5-2018

Evaluation of Using the Bootstrap Procedure to Estimate the Population Variance

Nghia Trong Nguyen
Stephen F Austin State University, nghianguyen94rand@yahoo.com

Evaluation of Using the Bootstrap Procedure to Estimate the Population Variance

By

NGHIA TRONG NGUYEN, Bachelor of Science

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

Of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

May 2018

Evaluation of Using the Bootstrap Procedure to Estimate the Population Variance

By

NGHIA TRONG NGUYEN, Bachelor of Science

APPROVED:

Robert K. Henderson, Thesis Director

Gregory K. Miller, Committee Member

Jacob Turner, Committee Member

Emiliano Giudici, Committee Member

Pauline Sampson, Ph.D.
Dean of Research and Graduate Studies

Abstract

The bootstrap procedure is widely used in nonparametric statistics to generate an empirical sampling distribution from a given sample data set for a statistic of interest. Generally, the results are good for location parameters such as population mean, median, and even for estimating a population correlation. However, the results for a population variance, which is a spread parameter, are not as good due to the resampling nature of the bootstrap method. Bootstrap samples are constructed using sampling with replacement; consequently, groups of observations with zero variance manifest in these samples. As a result, a bootstrap variance estimator will carry a bias to the low side. This work will attempt to demonstrate the bias issue with simulations, as well as explore possible approaches to correct for any such bias. In addition, these approaches will be evaluated for more general performance through simulations.

For Mother

Acknowledgement

This project would not have been possible without the support of many people. Many thanks to my adviser, Dr. Robert Henderson, for reading my revisions, offered guidance, and helped making sense of my many confusions. Also thanks to my committee members, Dr. Greg Miller, Dr. Jacob Turner, and Dr. Emiliano Giudici, who offered guidance and support. Thanks to my friends, the graduate students at the Department of Mathematics and Statistics 2016-2018, for always offering support and love. Special thanks to my best friend, Vicente Sanchez, for proofreading and support throughout my journey. Finally, thanks to my amazing family who endured this long process with me, always offering support and love.

Table of Content

List of Figures

List of Tables

Introduction

In this data-driven age, the following scenario is rare, but not entirely impossible: Given a random sample collected through a study or experiment, from a population with unknown probability distribution, is there a way to estimate a population parameter and construct a confidence interval for it? Moreover, is this possible if knowledge of the sampling distribution of the statistic is also unknown, and little to no prior knowledge is available to make any intelligent assumption about it (Efron, B. 1979)?

As an example of such a situation, consider a process engineer attempting to understand his process variation. The engineer might need an estimate of this to establish process control limits, evaluate conformance of product to specifications, or to determine if the current process variation is now different from some past historical experience. For example, the engineer might want to know if a new process improvement, or a new set of process equipment, has resulted in a truly less variable process and consequently a more consistent product. However, the available data to do this may be relatively sparse and not very well-behaved.

Such situations are the leading motivation for many simulation techniques in nonparametric statistics such as the jackknife method, delta method, crossvalidation method, subsampling method, and, the most popular method among them, the bootstrap method. The term “bootstrapping” originates from the idea of “pulling oneself up by one’s bootstrap”, this is especially true in this statistical situation where there is no exact solution to the problem (Chernick, M. and LaBudde, R. 2011).

What is a Bootstrap?

The bootstrap approach relies primarily on repeatedly re-sampling the original sample with replacement, with calculation of the statistic of interest for each re-sample. The resultant re-sample statistics then form an empirical estimate of the sampling distribution for this statistic.

Consider a random sample of size n from an unknown probability distribution: $x_1, \ldots, x_n$. Since this is a random sample, these observations are assumed to be independently and identically distributed. Now, imagine that while the probability distribution from which this sample was drawn is unknown, that the desire of the experimenter is to estimate a parameter of this distribution: θ. Further, assume that an estimate of this parameter that is a function of the sample observations - $\hat{\theta}$ is available.

The bootstrap approach to generating an estimate of θ would involve repeatedly sampling the original n observations with replacement. A typical such sample can be identified as $x_{1b}^*, \ldots, x_{nb}^*$, where b = 1, …, B = the total number of bootstrap samples to be considered. Efron (Efron, B. 1979) suggests that B be a value between 200 and 500; however, often B is 1000 or more with currently readily available computing and processing power.

With each bootstrap sample, an estimate of $\theta = \hat{\theta}_b^*$ can be obtained. A natural resulting point estimator of θ would then be given by:

$$\hat{\theta}_B = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}_b^*. \qquad (1)$$

In addition, and perhaps more importantly, a 100(1-α)% confidence interval estimate of θ can be determined from the α_1^{th} and α_2^{th} percentiles of the $\hat{\theta}_b^*$ values:

$$\hat{\theta}_{B_{(\alpha_1)}}^* \text{ to } \hat{\theta}_{B_{(\alpha_2)}}^*, \qquad (2)$$

where $\alpha_2 - \alpha_1 = 1 - \alpha$. Generally, $\alpha_1 = \frac{\alpha}{2}$ and $\alpha_2 = 1 - \frac{\alpha}{2}$.

The preceding paragraph essentially outlies the percentile method of obtaining a bootstrap interval estimate of an unknown parameter θ (Hollander, M. & Wolfe, D. A. 1999).

A similar, but slightly different bootstrap interval can be obtained using the residual method (Hollander, M. & Wolfe, D. A. 1999). With the observations

$x_1, \dots, x_n$, first compute the estimate $\hat{\theta}$ of the parameter of interest θ. Proceed with the bootstrap method as described above with a desired number of resamples B, and obtain an estimator $\hat{\theta}_b^*$ from each bootstrap replication for b = 1, …, B.

The residuals are calculated as follows:

$$e_b = \hat{\theta}_b^* - \hat{\theta}, \tag{3}$$

for b = 1, …, B. If $\hat{\theta}$ is an unbiased estimator of the unknown parameter θ, it follows that

$$e_b \approx \varepsilon_b = \hat{\theta}_b^* - \theta \tag{4}$$

for b = 1, …, B are indeed the residuals from each of the bootstrap replication.

Consequently, the bootstrap approach treats the observed sample as the "population" to derive an empirical sampling distribution for an estimator of an original population parameter. At first glance, the idea seems to be erroneous since a single realization of the population in the form of the sample may not accurately represent the underlying population, especially if extreme outliers are observed. However, researchers have shown that the bootstrapped confidence intervals for a population's measure-of-central-tendency parameter such as its mean, median, etc. do indeed cover the actual values with the stated confidence levels (Efron, B. 1979). These results are primarily based on the most powerful theorem in statistics, the Central Limit Theorem (Hogg, Robert V.; McKean, Joseph W.; and Craig, Allen T. 2013).

Objective

The challenge lies in the construction of confidence intervals for a population's measure-of-dispersion parameters such as its standard deviation, variance, etc. The re-sampling involved in bootstrapping, will tend to produce groups of repeated results. Each of these groups will have zero variation; consequently, the variances of those bootstrap samples are expected to be biased low compared to the variance of the relevant population. This, in turn, can lead to the actual coverage rates of confidence intervals for a variance parameter to be less than their stated nominal coverages.

Therefore, this work will seek to evaluate approaches to correct for the expected bias inherent in a bootstrap estimator of variance, as well as to move the actual coverage rates of corresponding confidence intervals closer to their stated nominal confidence coefficients.

Historical Notes

It appears many researchers have noted the low bias issue with bootstrap estimates of variances; however, few have offered clear and efficient solutions. Chernick and LaBudde (Chernick, M. and LaBudde, R. 2011) demonstrated the under-coverage rates of the bootstrap confidence intervals for the standard deviation of different underlying distributions. They specifically considered the Gamma(2, 3), Uniform(0, 1), Student ' s t with 5 degrees of freedom, Normal(0, 1), and Lognormal(0, 1) distributions. For the Lognormal(0, 1) distribution, which was the worst case, bootstrap intervals with stated and desired coverage rates of 90% and 95%, often only achieved actual coverages near 60%. For the specified nominal coverage probabilities considered, the Monte Carlo estimates of the actual coverage probabilities for the various intervals were the closest for the Uniform(0,1) distribution, the best case in relation to the other distributions listed, of 86.8% and 92% actual coverages for 90% and 95% nominal coverages, respectively. A similar example will be explored in the following section.

Example to Demonstrate the Issue

To prevent confusion due to the mixture of technical words and the English words, "replications" will always refer to the resamples obtained from the bootstrap process, and "trials" will refer to the repetitions of drawing a random sample of size n from an established and known distribution for simulation purposes.

In this example, 1000 trials of size 20 were drawn from a standard normal distribution with mean 0 and standard deviation (also variance) of 1. Then, for each trial, 1000 bootstrap replications (i.e., re-samples) were obtained. The two parameters to estimate in this example are the mean $(\mu = 0)$ and the variance $(\sigma^2 = 1)$. Clearly, the exact values are known here only because this is a simulation situation. They will rarely be known in a physical setting.

The first parameter, μ, is traditionally estimated with the statistic $\bar{X} = \frac{1}{20}\sum_{i=1}^{20} x_i$, the sample average when n = 20 (as in this simulation). Consequently, letting $\theta = \mu$, $\hat{\theta} = \bar{X}$, and B = 1000, using equations (1) and (2), both point and interval bootstrap estimates can be obtained for each of the 1000 trial samples. The point estimate so obtained is the mean statistic for each trial, and the 2.5th and 97.5th

percentiles of the bootstrap sample means provide an interval estimate with α = 0.05 (i.e., a 95% confidence interval) for $\theta = \mu$.

The second parameter to be estimated is the variance, σ^2. This parameter traditionally is estimated using the sample variance:

$$S^2 = \frac{1}{19}\sum_{i=1}^{20}(x_i - \bar{X})^2. \quad (5)$$

Equation (5) reflects the simulation sample size of n = 20.

Now, letting $\theta = \sigma^2$, $\hat{\theta} = S^2$, and, again, B = 1000, equations (1) and (2) can be used to provide bootstrap point and interval estimators of the parameter σ^2. Again, such estimates can be obtained across all 1000 trial samples, with the mean across replications at each trial providing a point estimate, and the 2.5th and 97.5th values across trials providing an approximate 95% bootstrap confidence interval for σ^2.

Figure 1 displays the bootstrap point and approximate 95% confidence interval estimates for the population mean across the 1000 trial samples using the percentile method. Note that the trial results have been ordered from smallest to largest by the respective point estimate. Also, observe that the upper limit falls below the actual parameter value (i.e., $\mu = 0$), and the lower limit lies above this value for several of the trials at each end of the figure. This is to be expected as the expected coverage rate for these intervals is only 95%. Now, note that the curve displaying the bootstrap point estimates crosses zero (the actual parameter value) very near the 500th ordered trial consistent with it being an unbiased estimator of the population mean.

Figure 2 is analogous to Figure 1; however, the parameter being estimated is the population variance (i.e., $\sigma^2 = 1$). While analogous, there are some distinct differences. Perhaps most notable is that the point estimate curve does not cross the actual parameter value until somewhere near the 600th ordered trial.

Figure 1: 1000 95% Percentile Bootstrap Confidence Intervals for the Mean (Samples of Size n = 20 from a Standard Normal Distribution) (Each Trial Sample Used B =1000 Bootstrap Re-Samples)

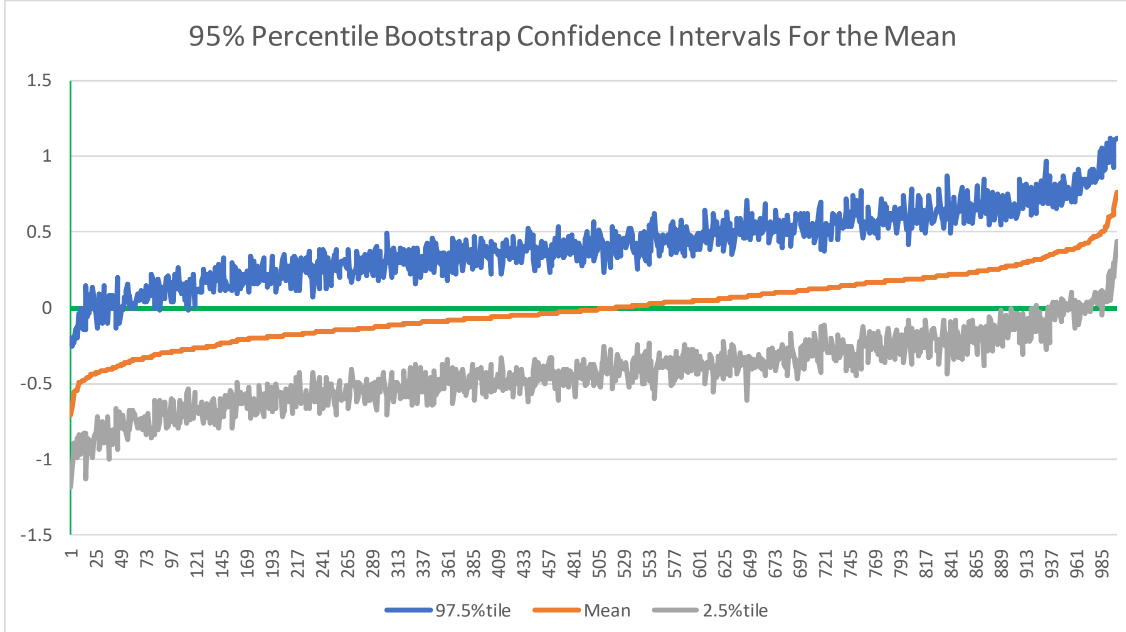
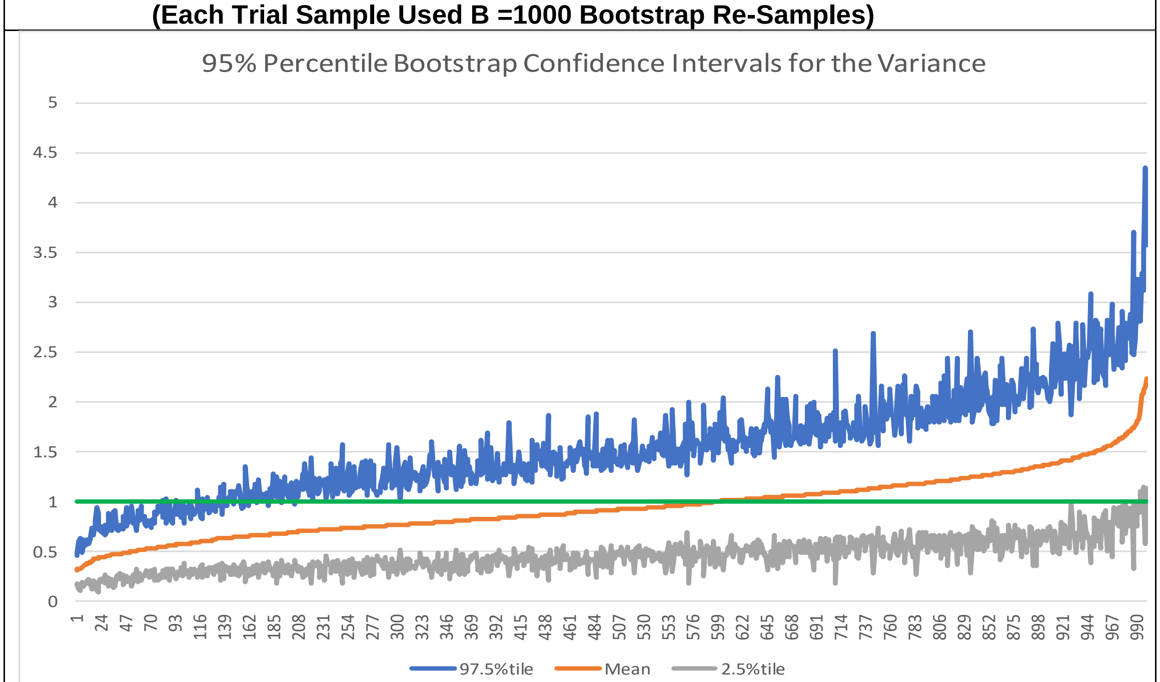


Figure 2: 1000 95% Percentile Bootstrap Confidence Intervals for the Variance (Samples of Size n = 20 from a Standard Normal Distribution) (Each Trial Sample Used B =1000 Bootstrap Re-Samples)



This indicates that the percentile bootstrap estimator of the variance tends to underestimate the true population variance (i.e., this estimator is biased low). An additional indication of this low bias is the relatively large number of upper confidence limits falling below the actual parameter value while very few lower limits lie above it.

