75)	(-8.335, 7.699)	(-7.340, 8.414)	(-5.157, 7.954)
TEE (0.50)	(-6.852, 6.901)	(-6.003, 7.533)	(-4.579, 6.931)
TEE (0.25)	(-6.895, 6.515)	(-5.709, 6.907)	(-4.794, 7.096)
OLS	(-7.371, 6.715)	(-5.719, 6.921)	(-4.974, 7.488)
Contam			
LAV	(-3.005, 4.390)	(-2.730, 4.278)	(-2.558, 4.666)
TEE (0.75)	(-2.876, 4.170)	(-2.685, 4.190)	(-2.528, 4.507)
TEE (0.50)	(-2.680, 3.965)	(-2.302, 6.644)	(-2.093, 4.187)
TEE (0.25)	(-2.517, 3.591)	(-1.948, 3.525)	(-1.635, 3.935)
OLS	(-2.470, 3.531)	(-1.807, 3.473)	(-1.672, 3.800)
T-distribution			
LAV	(-2.477, 3.895)	(-2.249, 3.555)	(-1.794, 4.330)
TEE (0.75)	(-2.554, 3.870)	(-2.161, 3.490)	(-1.842, 4.219)
TEE (0.50)	(-2.180, 3.537)	(-1.738, 3.164)	(-1.518, 3.894)
TEE (0.25)	(-1.808, 3.281)	(-1.482, 3.077)	(-1.288, 3.733)
OLS	(-1.746, 3.297)	(-1.447, 3.016)	(-1.280, 3.751)

Figure 2: Coverage probabilities of the 1,000 BC_a confidence intervals for β_1 when N=10 and p=2.