The actual coverage rates of these 95% bootstrap confidence intervals – the number of intervals that did not capture the true value of the mean of 0 or the variance of 1 – are displayed in Table 1. In this table, the "Missed High" column displays the number of intervals where the lower confidence limit was larger than the actual value of the parameter to be estimated. The "Missed Low" column displays the number of intervals where the upper confidence limit was less than the actual parameter value. The "Coverage" column indicates how many of the 1000 intervals captured the true value of the respective parameter.

Table 1: Summary of 1000 95% Percentile Bootstrap Confidence Intervals
(Samples of Size n = 20 from a Standard Normal Distribution)
(Each Trial Sample Used B =1000 Bootstrap Re-Samples)

Parameter	Missed High	Missed Low	Coverage
Mean	44	34	922
Variance	4	138	858

As noted earlier, the confidence intervals for the population variance tend to be biased low, and the actual coverage rate of the stated 95% bootstrap confidence intervals of the population variance is only 85.8%, from

(1000 – 4 – 138)%, in this simulation. This low coverage rate will be utilized as the reference rate in comparing various approaches to address this issue in the following sections.

Meanwhile, the actual coverage rate of the stated 95% bootstrap confidence intervals of the population mean is 92.2%. While it is not 95%, this rate is much closer to the nominal coverage rate than the actual coverage rate for the variance intervals.

Potential Approaches

Approach 1: Adjusting by Expected Bias

For the simple one population case, the model is $y_i = \mu + \varepsilon_i$, where y_i denotes the observed data point, μ denotes the mean of the underlying population, and ε_i denotes the noise, for $i = 1, \ldots, n$. Under the non-bootstrap situation, the following are assumed:

1) The expected value of the noise is 0, or $E[\varepsilon_i] = 0$,
2) The expected value of the variation in the noise is the variance of the underlying population, or $E[{\varepsilon_i}^2] = \sigma^2$, and
3) The specific noise values are uncorrelated from one observation to the next, or

   $E[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$.

Given the above assumptions:

$$E[y_i] = E[\mu + \varepsilon_i] = \mu + E[\varepsilon_i] = \mu, \tag{6}$$

and

$$\begin{aligned} Var(y_i) &= E[(y_i - \mu)^2] \\ &= E\left[y_i^2 - 2\mu y_i + \mu^2\right] \\ &= E\left[y_i^2\right] - 2\mu E[y_i] + \mu^2 \\ &= E[(\mu + \varepsilon_i)^2] - 2\mu^2 + \mu^2 \end{aligned}$$

$$= E[\mu^2] + 2\mu E[\varepsilon_i] + E\left[\varepsilon_i^2\right] - \mu^2$$

$$= \mu^2 + \sigma^2 - \mu^2$$

$$= \sigma^2. \qquad (7)$$

Given a random sample of $y_1, \ldots, y_n$ of n values from the model described above, the expected values of the traditional estimators for the parameters μ and σ^2 (i.e., $\hat{\mu} = \bar{Y}$ and $\widehat{\sigma^2} = S^2$) are then given by:

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum_{i=1}^n y_i\right] = \frac{1}{n}E[\sum_{i=1}^n y_i] = \frac{1}{n}\sum_{i=1}^n E[y_i] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}(n\mu) = \mu, \ (8)$$

and

$$E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{Y})^2\right]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^n (y_i^2 - 2\bar{Y}y_i + \bar{Y}^2)]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^n y_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2]$$

$$= \frac{1}{n-1}(\sum_{i=1}^n E[y_i^2] - nE[\bar{Y}^2])$$

$$= \frac{1}{n-1}\{n(\mu^2 + \sigma^2) - \frac{1}{n}E[(\sum_{i=1}^n y_i)^2]\}$$

$$= \frac{1}{n-1}\{n(\mu^2 + \sigma^2) - \frac{1}{n}\left(\sum_{i=1}^n E[y_i^2] + \sum_{i=1}^n \sum_{j\neq i}^n E\left[y_i y_j\right]\right)\}$$

$$= \frac{1}{n-1}\{n(\mu^2 + \sigma^2) - \frac{1}{n}n(\mu^2 + \sigma^2) - \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i}^n E\left[(\mu + \varepsilon_i)\left(\mu + \varepsilon_j\right)\right]\}$$

$$= \mu^2 + \sigma^2 - \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i}^n \left(\mu^2 + \mu E[\varepsilon_i] + \mu E\left[\varepsilon_j\right] + E\left[\varepsilon_i \varepsilon_j\right]\right) \qquad (9)$$

$$= \mu^2 + \sigma^2 - \frac{1}{n(n-1)}n(n-1)\mu^2$$

$$= \sigma^2. \quad (10)$$

Hence, both estimators are unbiased.

Under the model assumptions, and using only the original sample, all the expectations in equation (9) are zero. However, when bootstrap re-samples $y_1^*, \ldots, y_n^*$ of the original sample are considered, the last expectation in equation (9) is not zero if $y_j^* = y_i^*$.

Consequently, the expected value of the sample variance of a bootstrap re-sample of the original sample is given as:

$$E[S_b^2] = \sigma^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} E[\varepsilon_i^* \varepsilon_j^*],\ \varepsilon_i^* = y_i^* - \mu, i = 1, \ldots, n \quad (11)$$

In the bootstrapping re-sampling with replacement process, it is possible that even with j ≠ i, $\varepsilon_i^* = \varepsilon_j^*$. This occurs whenever the i^th^ re-sampled value is the same result as the j^th^ re-sampled value. In re-sampling with replacement, the probability that this occurs is $\frac{1}{n}$, and $E\left[\varepsilon_i^* \varepsilon_j^*\right] = \sigma^2$. When the i^th^ and j^th^ re-sampled points are not the same observation, then $E\left[\varepsilon_i^* \varepsilon_j^*\right] = 0$. Therefore,

$$E\left[\varepsilon_i^* \varepsilon_j^*\right] = \sigma^2 \left(\frac{1}{n}\right) + 0 \left(\frac{n-1}{n}\right) = \frac{\sigma^2}{n} \quad . \quad (12)$$

and

$$E[S_b^2] = \sigma^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\sigma^2}{n}$$

$$= \sigma^2 - \frac{1}{n(n-1)} n(n-1) \frac{\sigma^2}{n}$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2. \quad (13)$$

Consequently, a bootstrap re-sample sample variance is expected to be biased low by a factor of $\frac{n-1}{n}$; hence, a simple adjustment of multiplying S_b^2 by $\frac{n}{n-1}$ would be expected to provide an unbiased estimate of σ^2. The multiplier $\frac{n}{n-1}$ was found through the derivations above (Brennan, Robert L. 2007) to provide a suitable bias correction for a variance estimate obtained via a bootstrap procedure.

Approach 2: Bias Corrected and Accelerated Method (BCa)

The Bias Corrected and Accelerated (BCa) method (Efron and Tibshriani 1993) uses a form of bias correction and a jackknife approach to adjust for skewness (i.e., an "acceleration" factor). The bias correction factor is determined from the empirical bootstrap distribution as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{[\# \, \widehat{w}_b < \widehat{w}]}{B}\right), \quad (14)$$

where

- $\widehat{w}_b$ is the estimate of the parameter of interest from the bootstrap replication b, for $b = 1, \ldots, B$,

- $\widehat{w}$ is the estimate of the parameter of interest from the original sample of size n, and
- $\Phi^{-1}(p) = z_p$, with $\int_{-\infty}^{z_p} \left(\frac{1}{\sqrt{2\pi}}\right) e^{\frac{-t^2}{2}} dt = p$ the inverse function of the standard normal distribution.

The acceleration factor, which is based on the third sample moment (Chernick, M. R. and LaBudde, R. 2011), is given as

$$\hat{a} = \frac{\sum_{i=1}^{n} (\widehat{w}_{.} - \widehat{w}_{-i})^3}{6\left[\sum_{i=1}^{n} (\widehat{w}_{.} - \widehat{w}_{-i})^2\right]^{3/2}}, \tag{15}$$

where

- $\widehat{w}_{-i}$ is the estimator of w, the parameter of interest, without the i[th] data point Y_i, and so it uses only $Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$, and
- $\widehat{w}_{.} = \sum_{i=1}^{n} \frac{\widehat{w}_{-i}}{n}$.

This approach then uses

$$L_a = \frac{\hat{z}_0 + (\hat{z}_0 + z_{\alpha/2})}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})} \tag{16a}$$

and

$$U_a = \frac{\hat{z}_0 + (\hat{z}_0 + z_{1-\alpha/2})}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \tag{16b}$$

to obtain the percentiles:

$$p_{L_a} = \Phi(L_a) \quad \text{and} \quad p_{U_a} = \Phi(U_a).$$

Finally, the ~100(1-α)% BCA bootstrap confidence interval is given as:

$$\hat{w}_{p_{L_a}} \quad \text{to} \quad \hat{w}_{p_{U_a}} \; .$$

Since the theoretical distribution of the variance estimator, s^2, has a chi-square distribution, which is skewed to the right, the BCa method ideally would provide a suitable bias correction for the bootstrap variance estimator.

Approach 3: Shocking Bootstrap Resampled Observations

Another potential approach to correct for inherent bias in the bootstrap variance estimator would be to adjust the repeated data points in the bootstrap replications by a small "shock" or perturbation so that the variance between any two or more repeated data points is no longer zero. To minimize disruption of the original data, this "shock" could be obtained from a symmetric distribution centered at 0 with a suitable amount of spread.

Ostensibly, a suitable amount of spread could produce perturbations that increase the variance estimate and coverage rate while not seriously disturbing the original data. Too little spread would likely not account for the magnitude of the biases involved; however, too much spread would likely produce overestimates of the variance and coverage rates that diverge even further from the desired nominal level.

The shock idea is basically to stretch the confidence intervals by adding more noise so that more of them will include the true variance while hopefully adjust for the low bias of the point estimate. Thus, in addressing the coverage issue, it is possible that the bias problem will be indirectly resolved.

Shock Types and Sizes

In an initial attempt to determine an appropriate magnitude for the induced disturbances, shock sizes were chosen using uniform distributions from -0.01 to 0.01, -0.05 to 0.05, and -0.10 to 0.10 to limit disturbance of the original data with the extreme values having equally likely chances of being used. However, the resulting coverages for the example previously stated changed to about 87.1 to 87.5 for all three initial shock sizes, which were not significantly different from the reference coverage of 85.8%. Because the results above were obtained by shocking all the bootstrapped resampled points, shocking just the repeated values will have a lesser effect. Hence, this approach to setting shocks was not pursued.

However, these initial observations effectively indicated that a larger shock size was necessary. As a result, a variety of potential perturbation approaches were considered with varying shock sizes.

Uniformly Distributed Disturbances

The initial attempts at defining a range for uniformly distributed disturbances produced ranges that were clearly too small to generate the desired increase in variation within the bootstrap resamples. In order to establish more suitable (i.e., larger) ranges, multiples of an original sample estimate of the population standard deviation were considered.

The initial, and perhaps most obvious, estimator considered was the original sample standard deviation. However, understanding that generally bootstrap estimators will only be considered in situations where the data is either not well-behaved and/or the analyst is unwilling to adopt an assumption of a normality, a more robust estimator of standard deviation was also determined to be worthy of consideration.

The sample Inter-Quartile Range (IQR) is known to be a simple and robust estimator of the spread in the population distribution from which the sample was obtained. The IQR is simply the difference between the third sample quartile (Q_3)

and the first sample quartile (Q_1), and as such is not seriously impacted by extreme observations in the sample data, or the likelihood of their appearance when data is sampled from highly skewed or very heavily tailed population distribution.

If, on the other hand, the underlying population distribution happens to be a normal distribution, then the IQR can provide an estimate of the population standard deviation by setting it equal to its theoretical value:

$$IQR = Q_3 - Q_1 = z_{0.75}\sigma - z_{0.25}\sigma \approx 0.6745\sigma - (-0.6745)\sigma = 1.35\sigma,$$

(Note: z_p determined from $\int_{-\infty}^{z_p} \left(\frac{1}{\sqrt{2\pi}}\right) e^{\frac{-t^2}{2}} dt = p$ and the IQR value used in the calculations later in the simulations was 1.3489795 instead of 1.35, but for simplicity sake, 1.35 provides a useful approximate value.) producing an estimator of the population standard deviation given as:

$$\hat{\sigma}_{IQR} = \frac{IQR}{1.35}. \qquad (16)$$

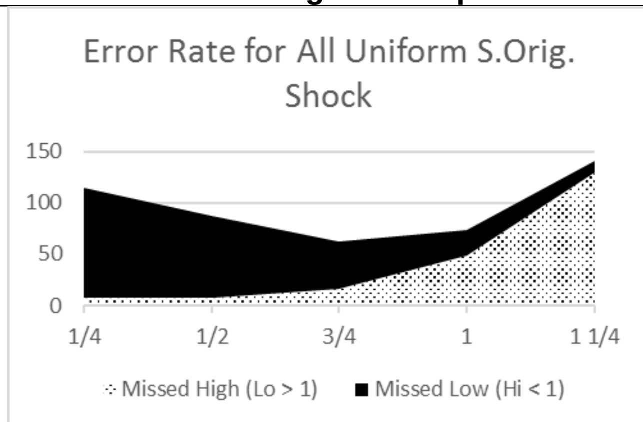
Since both S and $\hat{\sigma}_{IQR}$ can be obtained for the original sample, either can be used to assist in establishing a suitable range for uniformly distributed disturbances to be applied to the bootstrap sample results. The goal would be to determine a proper multiple of either of these estimators to produce disturbances of a sufficient magnitude to both

- reduce the bias in the bootstrap point estimator, and,

- more importantly (since the primary value of bootstrap estimators is in providing interval estimates), increase the coverage level of the bootstrap intervals to value much closer to the nominally stated level.

Initial exploratory work, the result of the shocking All the bootstrap results using a Uniform distribution for the spread with 1/2 times the standard deviation of the original data (S.Orig) as the shock size (i.e. Uniform(-1/2*S.Orig., 1/2*S.Orig)) yielded 8 Missed High and 79 Missed Low intervals. When changing the shock distribution to Uniform(-S.Orig, S.Orig), 48 intervals Missed High and 26 Missed Low. This phenomenon also occurred for other shocks, suggesting that there is a certain multiplier of the original standard deviation estimate that will balance out the number of intervals that Miss High and Low. Hence, values of a Multiplier were evaluated between 1/4 and 5/4 since a value near 3/4 was believed to be near an optimal value. Figure 3 displays the number of intervals missing both high and low, as well as the sum of these, or the total number of intervals failing to cover the actual variance for Multipliers covering the 1/4 to 5/4 range noted above. The fewest overall errors (i.e., the highest coverage rate does occur when the Multiplier = 3/4.

Figure 3: Error Rate (Sum of # Missed High and # Missed Low) out of 1000 trials with Shocks Based on the Original Sample Standard Deviation



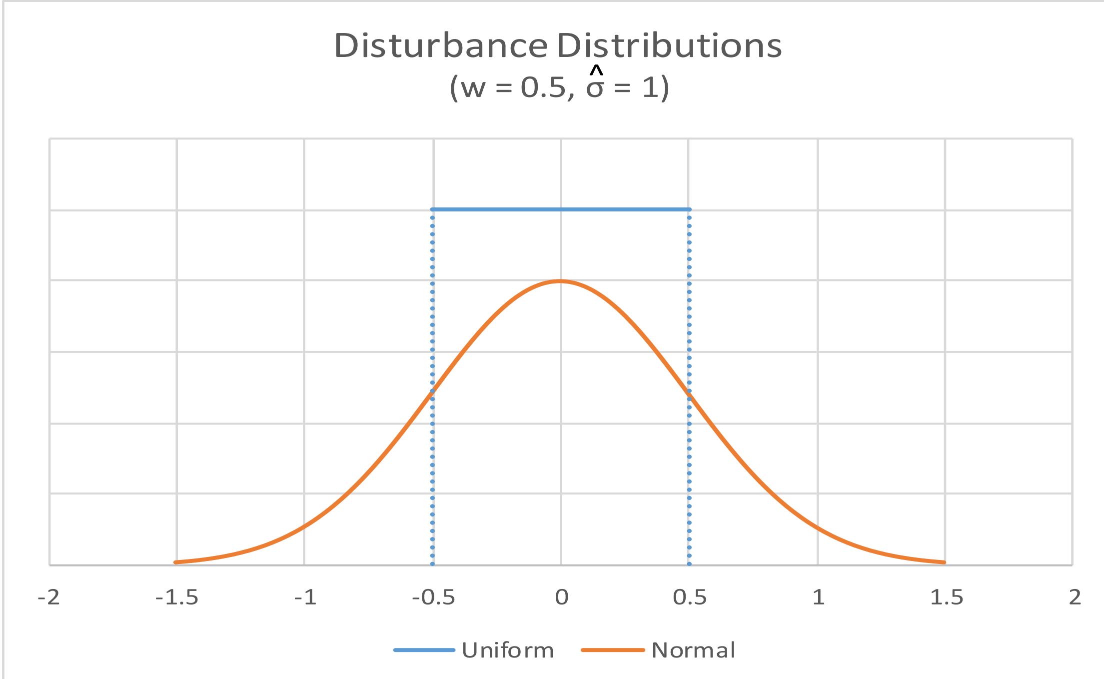
The multiples, w, that were explicitly evaluated were from w = 0.25, 0.50, …, 1.25 of the respective estimate of standard deviation, either S or IQR. This resulted in consideration of uniformly distributed disturbances centered at zero with ranges extending from $-w\hat{\sigma}$ to $+w\hat{\sigma}$, where $\hat{\sigma} = S$, or $\hat{\sigma} = \frac{IQR}{1.35}$. As discussed, multiples that are too small will be unlikely to sufficiently improve the bootstrap estimates. However, a multiple that is too large will likely add too much variation, and essentially overcompensate for the low bootstrap bias and produce estimates biased high.

Normally Distributed Disturbances

In addition to considering uniformly distributed disturbances, or shocks, normally distributed shocks also were explored. These distributions were centered at zero, and the standard deviations were again established using multiples of the original sample estimates of the population standard deviation.

Figure 5 displays generic disturbance distributions of both types. Note that for the same Multiplier value, w, and same original sample estimate of standard deviation, $\hat{\sigma}$, the actual magnitude of any shock is relatively restricted when using uniformly distributed disturbances. In fact, the actual shock magnitudes can be over twice as large when using normally distributed disturbances than when using uniformly distributed shocks.

Figure 4: Generic Disturbance Distributions



The table below displays the different scenarios being evaluated in the simulation process. (Note: The encrypted scenario numbers 1 to 40 are for coding purposes only, and are not ordinal or ranking values.)

Table 2: Scenarios for Shock Types and Sizes Evaluated for Approach 3

Scenario	Numbers	All		Repeated	
*	Shock Size	Distribution			
		Uniform	Normal	Uniform	Normal
Multiplier	1/4*S.Orig.	1	11	21	31
	2/4*S.Orig.	2	12	22	32
	3/4*S.Orig.	3	13	23	33
	4/4*S.Orig.	4	14	24	34
	5/4*S.Orig.	5	15	25	35
	1/4*IQR.Orig.	6	16	26	36
	2/4*IQR.Orig.	7	17	27	37
	3/4*IQR.Orig.	8	18	28	38
	4/4*IQR.Orig.	9	19	29	39
	5/4*IQR.Orig.	10	20	30	40

The R code used to conduct this experimental design can be found in the R Codes section.

Approach 4: Using Inter-Quartile Range Estimate

Because the IQR was believed to be a more robust measure of spread (Whaley, Dewey L. 2005) and the 95% bootstrapped IQR confidence interval yielded a coverage of 96.4% in the simulation study, another idea is to simply bootstrap the IQR of the original data and then scale it to a variance estimate by $\hat{\sigma}^2 = \left(\frac{IQR}{1.35}\right)^2$, such that the percentiles:

$$p_{L_a} = \left(\frac{IQR_{(a,B)}}{1.35}\right)^2 \quad \text{and} \quad p_{U_a} = \left(\frac{IQR_{(1-a,B)}}{1.35}\right)^2$$

For data that are from an approximate normal distribution, this approach should yield coverages of the variance estimate much closer to the nominal confidence level in relation to just bootstrapping the variance itself. Even if the data is assumed to not be from an approximately normal distribution: if the sample size is large, then the Central Limit Theorem will apply, and if the sample is small, then the available tests for normality would not have much power to detect a difference. In the example discussed, this approach gave a coverage of 96.4%, as noted above.

Combining Approaches 1 & 3

Because Approach 1 was derived theoretically to correct for the bias of the bootstrap variance estimate, and Approach 3 was proposed primarily to correct for the coverage issue, potentially the combination of these approaches would address both problems simultaneously.

To evaluate the approaches for performance, simulations will be conducted.

Simulation Specifics

Different Sample Sizes

Since for larger samples, a common sample variance (i.e., S^2) will converge in probability to σ^2, the simulation here will focus on situations where only relatively smaller sample sizes are available. The example utilized previously to demonstrate the low bias and coverage issues for the bootstrap variance estimator involved a sample size of 20. To assess the impact of varying sample sizes, sample sizes of 10 and 30 will also be evaluated in these simulations.

Different Underlying Distributions

The approaches above were designed considering only the most well-behaved data, data from a standard normal distribution. However, because bootstrap is a nonparametric method, it is more likely to be considered and utilized for data that are not well-behaved (i.e., not necessarily normally distributed). Examples of such distributional models include skewed distributions such as the exponential distribution or the log normal distribution. In addition, they might possibly be symmetric, but heavy-tailed distributions such as the double exponential (also known as Laplace) distribution.

Therefore, the simulations done in R consisted of the four different distributions described in more detail below with each evaluated for three different sample sizes. For each distribution, sample size combination, 1000 random samples were obtained (i.e., trials). For each trial, the minimum, 2,5th percentile, 25th percentile, mean, median, 75th percentile, 97.5th percentile, and maximum values across the 1000 bootstrap replications were obtained for both the sample mean and the sample variance. In addition, the number of times this primary sample statistic of interest (i.e., the sample mean or the sample variance) across the replications was less than the original sample statistic was calculated. The R statistical package was utilized to obtain the samples and calculations of the summary statistics for each replication of each trial. These results were then stored in Excel through R export commands for further analysis.

Standard Normal Distribution

Let $X_1, \ldots, X_n$ represent a random sample from a normal distribution with density function given as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{\sigma^2}}, \text{ for } \sigma > 0.$$

For normally distributed simulation results, the two density parameters, μ and σ were set to 0 and 1, respectively.

Shifted Exponential Distribution

Let $X_1, \dots, X_n$ represent a random sample from an exponential distribution with unknown location parameter θ and scale parameter λ, and having the following probability density function

$$f(x) = \lambda e^{-\lambda x}\text{, for } \lambda > 0 \text{ and } x > 0.$$

The mean and variance of this distribution can be found by integration by parts as λ^{-1} and λ^{-2}, respectively. Consequently, choosing a lambda value of 1 produces an exponential distribution where the mean and variance are both 1. However, to be consistent with the standard normal population mean and variance, this exponential distribution (λ = 1) was shifted by 1, producing a shifted exponential with mean $\lambda^{-1} - 1 = 0$ and variance as $\lambda^{-2} = 1$. This shift does not affect the spread of the distribution. Hence, simulated exponentially distributed results were obtained from the probability density

$$f(x) = e^{-(x+1)}\text{, for } x > -1.$$

To find the true IQR of this distribution for coverage assessment, we simply solve $.75 = 1 - e^{-\lambda x_{.75}}$ and $.25 = 1 - e^{-\lambda x_{.25}}$ and take the difference between $x_{.75}$ and $x_{.25}$, with $\lambda = 1$. This results in

$$IQR = -\lambda \ln .25 + \lambda \ln .75 = 1.098612289.$$

Double Exponential Distribution

Let $X_1, \dots, X_n$ represent a random sample from a double exponential distribution or Laplace distribution with unknown location parameter μ and scale parameter λ, and having the following probability density function

$$f(x) = \frac{1}{2\lambda} e^{\frac{-|x-\mu|}{\lambda}}, \text{ for } \lambda > 0 \text{ and } -\infty < \mu < \infty.$$

The mean and variance of this distribution are μ and $2\lambda^2$, respectively. Hence, for simulated double exponential samples, setting $\mu = 0$ and $\lambda = \frac{1}{\sqrt{2}}$ will produce results with a mean of zero and variance of one; consistent with the other distributions considered.

The IQR is found by solving

$$.25 = \frac{1}{2} e^{\left(\frac{x_{.75}-\mu}{\lambda}\right)} \text{ and } .75 = 1 - \frac{1}{2} e^{\left(\frac{-x_{.25}-\mu}{\lambda}\right)} \text{ for } x_{.75} \text{ and } x_{.25}$$

and take the difference, or simply solve for one and double the result by exploiting the symmetric property of the double exponential distribution. These yield the same answer of

$$IQR = 2\left(\frac{\lambda \ln(-2(.75-1)+\mu)}{-1}\right) = 0.980258143.$$

Shifted Log Normal Distribution

Let $X_1, \dots, X_n$ represent a random sample from a log normal distribution with unknown location parameter μ and scale parameter σ, and having the following probability density function

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}}, \text{ for } \sigma > 0,\ x > 0, \text{ and } -\infty < \mu < \infty.$$

The mean and variance of this distribution are $e^{\mu+\frac{\sigma^2}{2}}$ and $e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$, respectively. Setting $\mu = 0$ and solving $1 = e^{2\sigma^2} - e^{\sigma^2}$ for σ produces a value of $\sigma = 0.693694331$. This, in turn produces a distribution mean of $e^{\frac{\sigma^2}{2}} = 1.272019649$. Since for comparison purposes, it was desirable for the mean and variance of the simulated lognormal results to also have a zero mean and variance of one, this log normal distribution needed to be shifted by 1.272019649. Hence, the simulation results were obtained from the shifted log normal distribution,

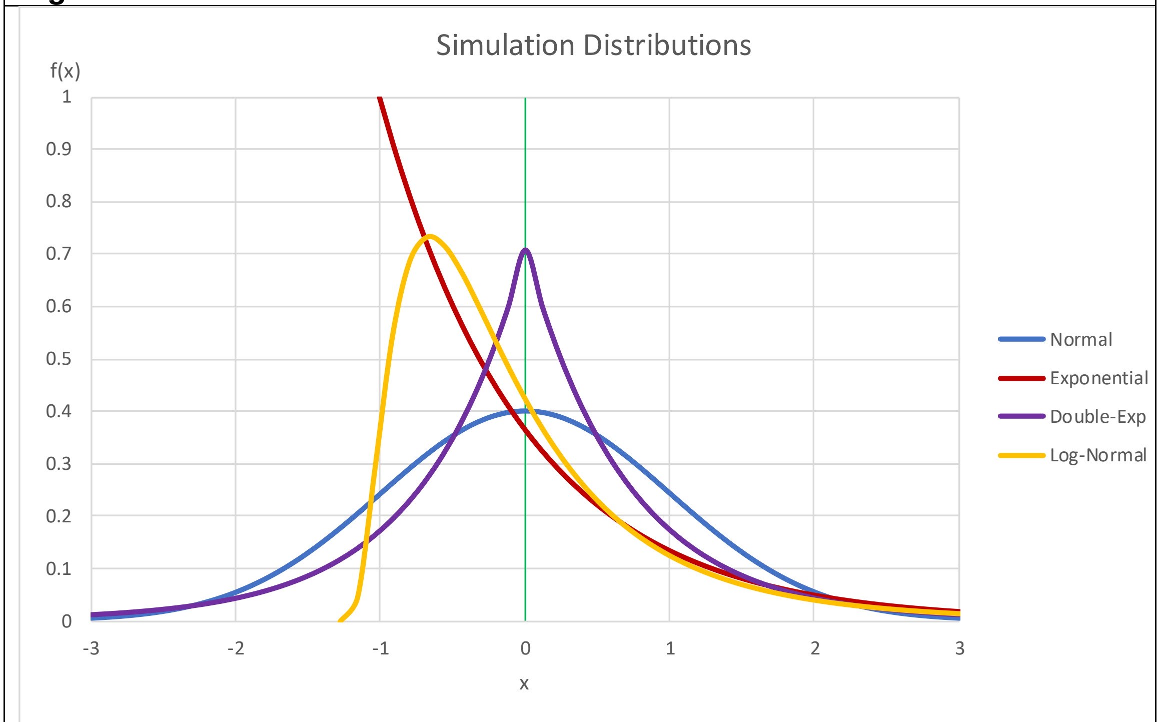
$$f(x) = \frac{1}{0.6937(x+1.272)\sqrt{2\pi}} e^{\frac{-(ln(x+1.272))^2}{2(0.6937)^2}}, \text{ for } x > -1.272.$$

Here, the true IQR is found by substituting $\mu = 0$ and $\sigma = 0.693694331$ into the equation (Whaley, Dewey L. 2005)

$$IQR = e^{\mu}(e^{.6745\sigma} - e^{-.6745\sigma}) = 0.970298718.$$

The simulation distributions are displayed in in Figure 5. Note that all the distributions have a mean of zero and a standard deviation of one; hence, vary only in shape.

Figure 5: Simulation Distributions



For the IQR calculations above, a priori knowledge of the actual underlying distribution would be required. Since this virtually never will be known in actual practice, the standardization of the IQR for Approach 4 will still use a divisor of 1.35. Again, the R code used to run this simulation procedure can be found in the R Codes section.

Evaluation of the Approaches

Reference

For each combination of sample size (n = 10, 20, 30) and probability distribution model (Normal, Exponential, Double-Exponential, and Log-Normal), 1000 trial samples were obtained. For each respective trial sample, B = 1000 bootstrap re-sample replications were generated.