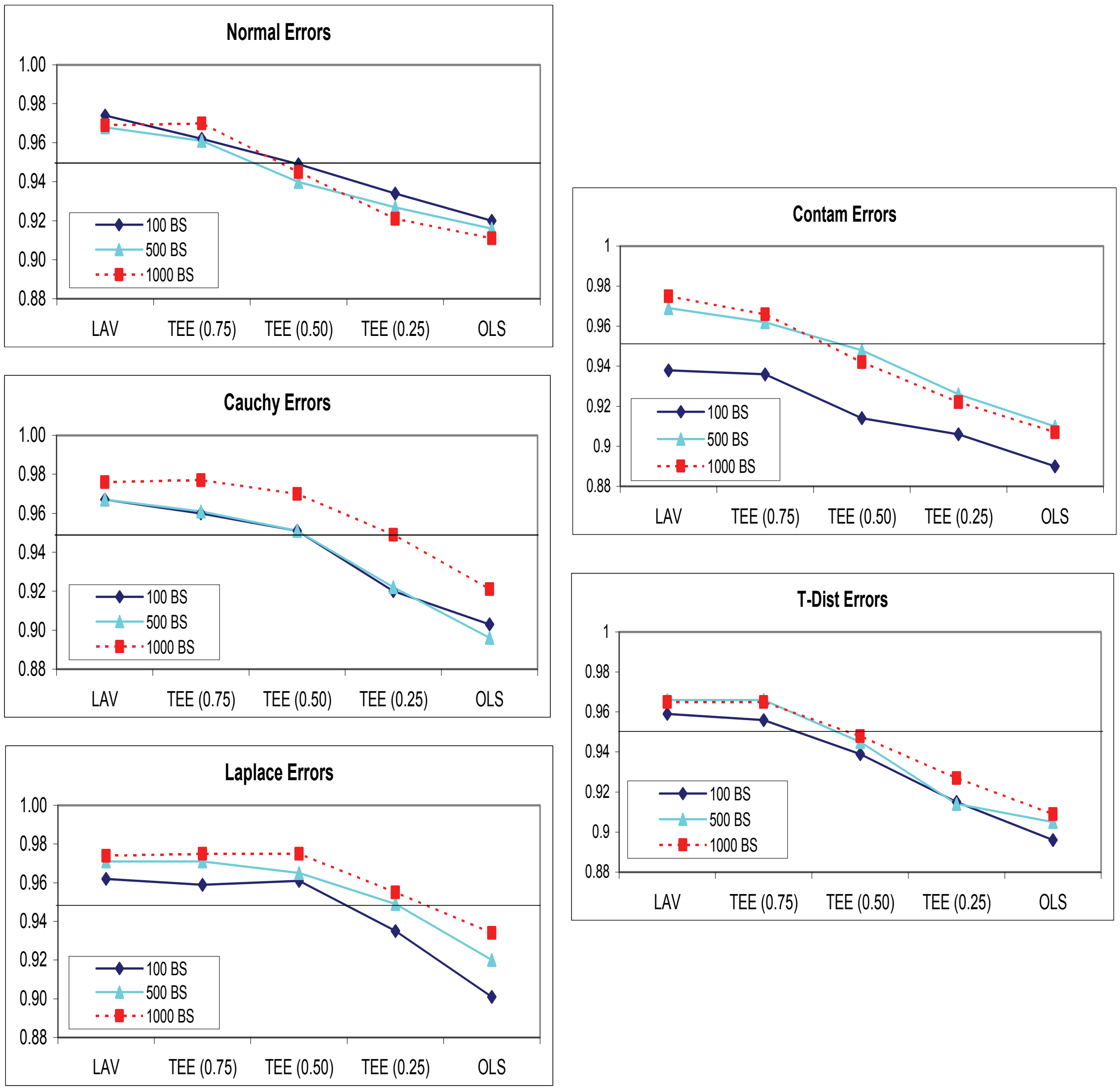


Figure 3: Coverage probabilities of the 1,000 BC_a confidence intervals for β_1 (column 1) and β_2 (column 2) when N=25 and p=3.