For reference, or baseline comparison results against which to evaluate the remedial approaches discussed above, simple percentile bootstrap estimates of the population variance were obtained using the common sample variance. The summary values for each sample size, distribution combination that were considered were

- Bias = $\frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}_i^2 - 1$, where $\hat{\sigma}_i^2 = \frac{1}{B}\sum_{b=1}^{B}S_{b_i}^2$, $S_{b_i}^2$ = the sample variance for the bth bootstrap re-sample of the ith simulation trial sample. Hence, for each trial, $\hat{\sigma}_i^2$ is essentially the bootstrap point estimate for the population variance, and the Bias is the average (mean) of all these estimates across the 1000 simulation trials minus the actual true population distribution variance. Consequently, a negative Bias indicates that the estimate was lower than the truth, and vice versa.

- Coverage = 1000 – Miss_{Low} – $\text{Miss}_{\text{High}}$, where

  Miss_{Low} = Number of $S^2_{(975)_i} < 1$, and

  $\text{Miss}_{\text{High}}$ = Number of $S^2_{(25)_i} > 1$, for i = 1, …, 1000;

  where $S^2_{(b)_i}$ = the b^{th} order statistic of the B bootstrap re-sample sample variance estimates based on the i^{th} simulation trial sample. Hence, Coverage, again, is calculated by subtracting the number of 95% percentile bootstrap confidence intervals that failed to capture the true variance, both low and high, across the 1000 simulation trial samples.

- Confidence Interval Width = $\frac{1}{1000}\sum_{i=1}^{1000}(S^2_{(975)_i} - S^2_{(25)_i})$, which is equal to the average interval width across all the respective simulation trials.

- Mean Squared Error (MSE) = Bias^2 + Variance, where Bias is defined as above, and Variance = $\frac{1}{1000}\sum_{i=1}^{1000}\left[\hat{\sigma}_i^2 - \overline{\hat{\sigma}_i^2}\right]^2$, where $\overline{\hat{\sigma}_i^2} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}_i^2$, and, as above $\hat{\sigma}_i^2 = \frac{1}{B}\sum_{b=1}^{B} S^2_{b_i}$; which is the variance of the bootstrap point estimators of variance across all 1000 simulation trials.

The biases for the reference case are all negative as expected from the issue discussed with bootstrapping variance estimates, and they generally decrease in magnitude (get closer to 0, the desired bias value) as the sample size increases.

In addition, relative to the normal distribution, the observed bias was only smaller in magnitude for the Double-Exponential distribution at sample sizes 20 and 30. For this distribution and a sample size of 10, the magnitude of the bias was largest of any of those displayed in Table 3.

Table 3: Bias Results for Simple Percentile Bootstrap Variance Estimate

Distribution	Sample Size (n)		
	10	20	30
Normal	-0.0943	-0.0498	-0.0266
Double-Exponential	-0.1446	-0.0425	-0.0157
Exponential	-0.1016	-0.0536	-0.0407
Log-Normal	-0.1103	-0.0609	-0.0451

The coverage results for the percentile bootstrap variance 95% confidence intervals are displayed in Table 4. There is a clear ordering for coverage with the normal distribution always having the greatest coverage, the other symmetric distribution, the double-exponential next, followed by the two skewed distributions. The log-normal distribution has the consistently lowest coverage. However, none of the observed coverages was close to the nominal 95% confidence level, ranging from a low of 51% for the log-normal distribution at a sample size of 10; to a high of 89.7% for the normal distribution at a sample size of 30.

Table 4 also indicates an increase in coverage rates as the sample size increases. Perhaps to be expected, the rate of increase was larger for the distributions having the generally lowest coverages than for those with the higher coverage rates.

Table 4: Coverage Results for Simple Percentile Bootstrap Variance Estimate			
Distribution	Sample Size (n)		
	10	20	30
Normal	775	858	897
Double-Exponential	655	773	814
Exponential	580	695	743
Log-Normal	510	633	689

Table 5 (Confidence Interval Width) and Table 6 (MSE) display results consistent in nature with Table 4. The results are consistently lowest for the normal distribution and highest for the log-normal, with the symmetric double exponential providing narrower intervals and smaller MSE than the skewed exponential. The results are also consistent across sample size since, as expected, the intervals become narrower and the MSE results lower as sample sizes increase.

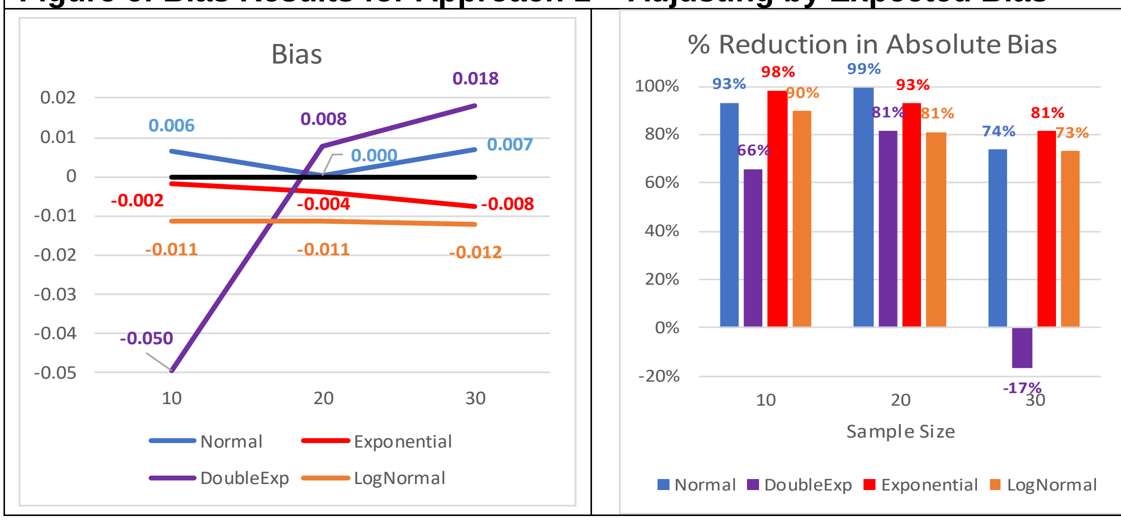
Table 5: Confidence Interval Width Results for Simple Percentile Bootstrap Variance Estimate			
Distribution	Sample Size (n)		
	10	20	30
Normal	1.348	1.088	0.926
Double-Exponential	1.521	1.411	1.273
Exponential	1.637	1.529	1.395
Log-Normal	1.737	1.650	1.560

Table 6: MSE Results for Simple Percentile Bootstrap Variance Estimate			
Distribution	Sample Size (n)		
	10	20	30
Normal	0.209	0.097	0.067
Double-Exponential	0.404	0.241	0.175
Exponential	0.616	0.363	0.255
Log-Normal	1.613	0.731	0.550

Results for Approach 1 – Adjusting by Expected Bias

The results for the change in bias when adjusting the variance estimator by the multiplier $\frac{n}{n-1}$, where n = sample size appear in Figure 6. As desired, it appears this adjustment does indeed reduce the magnitude of the bias in the simple percentile bootstrap variance estimator.

Figure 6: Bias Results for Approach 1 – Adjusting by Expected Bias



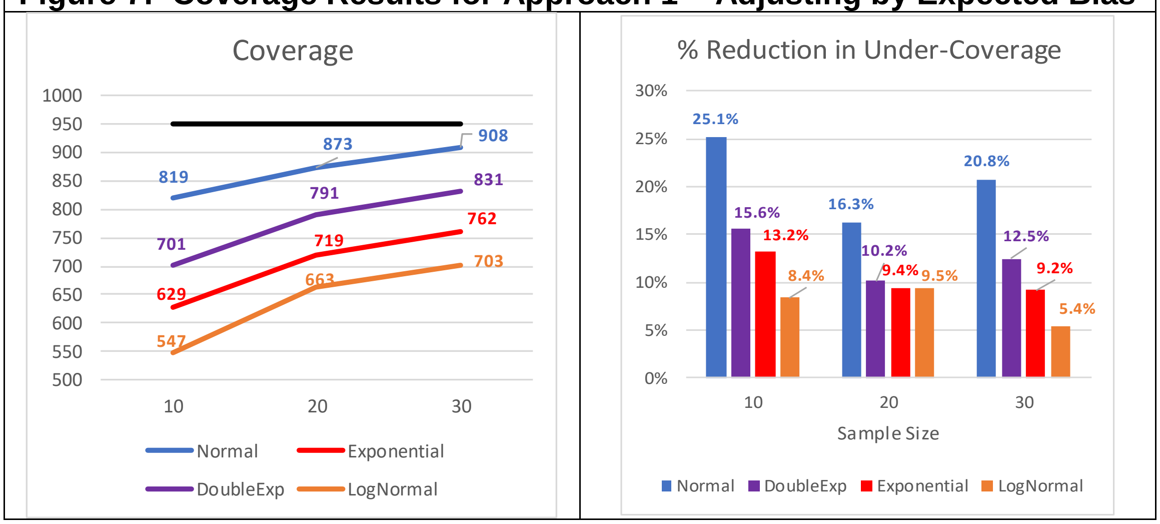
For the normal distribution, the adjustment moves the bias at all three sample sizes just into the positive range, but very near to zero. For the exponential distribution, the same results are observed; however, the bias is still low.

The double-exponential distribution is the clear outlier here as the adjustment does address the low bias of the simple percentile bootstrap estimator, but actually appears to over-correct as the sample size increases. The bias at a sample size of n = 30 is actually larger in magnitude (but positive) than the bias without this adjustment.

This approach does reduce the absolute magnitude of the bias in the simple percentile bootstrap variance estimate by a significant percentage, almost achieving zero bias for the normal distribution and sample size of 20. For the skewed distributions, the percentage reduction in absolute bias decreases as the sample size increases, but so does the actual magnitude of the bias. For the normal distribution, it appears the percent reduction also begins to become smaller as sample sizes increase beyond n = 20. However, it appears for the heavy-tailed, symmetric double-exponential distribution that the bias will actually become increasingly larger with larger sample sizes. This is not unexpected, as larger samples will be more likely to include more extreme values and naturally inflate a sample variance estimator such as S^2.

Figure 7 displays the Coverage results for this simple adjustment for the expected bias. The most obvious result is that this option does not generate coverage equal to the nominal 95%, as the highest realized coverage is 90.8% for the normal distribution and n = 30. As seen with the simple percentile bootstrap approach, coverages are uniformly higher for the symmetric distributions with normal coverages always larger than for the double-exponential. Coverages for the exponential case are uniformly higher than for the log-normal. Also, as expected, coverages uniformly increase with sample size.

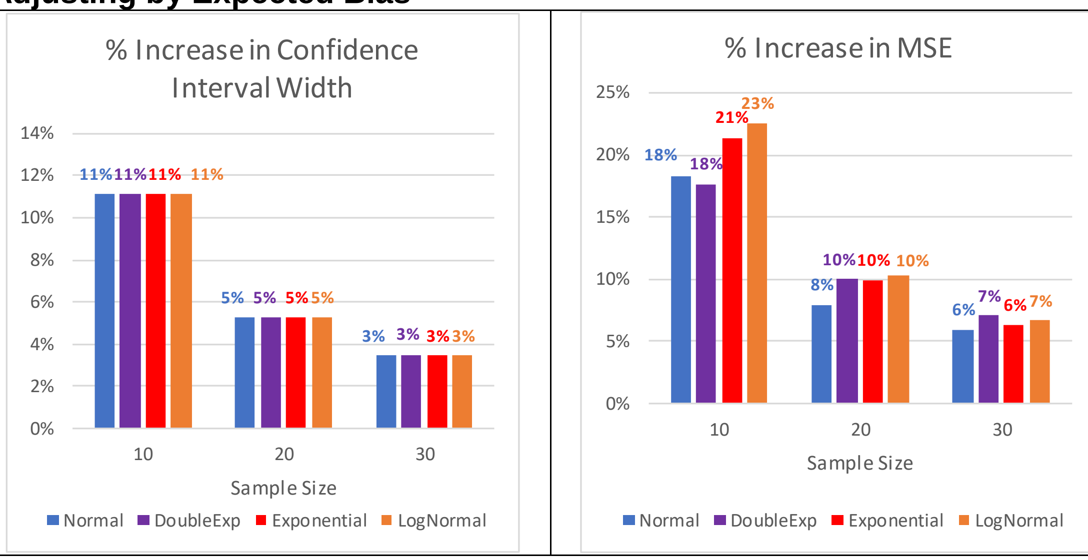
Figure 7: Coverage Results for Approach 1 – Adjusting by Expected Bias



The increases in coverage rates using this simple adjustment for the expected bias in the simple percentile bootstrap point estimate of the variance produces larger percentage reductions in the under-coverages observed for the associated simple percentile bootstrap 95% confidence intervals when the underlying data is symmetrically distributed. These reductions in under-coverage range from over 25% for the small sample (n=10) normally distributed data case to a low of only a little over a 5% reduction for the large sample (n=30) log-normal case.

Since this approach merely applies the bias adjustment multiple of $\frac{n}{n-1}$ to each of the bootstrap re-sample variance estimates, the associated percentile bootstrap confidence intervals will be $100\frac{1}{n-1}\%$ wider than the usual bootstrap intervals. This can be seen in the left-side chart of Figure 8.

Figure 8: Confidence Interval Width and MSE Results for Approach 1 – Adjusting by Expected Bias



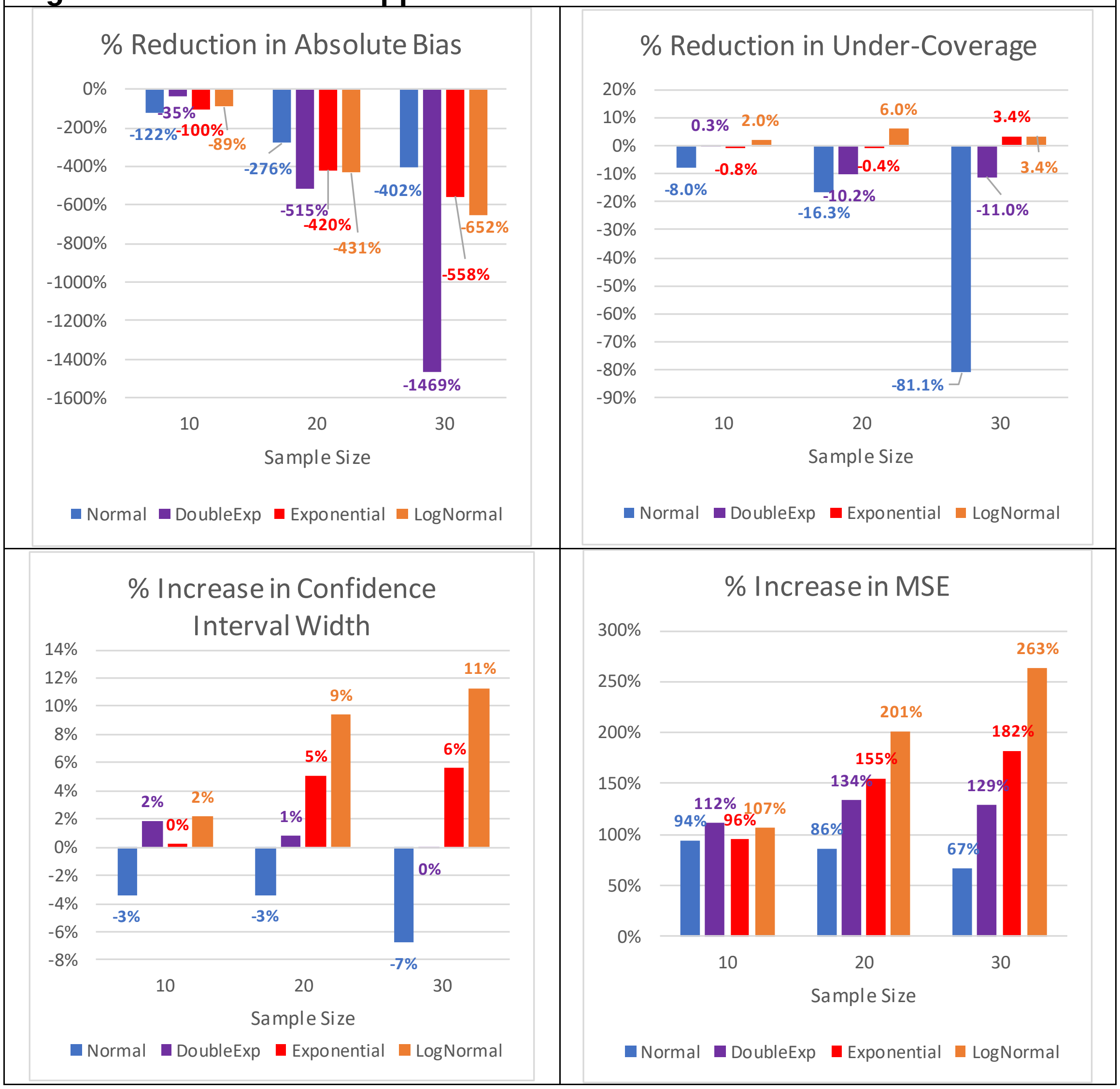
The right side of this figure displays the percentage increases in MSE observed for this approach. As with the confidence interval width, these increases become smaller as sample size increases; however, they are not constant across different distributions. For the small sample size (n=10), the increases in MSE are larger for the skewed distributions, but for the other sample sizes, the percentage increases tend to become more similar in magnitude.

Results for Approach 2 – Bias Corrected and Accelerated Method (BCa)

The Bias and Accelerated (BCa) method of Efron and Tibshriani (1993) might be expected to reduce the observed low bias for the simple percentile bootstrap variance estimate. Given that the distribution of a sample variance is known to be a skewed distribution, and that the BCa method includes an attempt to correct for skewness makes this approach seem potentially even more promising.

Unfortunately, as the negative percent reduction results in Figure 9 indicate, this approach actually over-corrects for the low bias, and produces biases to the high side of even larger magnitudes than those observed for the simple percentile bootstrap variance estimate. In addition, for many of the distribution, sample size situations, the BCa interval coverage rates are even lower than those observed using the simple percentile bootstrap approach. For those situations where they are not, the coverage rates are only minimally improved.

Figure 9: Evaluation of Approach 2 – Bias Corrected & Accelerated Method



Another observation in Figure 9 is that all of the evaluation metrics suggest that performance for this approach deteriorates as sample size increases. Over-correction for the low bias gets more pronounced (especially for the heavy-tailed double-exponential distribution); and under-coverage is not reduced (and seems to increase significantly for the normal distribution case, bootstrap is 5.3% low while BCa is 9.6% low). In addition, the width of BCa intervals and the MSE of the BCa estimators do not decrease at the same rate as the usual bootstrap intervals with increasing sample size. Width of these intervals for the lognormal case seem to minimal for the sample size range evaluated, and hence, in comparison to the usual bootstrap intervals are increasingly wider.

Results for Approach 3 – Shocking Bootstrap Observations

Since there are 40 different shock sizes-and-types combinations (Table 2) being applied to each of the 12 simulation settings (four underlying distributions by three different sample sizes), it was necessary to create a criterion to evaluate across all 480 different results to determine a shock size and type that would perform reasonably well across all 12 simulation settings.

Criteria for Identifying a Widely Useful Set of Shock Parameters

Since four different evaluation criteria are being considered (Bias, Coverage, Interval Width, and MSE), and it is unlikely that a single shock size and type will universally outperform all other size and type scenarios, a weighted average of the evaluation metrics was used to generate a single performance metric for the 40 respective shock scenarios. Since the MSE and the Confidence

Interval Width metrics generally carry similar information, the weight for the MSE evaluation metric was set to zero for this weighted average.

The other three evaluation metrics – Bias, Coverage, and Interval Width – were given weights of 0.3, 0.5, and 0.2, respectively. Coverage received the highest weight since generally the most value in a bootstrap approach is the generation of a confidence interval for the parameter of interest. Bias was given the next largest weight given the generally wide use of point estimators, and that the primary motivating factor for this work was the low bias of the simple percentile bootstrap variance estimator. The remaining weight was allocated to Interval Width recognizing that very wide intervals can ensure very high coverage rates.

In order to avoid different scales for the evaluation metrics to dominate the weighted average, z-scores of each metric evaluated across the 40 shocking scenarios were obtained. These z-scores were then what was utilized to create the weighted average performance metric for each specific scenario.

For each distribution, sample size combination, the z-score for Bias was the absolute value of the difference between the point estimate (mean of each bootstrap trial result) and 1 (the true population variance). Consequently, any deviation in the estimator from its associated parameter would produce a larger z-score for Bias. These values were obtained for all 40 shock scenarios, then the z-scores were obtained in the usual manner (subtracting the average of the 40 values and dividing by the standard deviation of the 40 values). Low (i.e., values well below zero) z-scores are desirable.

The z-scores for the Coverage metric used the difference between 950 and the relevant observed coverage of the result. Small values here represent coverages near the nominal 95%; large values would be much lower coverage rates; and results less than zero indicate coverages larger than the nominal 95%. Again, the z-scores were obtained by subtracting the average of these differences across the 40 shock scenarios and dividing by the standard deviation of the 40 differences. Again, low z-scores are desirable.

The z-scores for the Confidence Interval Widths directly utilized the widths, again, subtracting the average width across the 40 shock scenarios and dividing by the standard deviation of the 40 widths. Again, low z-scores are desirable as they indicate narrower intervals for a specific shock scenario.

For example, consider shock size and type Scenario 1 when the simulation is obtaining sample sizes of n = 10 from a normal distribution. For this scenario, all bootstrap replications were shocked with the Uniform shock type, and a shock size utilizing 1/4 of the Original Sample standard deviation to set the bounds of this uniform shock distribution. The observed Bias, Coverage, and Interval Width for this scenario appear in Table 7.

Table 7: Bias, Coverage, and Interval Width Results for Shock Type and Size Scenario 1 of the 40 Considered					
Shock Scenario	Extent of Shocks	Shock Type and Size	Bias	Coverage	Interval Width
1	All	Uniform 1/4*S.$_{Orig}$	0.9261	789	1.398

The observed Bias produces an absolute difference of 1 – 0.961 = 0.0739. The average of all 40 point such differences (i.e., those for all 40 shock scenarios) was 0.2365 and the corresponding variance was 0.1129, generating a z-score for Bias for this specific scenario of -0.4838.

The difference between the nominal and the observed Coverage for this scenario is 950 – 789 = 161. The average of all 40 such results was 93.925 and the variance was 2535.148, producing a z-score for Coverage for this scenario of 1.3322. Similarly, with the average of all 40 widths being 1.9617 and the variance being 0.5132, the Interval Width z-score for this scenario was -0.7873.

Using three z-scores so obtained, a weighted average overall z-score with weights 0.3(Bias) + 0.5(Coverage) + 0.2(Interval Width) was generated for each of the 40 shock scenarios. This was done for each of the 12 distributions, sample size combinations considered in the simulation effort.

Different weights could be used here; however, more weight was put toward the Bias and Coverage than the Confidence Interval Width since the objective is to correct for both low bias and coverage. For this specific example, the overall weighted average z-score was 0.3635.

Identifying a Widely Useful Set of Shock Parameters

Weighted average z-scores were obtained for all 12 distributions, sample size combinations evaluated in the simulation. The “best” shocking scheme should not only work well, have small weighted overall z-score, under the standard normal case where the data is well-behaved, but also for distributions that are non-normal, where the bootstrap method is generally of most use. It also should perform well especially when the sample sizes are small, as this is another critical situation where the bootstrap method is more commonly utilized for analysis.

Table 8 displays the weighted average z-scores for all 40 shock scenarios across all 12 distributions, sample size combinations. The cells in the table are color coded with green representing the low values that indicate generally more desirable performance for that shock scenario. Analogously, red cells hold high values representing shock scenarios that performed less desirably.

Perhaps the most apparent observation in Table 8 is that shocking all bootstrap sample results with large shocks (i.e., using shocks that are normally distributed with a standard deviation 5/4 the magnitude of the original sample standard deviation) has the least desirable performance universally across all the distribution, sample size combinations.

Table 8: Weighted Average Z-Scores for Shock Scenarios

Scenario	Shock Extent	Type & Size	Standard Normal			Exponential			Double Exponential			Log Normal			Average Z-Score	Rank
			10	20	30	10	20	30	10	20	30	10	20	30		
1	All	Uniform 1/4*S.Orig.	0.3635	-0.4683	-0.573	0.394957	0.062384	-0.25723	0.31867	-0.15253	-0.38665	0.322291	0.234795	-0.03356	-0.01458	30
2		Uniform 2/4*S.Orig.	-0.0812	-0.5005	-0.566	0.008535	-0.13332	-0.38165	0.039949	-0.25173	-0.40614	0.069231	-0.0565	-0.24127	-0.20843	19
3		Uniform 3/4*S.Orig.	-0.4466	-0.4546	-0.428	-0.19147	-0.31218	-0.36932	-0.26441	-0.3904	-0.3303	-0.0765	-0.05754	-0.29136	-0.30109	12
4		Uniform 4/4*S.Orig.	-0.5704	-0.2192	-0.117	-0.13605	-0.20601	-0.11462	-0.32192	-0.25766	-0.07185	-0.10769	0.08784	0.058527	-0.16467	22
5		Uniform 5/4*S.Orig.	-0.4332	0.26979	0.437	-0.07126	0.337097	0.441094	-0.16917	0.467975	0.635708	-0.04432	0.235335	0.510214	0.218022	36
6		Uniform 1/4*IQR.Orig.	0.335	-0.4565	-0.58	0.384781	0.047068	-0.22897	0.372576	-0.11928	-0.39176	0.321173	0.244481	-0.06477	-0.01138	31
7		Uniform 2/4*IQR.Orig.	-0.0664	-0.5066	-0.557	0.089854	-0.20359	-0.40192	0.173622	-0.22364	-0.43648	0.148017	-0.01583	-0.25588	-0.18797	20
8		Uniform 3/4*IQR.Orig.	-0.4572	-0.4539	-0.436	-0.25649	-0.41277	-0.46596	-0.0864	-0.44473	-0.41538	-0.14032	-0.25323	-0.44346	-0.35549	6
9		Uniform 4/4*IQR.Orig.	-0.5567	-0.2102	-0.121	-0.40955	-0.49601	-0.43267	-0.33111	-0.49956	-0.42238	-0.34676	-0.56856	-0.61563	-0.41752	2
10		Uniform 5/4*IQR.Orig.	-0.3348	0.26353	0.3931	-0.48337	-0.5573	-0.25573	-0.44919	-0.44846	-0.2097	-0.51858	-0.66407	-0.59039	-0.32125	9
11		Normal 1/4*S.Orig.	0.0346	-0.4902	-0.57	0.129542	-0.09134	-0.35792	0.148627	-0.21227	-0.39382	0.156252	0.000332	-0.15885	-0.15046	24
12		Normal 2/4*S.Orig.	-0.5771	-0.3793	-0.31	-0.26941	-0.35033	-0.33288	-0.29711	-0.42272	-0.26224	-0.1481	-0.13583	-0.18968	-0.30624	11
13		Normal 3/4*S.Orig.	-0.3732	0.3736	0.5379	-0.24048	0.259784	0.521712	-0.26981	0.520178	0.721233	-0.14898	0.124407	0.439025	0.205447	35
14		Normal 4/4*S.Orig.	0.6206	1.94235	2.0477	0.493194	1.780065	2.344129	0.413693	2.237012	2.566123	0.393367	1.197148	2.095146	1.510878	39
15		Normal 5/4*S.Orig.	2.5597	3.73541	3.5498	1.757346	3.971528	4.265072	1.999802	4.47119	4.454212	1.634223	3.162701	4.209544	3.314211	40
16		Normal 1/4*IQR.Orig.	0.0587	-0.5061	-0.572	0.199437	-0.14958	-0.37936	0.211121	-0.22622	-0.42234	0.190753	0.085641	-0.17001	-0.13997	25
17		Normal 2/4*IQR.Orig.	-0.5318	-0.3737	-0.322	-0.32505	-0.47889	-0.45526	-0.2922	-0.52994	-0.46597	-0.24365	-0.41701	-0.51042	-0.41214	3
18		Normal 3/4*IQR.Orig.	-0.3908	0.34715	0.4747	-0.51286	-0.60201	-0.22296	-0.56102	-0.44177	-0.16408	-0.56619	-0.72487	-0.67807	-0.3369	8
19		Normal 4/4*IQR.Orig.	0.5236	1.63359	1.7686	-0.23855	0.174879	0.819567	-0.49637	0.383892	0.644451	-0.45985	-0.56863	-0.26736	0.326488	37
20		Normal 5/4*IQR.Orig.	2.064	3.17591	3.1344	0.380745	1.643406	2.240461	-0.04602	1.496967	1.848291	-0.1511	0.296465	1.119749	1.433598	38
21	Repeated	Uniform 1/4*S.Orig.	0.4953	-0.4239	-0.563	0.476463	0.10878	-0.18532	0.437356	-0.1028	-0.36137	0.354279	0.32987	0.009369	0.047925	33
22		Uniform 2/4*S.Orig.	0.3008	-0.4799	-0.591	0.275611	-0.01008	-0.27644	0.347335	-0.15977	-0.37651	0.282069	0.175803	-0.08204	-0.04955	26
23		Uniform 3/4*S.Orig.	-0.0081	-0.492	-0.567	0.042418	-0.09598	-0.3888	0.115109	-0.23854	-0.40528	0.113301	-0.05641	-0.15486	-0.17797	21
24		Uniform 4/4*S.Orig.	-0.2715	-0.4923	-0.512	-0.11368	-0.1882	-0.42461	-0.14185	-0.3606	-0.38673	-0.02764	-0.12427	-0.29575	-0.27829	14
25		Uniform 5/4*S.Orig.	-0.4465	-0.4072	-0.393	-0.21028	-0.29058	-0.34672	-0.29097	-0.38992	-0.32234	-0.10293	-0.13118	-0.24394	-0.29795	13
26		Uniform 1/4*IQR.Orig.	0.4759	-0.4278	-0.576	0.462949	0.10899	-0.18466	0.454787	-0.10848	-0.34708	0.368515	0.345954	-0.01635	0.046425	32
27		Uniform 2/4*IQR.Orig.	0.3026	-0.4927	-0.586	0.317119	0.006827	-0.26781	0.385901	-0.16102	-0.4015	0.279864	0.210336	-0.08463	-0.04092	27
28		Uniform 3/4*IQR.Orig.	-0.0051	-0.5235	-0.568	0.146415	-0.18766	-0.39799	0.248938	-0.2395	-0.4176	0.163931	0.027357	-0.17191	-0.16042	23
29		Uniform 4/4*IQR.Orig.	-0.2637	-0.4835	-0.5	-0.13004	-0.30869	-0.46268	0.068456	-0.30306	-0.45194	0.018406	-0.18003	-0.32903	-0.27716	15
30		Uniform 5/4*IQR.Orig.	-0.3769	-0.4198	-0.405	-0.25155	-0.45419	-0.46217	-0.10795	-0.40513	-0.44332	-0.17387	-0.36131	0.156243	-0.30876	10
31		Normal 1/4*S.Orig.	0.3726	-0.4766	-0.59	0.338466	0.040102	-0.25188	0.399601	-0.13508	-0.37049	0.291579	0.256649	-0.06709	-0.01605	28
32		Normal 2/4*S.Orig.	-0.1594	-0.4915	-0.546	-0.05348	-0.18049	-0.39268	0.013661	-0.26851	-0.40114	0.036136	-0.15158	-0.23071	-0.23544	16
33		Normal 3/4*S.Orig.	-0.5357	-0.41	-0.366	-0.26392	-0.32927	-0.33882	-0.37431	-0.41027	-0.33352	-0.20385	-0.19427	-0.3186	-0.33989	7
34		Normal 4/4*S.Orig.	-0.4675	-0.082	0.0182	-0.36222	-0.31953	-0.18702	-0.41932	-0.2284	0.029867	-0.34151	-0.29109	-0.10825	-0.2299	17
35		Normal 5/4*S.Orig.	-0.1747	0.45594	0.5844	-0.36246	-0.05245	0.427734	-0.38589	0.420192	0.73972	-0.34847	-0.086	0.208742	0.118895	34
36		Normal 1/4*IQR.Orig.	0.3338	-0.4678	-0.597	0.371457	0.016487	-0.22414	0.379819	-0.12467	-0.36675	0.300831	0.237901	-0.04642	-0.01553	29
37		Normal 2/4*IQR.Orig.	-0.145	-0.5211	-0.54	0.013043	-0.25482	-0.43625	0.158308	-0.31355	-0.44088	0.104272	-0.09881	-0.24571	-0.2267	18
38		Normal 3/4*IQR.Orig.	-0.5024	-0.4064	-0.374	-0.3197	-0.50391	-0.45285	-0.22982	-0.46332	-0.45222	-0.22395	-0.42745	-0.47657	-0.40271	4
39		Normal 4/4*IQR.Orig.	-0.4333	-0.0961	-0.019	-0.49967	-0.65664	-0.46235	-0.48775	-0.57284	-0.41608	-0.54446	-0.80046	-0.67911	-0.47235	1
40		Normal 5/4*IQR.Orig.	-0.2315	0.4159	0.5016	-0.58079	-0.7316	-0.26009	-0.66474	-0.39106	-0.16577	-0.62976	-0.88809	-0.74486	-0.36423	5