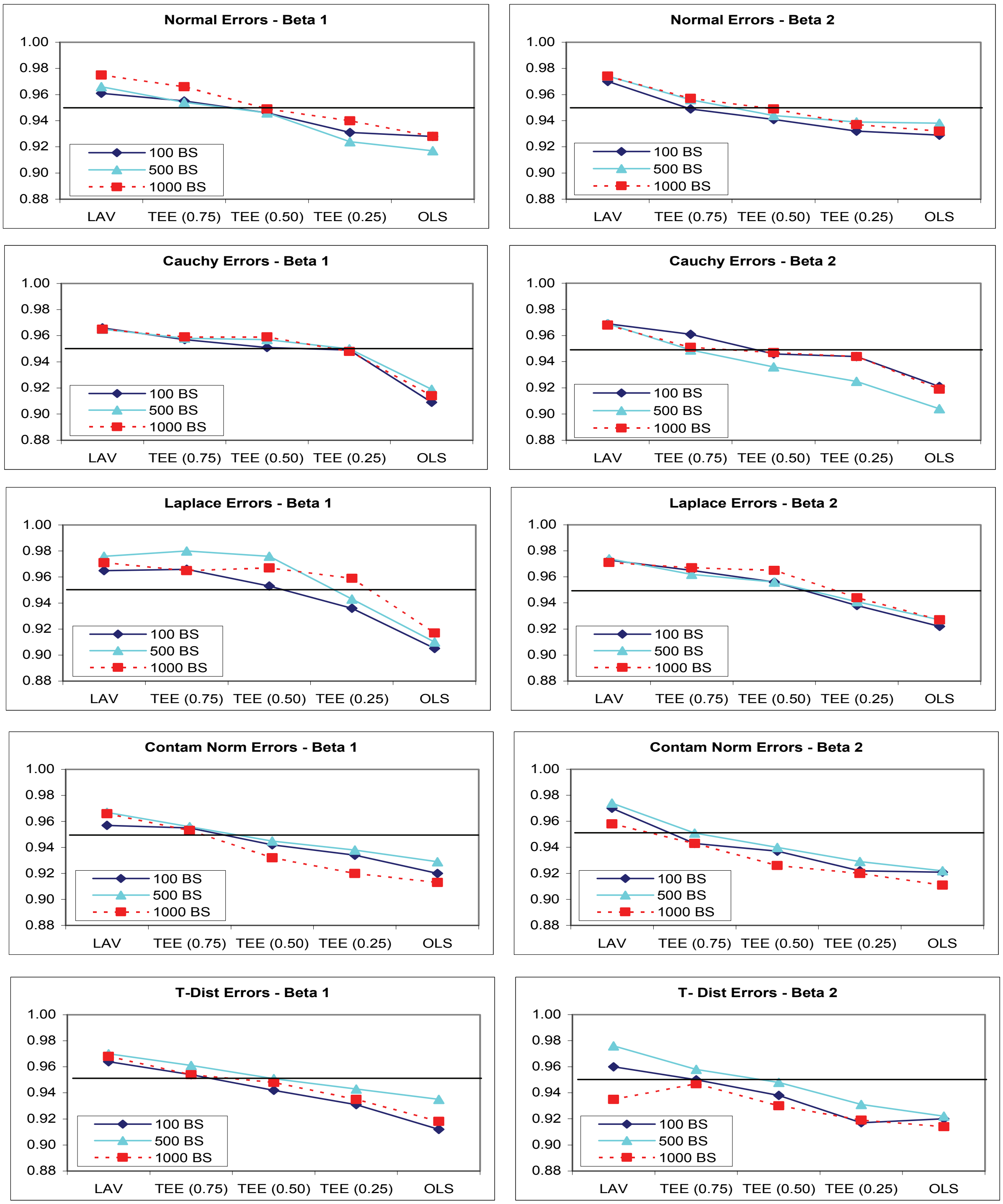
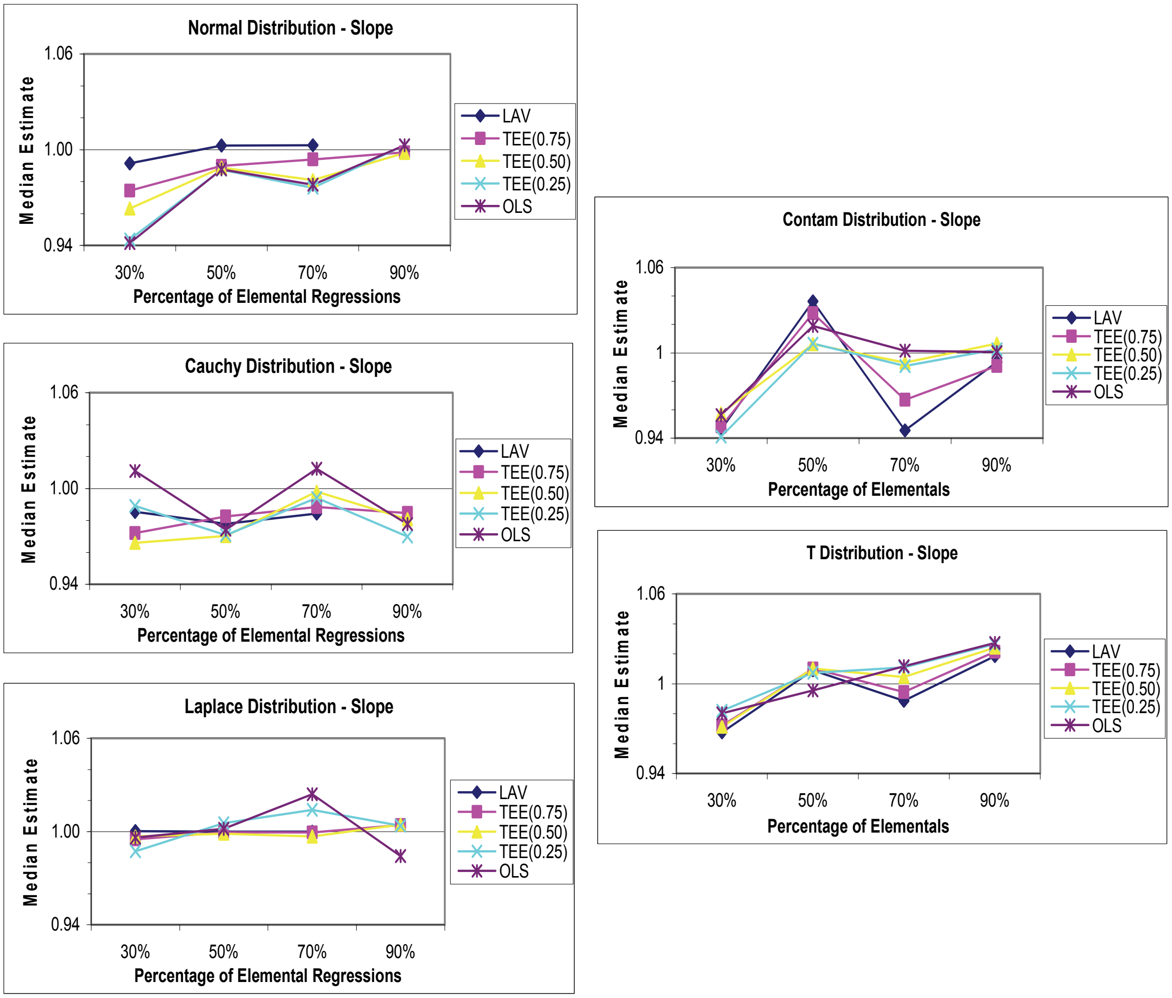


Figure 4: Median estimates for β_1 of 10,000 simulated data sets (N=10, p=2) using random selection of elemental regressions. The true value is one.