However, in looking for the desirable green cells, it appears that either

- shocking all bootstrap samples with uniformly distributed disturbances with a range set a larger multiple (75% - 125%) of the original sample IQR sample standard deviation estimator (16), or
- shocking only the repeated samples with normally distributed disturbances with a standard deviation again determined from a larger multiple of the original sample IQR sample standard deviation estimator.

Note that the last column ranks the average z-score across all 12 distributions, sample size combinations in ascending order from 1 (most desirable performance) to 40 (least desirable performance). These ranks are also color coded from green (low ranks) to red (high ranks), and since the ranks are uniformly distributed, there is a more uniform distribution of the colors from green to red across the shock scenarios than was obtained simply using the average z-scores. When using this column (next to last in Table 8), the three largest average z-scores command the red and orange cells, as their values are 3.3142 (All-Normal 5/4*S_{Orig}), 1.5119 (All-Normal 4/4*S_{Orig}), and 1.4336 (All-Normal 5/4*IQR_{Orig}). The next largest average z-score is 0.3265 (All-Normal 4/4*IQR_{Orig}) which is more than a full standard deviation lower.

Closer inspection of Table 8 indicates that 32 of the 40 scenarios average z-scores are less than zero, and the smallest z-score is -0.4724 (Repeated-

Normal 4/4*IQR_{Orig}). Hence, this distribution of averages of weighted average z-scores is still highly skewed to the right.

The ten most desirable shock scenarios are displayed in Table 9. A clear observation is that basing the spread of the disturbance distribution on a generally larger multiple of the original sample IQR estimator of the standard deviation (16) is preferred over using the common sample variance estimator, S.

Table 9: Top Ten Performing Shock Scenarios

Rank	Average Z-Score	Shock Extent	Disturbance Distribution	Multiplier	Standard Deviation Estimator
1	-0.4724	Repeated	Normal	4/4 = 100%	IQR_{Orig}
2	-0.4175	All	Uniform	4/4 = 100%	IQR_{Orig}
3	-0.4121	All	Normal	2/4 = 50%	IQR_{Orig}
4	-0.4027	Repeated	Normal	3/4 = 75%	IQR_{Orig}
5	-0.3642	Repeated	Normal	5/4 = 125%	IQR_{Orig}
6	-0.3555	All	Uniform	3/4 = 75%	IQR_{Orig}
7	-0.3399	Repeated	Normal	3/4 = 75%	S_{Orig}
8	-0.3369	All	Normal	3/4 = 75%	IQR_{Orig}
9	-0.3213	All	Uniform	5/4 = 125%	IQR_{Orig}
10	-0.3088	Repeated	Uniform	5/4 = 125%	IQR_{Orig}

Three of the top five performing scenarios involve shocking only repeated bootstrap re-sample observations using normally distributed disturbances with a mean of zero and a standard deviation estimated using a larger multiple (75% - 125%) of the original sample IQR divided by 1.35. While the “best” performing of these indicates using 100% of this estimator as the standard deviation of the shock distribution, it appears that using only a 75% multiple generates better performance for the normal distribution across all sample sizes than for either of the larger multipliers. Consequently, with a desire to not sacrifice too much performance when the data truly are more nearly normally distributed, a shock scenario

- utilizing a normally distributed disturbance distribution with
- a standard deviation of ¾ the size of the original sample IQR divided by 1.35
- applied only to repeated bootstrap sample observations

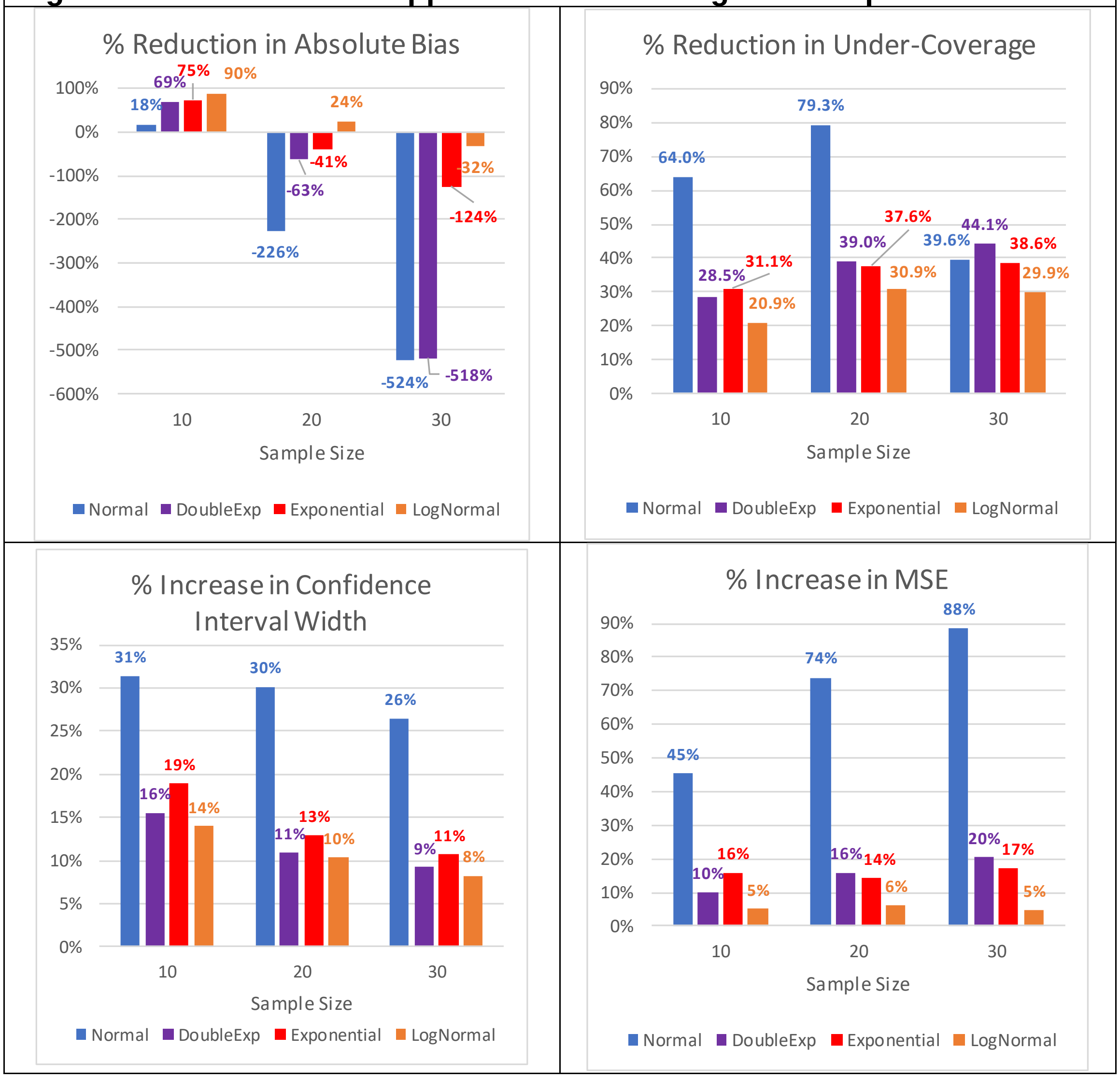
was selected as the most appropriate set of shock parameters for this approach.

Evaluation of Performance for Approach 3 – Shocking Bootstrap Observations

The actual performance metric results for the selected shock scenario for this approach relative to the simple percentile bootstrap variance estimator appear in Figure 10. It is clear that this approach sacrifices bias in the point estimator to improve coverage of the corresponding interval estimate.

While the first approach (merely adjusting for the expected bias) does a relatively good job of reducing the amount of low bias, this approach (shocking bootstrap observations) actually produces larger biases for sample sizes greater than n = 10. These increases are relatively large for the symmetric distributions for a sample size n = 30. The increase in bias for the normal distribution with a sample size of n = 30 is even larger than the associated increase in bias for the BCa method.

Figure 10: Evaluation of Approach 3 – Shocking Bootstrap Observations



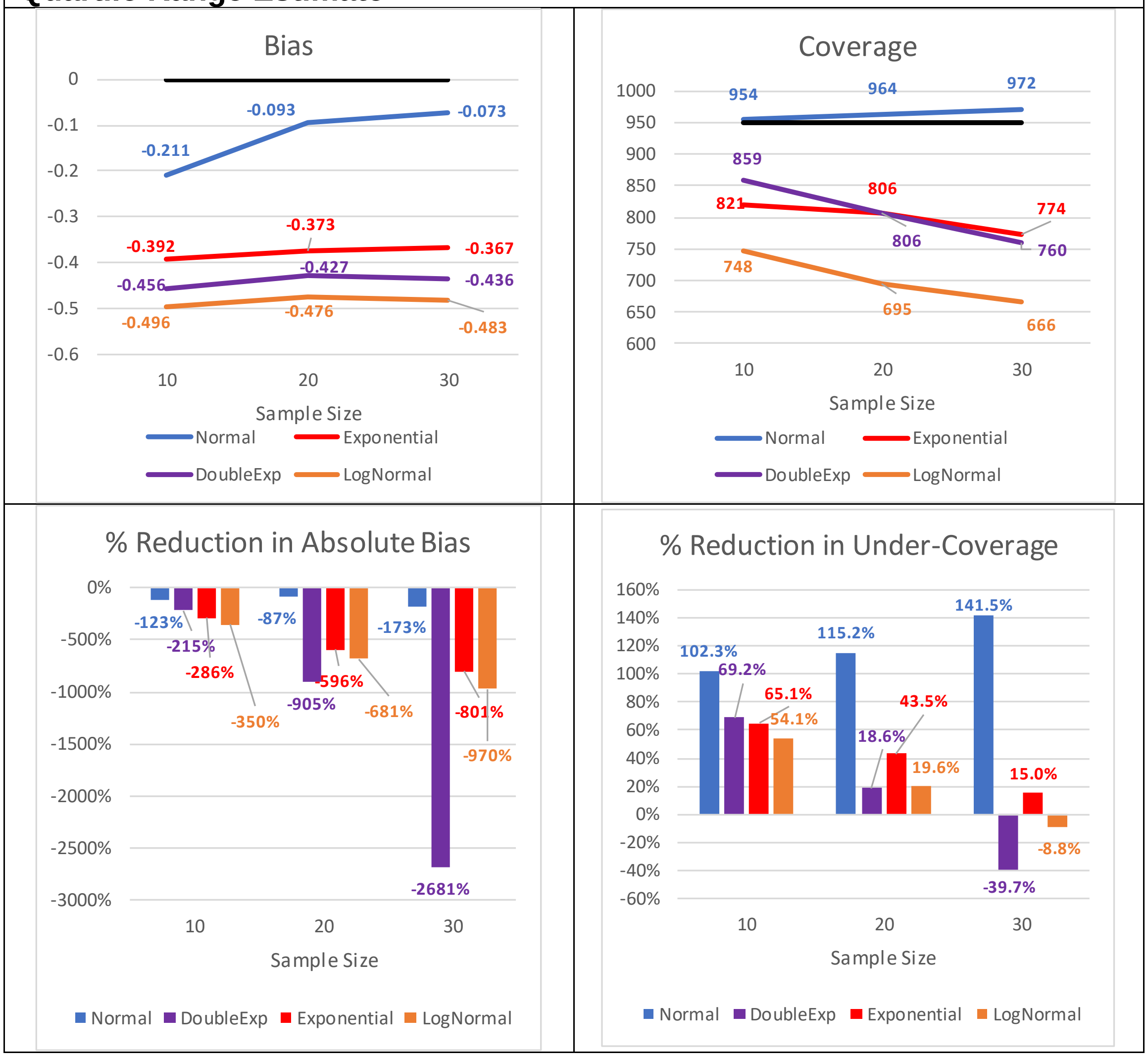
However, this sacrifice is made in order to generate higher coverage values. For the first approach (adjusting by expected bias), the largest observed reduction in under-coverage was ~25%, yet for this shocking approach, virtually all of the reductions in under-coverage are larger than 25% (the only exception is for the log-normal distribution with a sample size of n = 10). None of the coverage rates achieve the desired nominal 95%; however, for a sample size of n = 20, in the normal distribution case, the coverage observed was 93.1%, similar to the coverage rate for a 95% bootstrap percentile interval for the population mean. Of course, these improvements in coverage will necessarily require wider confidence intervals and generate larger MSE values.

Results for Approach 4 – Using Inter-Quartile Range Estimate

This approach has the benefit of simplicity, as it merely requires a bootstrap of the IQR, then estimation of the population variance using $\hat{\sigma}^2 = \left(\frac{IQR}{1.35}\right)^2$(16) for each bootstrap re-sample. As Figure 11 displays, the bootstrap point estimator is not very good and, unfortunately, is biased low to an even larger degree than when bootstrapping the sample variance S^2. However, the coverage of the associated bootstrap percentile confidence interval for this approach is generally higher than that achieved for the other approaches at smaller sample sizes.

For the normal distribution case, the simple percentile bootstrap confidence interval, using the modified IQR statistic (16) to estimate the variance, actually achieves the stated nominal 95%. Moreover, as the sample size increases in this case, the coverage actually also increases, continues to achieve the nominal rate, and actually might be considered conservative.

Figure 11: Bias and Coverage Results for Approach 4 – Using the Inter-Quartile Range Estimate



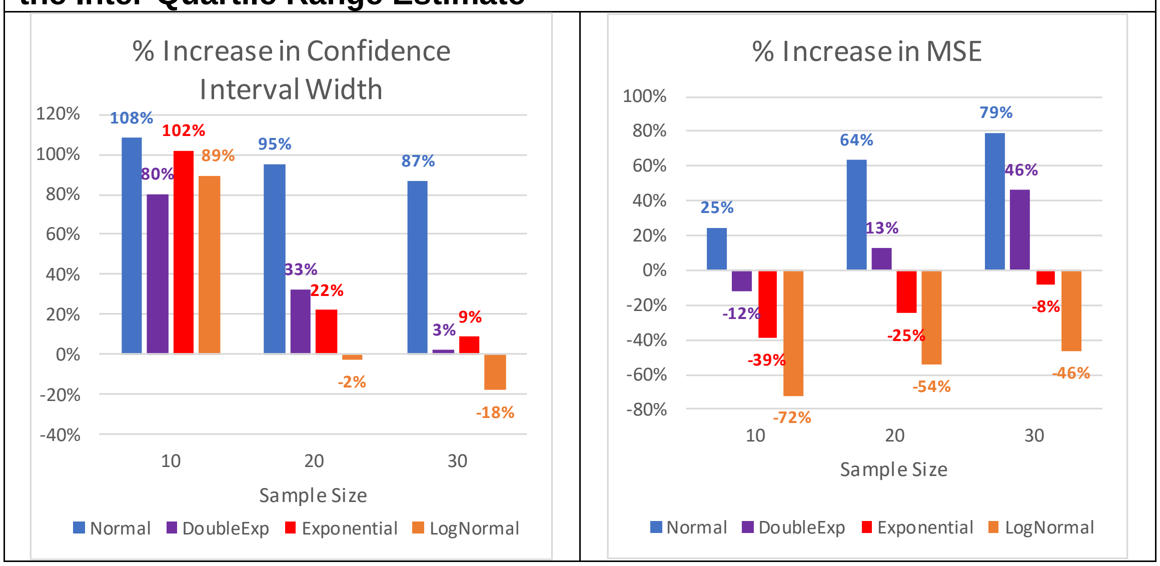
For the non-normal distributions, while still falling below the nominal 95% rate, the coverage is higher than all the other approaches for sample size n = 10. As sample sizes increase, however, coverage also deteriorates. At sample size n = 30, both the double-exponential and log-normal distributions have coverages that are even lower than those achieved by the simple percentile bootstrap interval based on the sample variance. This is due to all of the confidence intervals missing low (upper limit being less than the true variance of 1).

The difference in the behavior between the normal distribution and the non-normal distributions is understandable. The variance estimator being used here divides the respective sample IQR by the value that is appropriate if the data is indeed normally distributed (i.e., 1.35). As noted previously, the proper divisors are all smaller than this for the other distributions considered here (1.1 – Exponential, 0.98 – Double-Exponential, and 0.97 – Log-Normal). This does not seem to be an issue for smaller sample sizes, but begins to be an issue as sample sizes increase.

Of course, if the underlying distribution is known, then the proper adjustment divisor for the IQR could be applied. However, this pre-supposes knowledge that generally would not be available to an analyst. Moreover, if the under-lying distribution giving rise to the data was known, it is likely the analyst would favor some estimation approach other than a bootstrap. Although not considered here, a divisor different than 1.35 could be utilized (e.g., 1) with this approach.

Interestingly, for this approach, where improved coverages are achieved (normal – all sample sizes and non-normal – sample size n = 10), the interval widths are indeed wider than all the other approaches (see Figure 12). However, the increases in the MSE of the estimator are generally less than the other approaches. For the skewed distributions, the MSE is less than that for the simple percentile bootstrap sample variance estimator at all sample sizes considered, and this is also true for the double-exponential distribution at sample size n = 10.

Figure 12: Interval Width and MSE Performance for Approach 4 – Using the Inter-Quartile Range Estimate



Results When Combining Approaches 1 & 3

Considering the evaluation for both Approach 1 – Adjusting by Expected Bias -- and Approach 3 – Shocking Bootstrap Observations, the only reason to combine them would be to potentially increase coverage. Approach 1 effectively addresses the low bias issue alone; however, shocking the bootstrap observations will increase variation in the data set and force the bias towards the high side.

Another potential advantage of combining these approaches is that the bias adjustment (Approach 1) might allow for smaller shocks to be applied for Approach 3. In evaluating the 40 shock scenarios after making the Approach 1 bias adjustment in the same fashion as described previously for Approach 3 alone, it was found that more of the better performing shock scenarios did involve marginally lower weights (see Table 10). However, the results were not appreciably different than when applying shocks to the data not subject to the bias adjustment of Approach 1. Consequently, again, the selected shock scenario for the combination of approaches

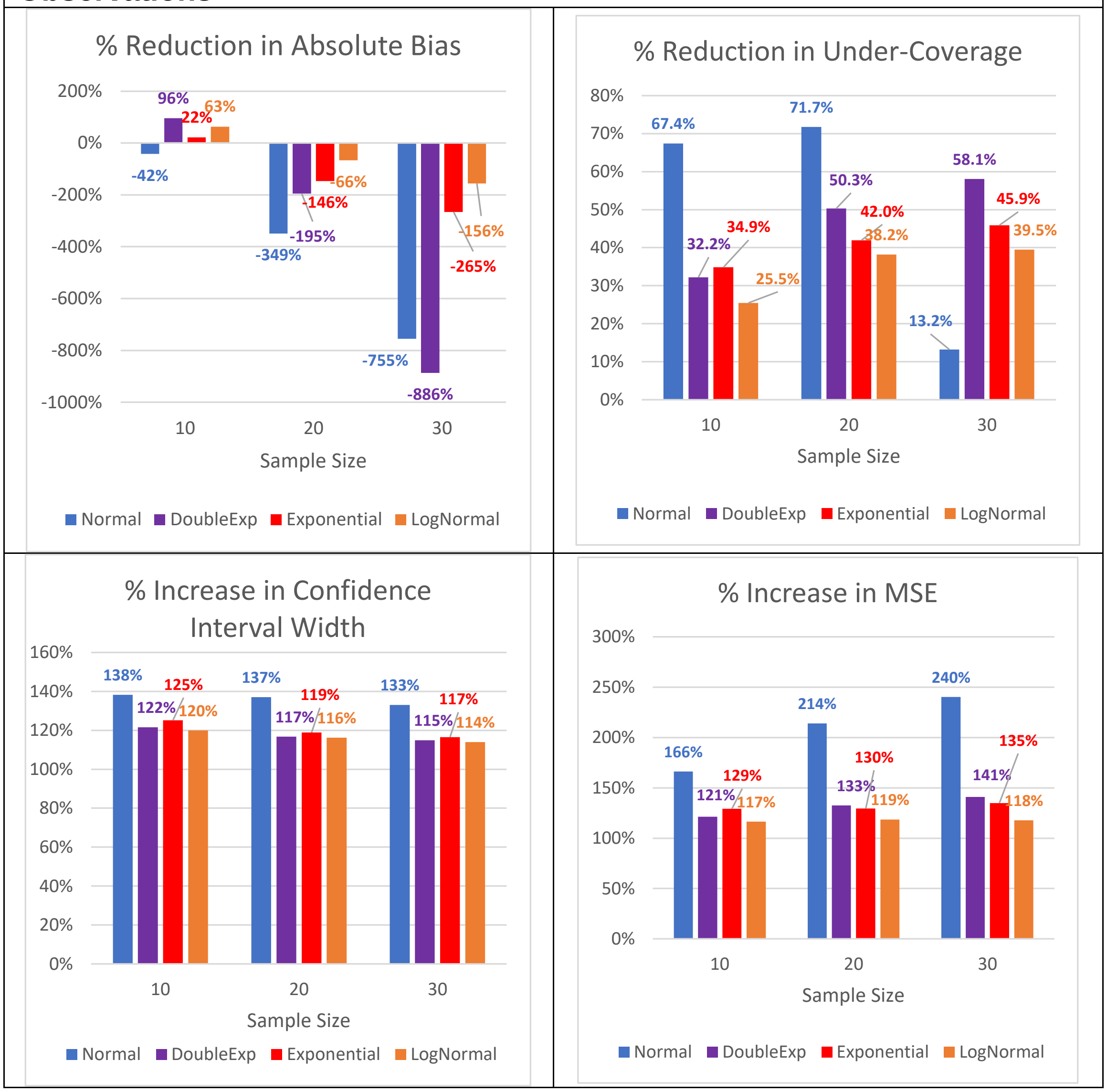
- utilized a normally distributed disturbance distribution with
- a standard deviation of 3/4 the size of the original sample IQR divided by 1.35, and was
- applied only to repeated bootstrap sample observations.

Table 10: Weighted Average Z-Scores for Shock Scenarios After Making Adjustment for Expected Bias

Scenario	Shock Extent	Type and Size	Standard Normal			Double Exponential			Exponential			Log Normal			Average Z-Score	Rank
			10	20	30	10	20	30	10	20	30	10	20	30		
1	All	Uniform 1/4*S.Orig.	0.20328	-0.5599	-0.6514	0.28373	-0.2733	-0.4506	0.33328	-0.0848	-0.3875	0.24758	0.12017	-0.1725	-0.115989	29
2		Uniform 2/4*S.Orig.	-0.1794	-0.5292	-0.59	0.0389	-0.3693	-0.4369	0.07041	-0.1675	-0.4092	0.09477	-0.0287	-0.3171	-0.235269	19
3		Uniform 3/4*S.Orig.	-0.4228	-0.434	-0.4003	-0.2469	-0.349	-0.3285	-0.1525	-0.2983	-0.3548	-0.0583	-0.0459	-0.2163	-0.275632	12
4		Uniform 4/4*S.Orig.	-0.5594	-0.1768	-0.0428	-0.287	-0.0971	0.01671	-0.1403	-0.1669	-0.0648	-0.0963	0.03011	0.10757	-0.123085	27
5		Uniform 5/4*S.Orig.	-0.3451	0.37436	0.50494	-0.1592	0.63102	0.75951	-0.0855	0.50117	0.56187	0.01189	0.39079	0.57923	0.3104193	36
6		Uniform 1/4*IQR.Orig.	0.19336	-0.5617	-0.6525	0.30629	-0.2426	-0.4647	0.34598	-0.0975	-0.4093	0.2856	0.12122	-0.1779	-0.112811	31
7		Uniform 2/4*IQR.Orig.	-0.1597	-0.5353	-0.5983	0.1449	-0.3571	-0.4627	-0.0235	-0.2586	-0.4457	0.10939	-0.0532	-0.3479	-0.248979	18
8		Uniform 3/4*IQR.Orig.	-0.4633	-0.4313	-0.4153	-0.1302	-0.4911	-0.4529	-0.1822	-0.4119	-0.4304	-0.1266	-0.2616	-0.4373	-0.35284	6
9		Uniform 4/4*IQR.Orig.	-0.5465	-0.1511	-0.0938	-0.2656	-0.4885	-0.3943	-0.4122	-0.4766	-0.4161	-0.3489	-0.6025	-0.5431	-0.394937	4
10		Uniform 5/4*IQR.Orig.	-0.2765	0.36801	0.42323	-0.3868	-0.3022	-0.1394	-0.4559	-0.4804	-0.1259	-0.4456	-0.6917	-0.5219	-0.252903	16
11		Normal 1/4*S.Orig.	-0.0868	-0.536	-0.6245	0.11994	-0.3444	-0.4492	0.12541	-0.1916	-0.4054	0.08872	0.0067	-0.2583	-0.212944	23
12		Normal 2/4*S.Orig.	-0.6394	-0.344	-0.2803	-0.3185	-0.3288	-0.2177	-0.2364	-0.3221	-0.2746	-0.1209	-0.034	-0.0796	-0.266352	15
13		Normal 3/4*S.Orig.	-0.335	0.44895	0.60379	-0.2581	0.7071	0.88897	-0.2046	0.44875	0.65789	-0.0968	0.27208	0.57134	0.3086934	35
14		Normal 4/4*S.Orig.	0.87217	1.99292	2.09113	0.53376	2.3433	2.59103	0.5324	1.97346	2.40562	0.44633	1.4871	2.27233	1.6284621	39
15		Normal 5/4*S.Orig.	3.01212	3.65246	3.41653	2.18018	4.53085	4.34432	1.88422	4.1371	4.20755	1.68069	3.56175	4.39026	3.4165015	40
16		Normal 1/4*IQR.Orig.	-0.0636	-0.5473	-0.6114	0.15813	-0.3247	-0.4677	0.07614	-0.2324	-0.4468	0.14595	-0.0241	-0.2685	-0.217195	22
17		Normal 2/4*IQR.Orig.	-0.5596	-0.3443	-0.2921	-0.2374	-0.5035	-0.4559	-0.2914	-0.4944	-0.4435	-0.277	-0.4431	-0.48	-0.401848	2
18		Normal 3/4*IQR.Orig.	-0.3509	0.4273	0.54772	-0.5353	-0.2845	-0.0971	-0.544	-0.5193	-0.0908	-0.5451	-0.7619	-0.5629	-0.276398	10
19		Normal 4/4*IQR.Orig.	0.7016	1.65861	1.82209	-0.4752	0.54336	0.77297	-0.2218	0.46089	0.94907	-0.4451	-0.4299	0.03352	0.4475096	37
20		Normal 5/4*IQR.Orig.	2.54379	3.16186	3.0455	0.00037	1.62265	1.92643	0.42241	1.93268	2.38705	-0.1509	0.53521	1.44925	1.5730201	38
21	Repeated	Uniform 1/4*S.Orig.	0.36113	-0.5618	-0.6488	0.4212	-0.2296	-0.4672	0.40783	-0.0745	-0.388	0.33295	0.1929	-0.1289	-0.065217	33
22		Uniform 2/4*S.Orig.	0.16716	-0.5582	-0.6397	0.26346	-0.2874	-0.4592	0.26479	-0.1132	-0.3944	0.23341	0.00692	-0.1979	-0.142849	26
23		Uniform 3/4*S.Orig.	-0.1395	-0.5337	-0.606	0.07468	-0.3228	-0.4503	0.05218	-0.1876	-0.4082	0.08443	-0.0266	-0.2601	-0.226951	20
24		Uniform 4/4*S.Orig.	-0.3111	-0.4968	-0.5221	-0.113	-0.3957	-0.3973	-0.119	-0.2863	-0.3984	0.01464	-0.0692	-0.305	-0.283286	9
25		Uniform 5/4*S.Orig.	-0.3768	-0.3741	-0.36	-0.196	-0.3352	-0.2969	-0.1145	-0.3083	-0.3382	-0.0923	-0.0586	-0.1778	-0.252375	17
26		Uniform 1/4*IQR.Orig.	0.33934	-0.5526	-0.6514	0.4311	-0.2471	-0.4726	0.38343	-0.1025	-0.3825	0.34735	0.17161	-0.1314	-0.072278	32
27		Uniform 2/4*IQR.Orig.	0.13542	-0.5605	-0.6374	0.27754	-0.2822	-0.4682	0.26005	-0.156	-0.4054	0.24333	-0.0118	-0.1715	-0.148048	25
28		Uniform 3/4*IQR.Orig.	-0.1181	-0.549	-0.6008	0.19124	-0.3564	-0.4745	0.10057	-0.2553	-0.4343	0.13844	-0.0725	-0.2917	-0.226869	21
29		Uniform 4/4*IQR.Orig.	-0.3436	-0.4999	-0.5187	0.0141	-0.4105	-0.4761	-0.1263	-0.3786	-0.4531	-0.0097	-0.1681	-0.4067	-0.314769	7
30		Uniform 5/4*IQR.Orig.	-0.361	-0.3632	-0.377	-0.1839	-0.487	-0.4809	-0.2014	-0.4424	-0.4169	-0.1685	-0.3203	-0.4709	-0.35611	5
31		Normal 1/4*S.Orig.	0.20251	-0.5549	-0.638	0.28462	-0.267	-0.4515	0.31047	-0.0802	-0.41	0.26938	0.03079	-0.1558	-0.121634	28
32		Normal 2/4*S.Orig.	-0.2593	-0.5232	-0.5699	-0.0344	-0.3507	-0.4342	-0.0155	-0.2652	-0.4223	0.01621	-0.0747	-0.3193	-0.271037	13
33		Normal 3/4*S.Orig.	-0.5044	-0.3868	-0.347	-0.32	-0.3791	-0.3016	-0.243	-0.3268	-0.3464	-0.1924	-0.159	-0.1373	-0.30366	8
34		Normal 4/4*S.Orig.	-0.4091	-0.0005	0.08866	-0.3963	-0.0268	0.11278	-0.3394	-0.2592	-0.0499	-0.3296	-0.2906	-0.0477	-0.162301	24
35		Normal 5/4*S.Orig.	-0.0621	0.57875	0.63979	-0.3824	0.6335	0.88955	-0.3285	0.17073	0.60223	-0.2775	-0.0207	0.46403	0.2422791	34
36		Normal 1/4*IQR.Orig.	0.21442	-0.5561	-0.6485	0.35846	-0.2521	-0.4785	0.32184	-0.1248	-0.3852	0.27136	0.05792	-0.1651	-0.115524	30
37		Normal 2/4*IQR.Orig.	-0.1966	-0.5336	-0.5697	0.11052	-0.3664	-0.477	-0.0673	-0.337	-0.4706	0.05762	-0.1169	-0.3464	-0.276118	11
38		Normal 3/4*IQR.Orig.	-0.4617	-0.3782	-0.365	-0.2259	-0.4984	-0.4861	-0.276	-0.4801	-0.4398	-0.2186	-0.48	-0.5069	-0.401383	3
39		Normal 4/4*IQR.Orig.	-0.3498	-0.0231	0.04251	-0.399	-0.5152	-0.3513	-0.5165	-0.6701	-0.4357	-0.4953	-0.7986	-0.6309	-0.428574	1
40		Normal 5/4*IQR.Orig.	-0.0652	0.49378	0.72699	-0.6421	-0.2462	-0.0615	-0.5937	-0.5746	-0.0873	-0.6247	-0.9413	-0.6329	-0.270719	14