Thus, it appears that randomly selecting at least 50% of the elemental regressions is sufficient for producing accurate estimates. These results are similar for the intercept (data not shown) with the exception of using 50% of the elemental regressions with Laplace errors. In this situation, TEE(0.25) and OLS overestimated the intercept considerably. However, at 70%, the estimates behaved much more like those seen in Figure 4.

Figure 5 shows the coverage probabilities for the 95% BC_a confidence intervals using various quantities of bootstrap samples when n = 10, p = 2, and 50% of the elemental regressions are randomly selected. When similar simulations (n = 10, p = 2) are compared between Figure 5 (coverage probabilities when 50% of the elemental regressions are randomly selected) and Figure 2 (coverage probabilities without random selection), it is observed that results are quite similar. That is, while the coverage probabilities in Figure 5 are slightly higher than those in Figure 2, the trends seem similar. As was the case in Figure 2, generally speaking, LAV and TEE(0.75) over perform at the 95% level, TEE(0.50) and TEE(0.25) performed consistently at the 95% level, and OLS performed below the 95% level. Coverage probabilities from randomly selecting 70% and 90% of the elemental regressions produced similar results with the lines generally moving closer (as the percentage increased) to those observed in Figure 2.

Conclusion

The construction of BC_a confidence intervals for the trimmed elemental estimators have been demonstrated and their coverage probabilities have been. These are necessary extensions to Mayo and Grays original work and are additions to the development of TEE for inference purposes. In agreement with Mayo and Gray, this article showed that the trimmed elemental estimators are desirable in many situations. In fact, among those considered, they seem to be the clear choice when the error distribution is Cauchy or Laplace. Furthermore, for the Normal, 10% Contaminated Normal, or Student's t error distributions, trimmed elemental estimators were found to be almost indistinguishable from OLS. In addition, TEE(0.50) performed consistently well in terms of estimation and coverage probabilities for all of the error distributions under consideration. It appears that a researcher could be fairly comfortable in choosing TEE(0.50), however knowledge of the process should guide this and utilization of traditional graphical procedures, such as residual and fitted value plots, might aid in determining the trimming constant. The TEE requires a large number of calculations as compared with OLS, therefore, it is desirable to use OLS when it is known that the assumptions for OLS are not violated and that there are no outliers present.

When data sets become larger and the number of parameters increases, increasing computational difficulties for LRWE estimators are present. Since there are ${}_nC_p$ elemental subsets that must be fit, ways must be found to decrease the number of computations. Hawkins (1993) suggested that using a random subsample of the elemental subsets would produce a good estimate for the best elemental estimator. This article examined such random subsamples to determine if this method is appropriate for reducing the number of calculations required for the trimmed elemental estimator. It was found that utilizing at least 50% of the elemental regressions generally provides good results as long as the error distribution is Normal, Cauchy, Laplace, 10% Contaminated Normal, or Student's t. It was also observed that estimates tend to drift from the true value when random sampling falls to 30%.

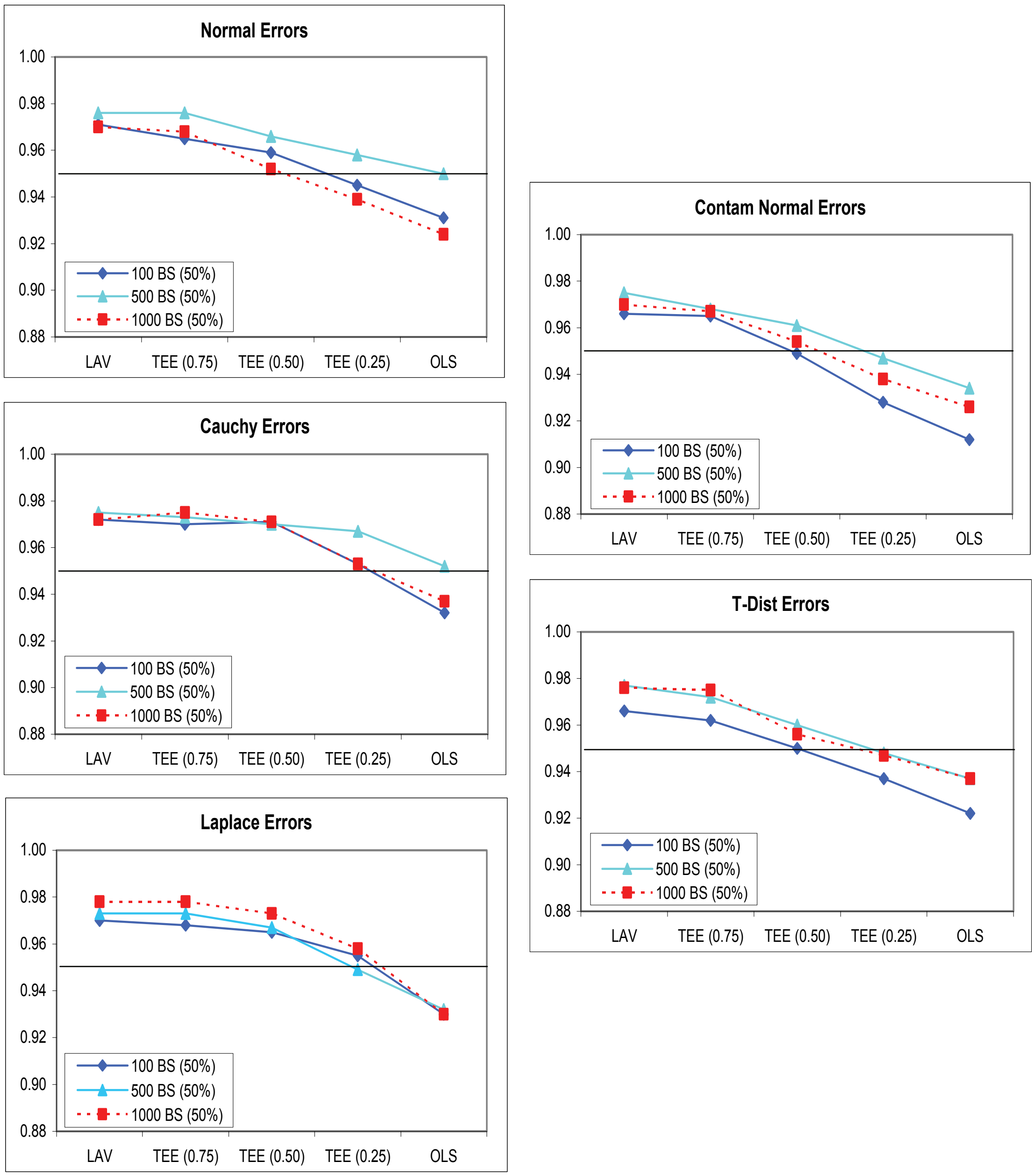
References

DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*, 189-212.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, *82*, 171-185.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *Conference Board of the Mathematical Sciences*, *38,* Society for Industrial and Applied Mathematics – National Science Foundation.

Figure 5: Coverage probabilities of the 1,000 BC_a confidence intervals for β_1 (N=10 and p=2) when randomly selecting 50% of the elemental regressions.



Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. NY: Chapman and Hall.

Hawkins, D. (1993). The accuracy of elemental set approximations for regression. *Journal of American Statistical Association*, *88*, 580-589.

Hawkins, D., & Olive, D. (2002). Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algorithm. *Journal of the American Statistical Association*, *97*, 136-148.

Mayo, M., & Gray, B. (1997). Elemental subsets: the building blocks of regression. *The American Statistician*, *51*, 122-129.

Mayo, M., & Gray, B. (2001). The robustness and efficiency of trimmed elemental estimation in regression analysis: a Monte Carlo simulation study. *Probabilistic Engineering Mechanics*, *16*, 323-330.

Rousseeuw, P., & Bassett, G. (1991). Robustness of the p-subset algorithm for regression with higher breakdown points. In W. Stahel (Ed.), *Robustness, Diagnostics, Computing and Graphics in* Statistics. NY: Springer.

Rubin, D. (1980). Composite points in weighted least squares regression. *Technometrics*, *22*, 343-348.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*, 1379-1389.

Sheynin, O.B. R.J. (1973). Boscovich's work on probability. *Archive for History of Exact Sciences*, *9*, 306-324.

Theil, H. (1950). A rank invariant method of linear and polynomial regression analysis III. *Proceedings of the Royal Society A*, *53*, 1397-1412.

Welsch, R. (1986). Comment. *Statistical Science*, *1*, 403-405.