In evaluation of this combination of approaches, Figure 13 suggests that for sample size n = 10, the combined effect on the bias is perhaps tolerable, but it gets progressively worse as the sample size increases. The coverages are marginally higher for all the distributional models at sample size n = 10, and also for all the non-normal distributions at the other sample sizes considered.

Figure 13: Performance Results for the Combination of Approach 1 – Expected Bias Adjustment and Approach 3 – Shocking Bootstrap Observations



However, this improved coverage comes at the price of much wider intervals than previously observed for any of the other approaches. The associated increase in MSE is also largest versus all other approaches except for a few situations when the BCa method was applied.

Conclusion

In order to draw some general conclusions, it will be necessary to compare performance across the five approaches considered to address the low bias and under-coverage of the simple percentile bootstrap variance estimate. To review these five approaches can generally be described as follows:

- Approach 1: Adjusting by Expected Bias,
- Approach 2: Bias Corrected and Accelerated Method (BCa),
- Approach 3: Shocking Bootstrap Resampled Observations,
- Approach 4: Using Inter-Quartile Range Estimate, and
- Approaches 1 & 3 combined.

The four performance metrics evaluated were as follows:

- % Reduction in Bias,
- % Reduction in Under-Coverage,
- % Increase in Confidence Interval Width, and
- % Increase in MSE.

For all of these metrics, the reference comparison is the simple percentile bootstrap variance estimate. It is desirable for the first two metrics to be as large as possible, and negative results indicate performance worse the simple percentile bootstrap variance estimate (larger absolute bias or less coverage). It is desirable for the second two metrics to be as small as possible. Negative results for these metrics indicate narrower intervals and smaller MSE values than observed for the simple percentile bootstrap variance estimate.

Table 11 displays a comparison of the approaches by performance metric across all the simulation distributional model and sample size combinations. The cells are color coded from green = desirable performance to red = poor performance within each performance metric across approaches, but also across all the simulation combinations.

Table 11: Comparison of Approaches by Performance Metric

Performance Metric	Approach	Standard Normal			Double Exponential			Exponential			Log Normal		
		10	20	30	10	20	30	10	20	30	10	20	30
% Bias Reduction	Adjust by Expected Bias	93%	99%	74%	66%	81%	-17%	98%	93%	81%	90%	81%	73%
	Bias Corrected & Accelerated	-122%	-276%	-402%	-35%	-515%	-1469%	-100%	-420%	-558%	-89%	-431%	-652%
	Shocking Bootstrap Obs	18%	-226%	-524%	69%	-63%	-518%	75%	-41%	-124%	90%	24%	-32%
	Adjust by Bias & Shock	18%	-349%	-755%	96%	-195%	-886%	22%	-146%	-265%	63%	-66%	-156%
	Use Inter-Quartile Range	-123%	-87%	-173%	-215%	-905%	-2681%	-350%	-681%	-970%	-350%	-681%	-970%
% Undercoverage Reduction	Adjust by Expected Bias	25%	16%	21%	16%	10%	13%	13%	9%	9%	8%	9%	5%
	Bias Corrected & Accelerated	-8%	-16%	-81%	0%	-10%	-11%	-1%	0%	3%	2%	6%	3%
	Shocking Bootstrap Obs	64%	79%	40%	28%	39%	44%	31%	38%	39%	21%	31%	30%
	Adjust by Bias & Shock	67%	72%	13%	32%	50%	58%	35%	42%	46%	25%	38%	39%
	Use Inter-Quartile Range	102%	115%	142%	69%	19%	-40%	65%	44%	15%	54%	20%	-9%
% Increase C.I. Width	Adjust by Expected Bias	11%	5%	3%	11%	5%	3%	11%	5%	3%	11%	5%	3%
	Bias Corrected & Accelerated	-3%	-3%	-7%	2%	1%	0%	0%	5%	6%	2%	9%	11%
	Shocking Bootstrap Obs	31%	30%	26%	16%	11%	9%	19%	13%	11%	14%	10%	8%
	Adjust by Bias & Shock	138%	137%	133%	122%	117%	115%	125%	119%	117%	120%	116%	114%
	Use Inter-Quartile Range	108%	95%	87%	80%	33%	3%	102%	22%	9%	89%	-2%	-18%
% Increase MSE	Adjust by Expected Bias	18%	8%	6%	18%	10%	7%	21%	10%	6%	23%	10%	7%
	Bias Corrected & Accelerated	94%	86%	67%	112%	134%	129%	96%	155%	182%	107%	201%	263%
	Shocking Bootstrap Obs	45%	74%	88%	10%	16%	20%	16%	14%	17%	5%	6%	5%
	Adjust by Bias & Shock	166%	214%	240%	121%	133%	141%	129%	130%	135%	117%	119%	118%
	Use Inter-Quartile Range	25%	64%	79%	-12%	13%	46%	-39%	-25%	-8%	-72%	-54%	-46%

The most obvious conclusions from Table 11 are that Approach 1 – Adjusting by the Expected Bias performs universally best in reducing the low bias in the simple percentile bootstrap variance point estimator. In addition, it is readily apparent that the combination of this approach with shocking the bootstrap samples generates the universally widest confidence intervals (and largest MSE estimates).

Perhaps marginally less obvious is that Approach 2 – the Bias Corrected and Accelerated (BCa) method does the least to improve confidence interval coverage, primarily due to it having the narrowest intervals. It also produces some of the largest MSE values, which given the narrow intervals noted here, is driven primarily due to it also producing point estimates biased significantly to the high side (i.e., its gross over-correction of the low side bias observed for the simple percentile bootstrap variance point estimator).

As noted previously, Approach 4 – Using the Inter-Quartile Range Estimate does produce higher coverage rates for the normal case, as well as for all distributions considered with sample size n = 10. Not surprisingly, the approach requires the widest intervals to obtain these high coverage rates. In addition, this approach appears to perform worst in bias reduction, and contrary to the over-correction of the BCa method, this approach actually has larger low side bias than the simple percentile bootstrap variance estimator.

Perhaps an even more subtle observation in Table 11 is that Approaches 3 – Shocking Bootstrap Observations and the combination of this approach with Approach 1 – Adjusting by Expected Bias generally provide the most improvement in coverage rates across all the simulated conditions. However, as noted above, the interval widths are generally much wider for the combined approaches than when simply shocking the bootstrap observations.

Table 12 simply rearranges the rows of Table 11 (color-coding has not been altered) to provide another perspective on comparing the approaches. In this table, it is fairly obvious that Approach 1 – Adjusting by Expected Bias and Approach 3 – Shocking Bootstrap Observations have no red and generally fewer orange (i.e., fewer cells indicating poor relative performance) than the other approaches. Interestingly, from this perspective, combining these approaches does not generally improve on implementing either one separately, and actually appears to be the poorest performing of the options considered.

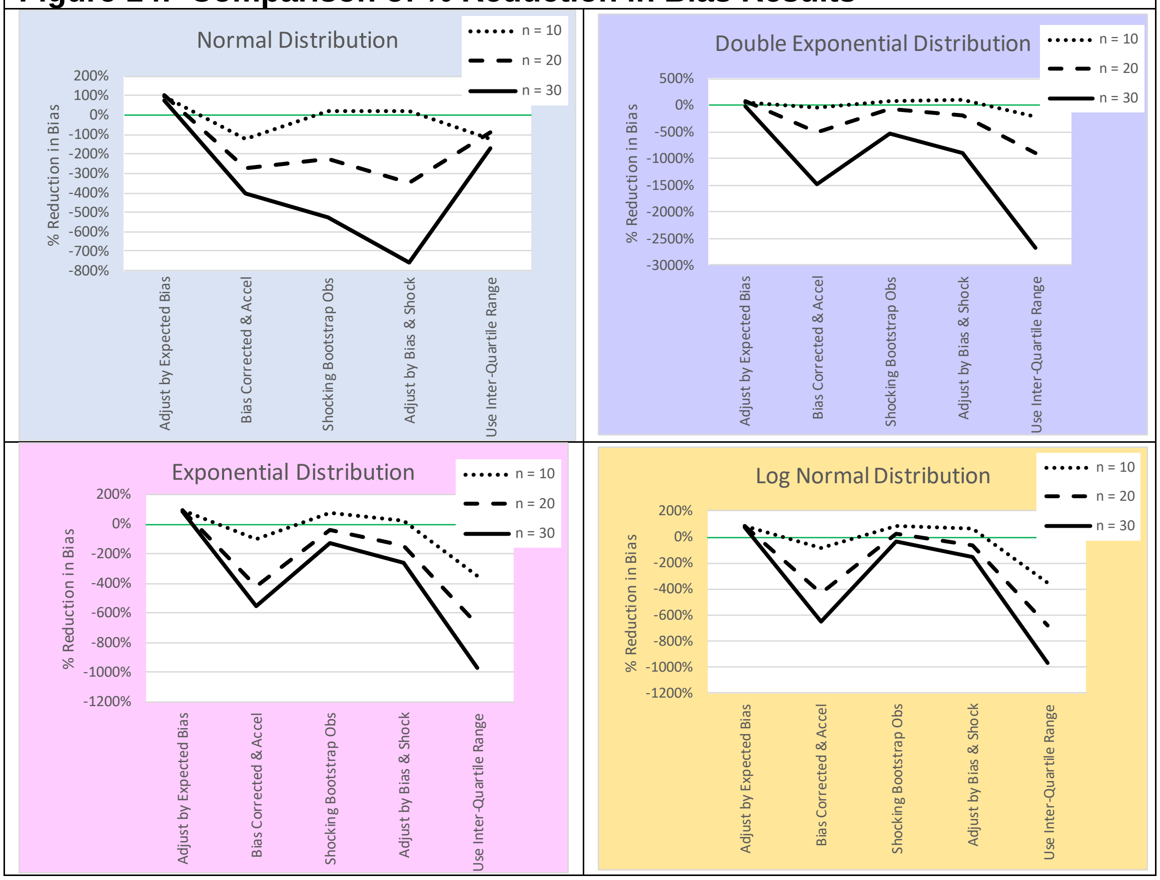
Figures 14 and 15 display the two primary performance metrics - % Bias Reduction and % Reduction in Under-Coverage for all five approaches. From these figures, it is clear that the first approach (Adjusting by Expected Bias) performs best in reducing the low bias in the simple percentile bootstrap variance estimate. However, the approaches involving shocks are the only others that seem to even be reasonably competitive alternatives for this metric.

Table 12: Comparison of Approaches Across Performance Metrics

Approach	Performance Metric	Standard Normal			Double Exponential			Exponential			Log Normal		
		10	20	30	10	20	30	10	20	30	10	20	30
Adjust by Expected Bias	% Bias Reduction	93%	99%	74%	66%	81%	-17%	98%	93%	81%	90%	81%	73%
	% Undercoverage Reduction	25%	16%	21%	16%	10%	13%	13%	9%	9%	8%	9%	5%
	% Increase C.I. Width	11%	5%	3%	11%	5%	3%	11%	5%	3%	11%	5%	3%
	% Increase MSE	18%	8%	6%	18%	10%	7%	21%	10%	6%	23%	10%	7%
Bias Corrected & Accelarated	% Bias Reduction	-122%	-276%	-402%	-35%	-515%	-1469%	-100%	-420%	-558%	-89%	-431%	-652%
	% Undercoverage Reduction	-8%	-16%	-81%	0%	-10%	-11%	-1%	0%	3%	2%	6%	3%
	% Increase C.I. Width	-3%	-3%	-7%	2%	1%	0%	0%	5%	6%	2%	9%	11%
	% Increase MSE	94%	86%	67%	112%	134%	129%	96%	155%	182%	107%	201%	263%
Shocking Bootstrap Observations	% Bias Reduction	18%	-226%	-524%	69%	-63%	-518%	75%	-41%	-124%	90%	24%	-32%
	% Undercoverage Reduction	64%	79%	40%	28%	39%	44%	31%	38%	39%	21%	31%	30%
	% Increase C.I. Width	31%	30%	26%	16%	11%	9%	19%	13%	11%	14%	10%	8%
	% Increase MSE	45%	74%	88%	10%	16%	20%	16%	14%	17%	5%	6%	5%
Adjust by Expected Bias AND Shock	% Bias Reduction	18%	-349%	-755%	96%	-195%	-886%	22%	-146%	-265%	63%	-66%	-156%
	% Undercoverage Reduction	67%	72%	13%	32%	50%	58%	35%	42%	46%	25%	38%	39%
	% Increase C.I. Width	138%	137%	133%	122%	117%	115%	125%	119%	117%	120%	116%	114%
	% Increase MSE	166%	214%	240%	121%	133%	141%	129%	130%	135%	117%	119%	118%
Using Inter-Quartile Range Estimate	% Bias Reduction	-123%	-87%	-173%	-215%	-905%	-2681%	-350%	-681%	-970%	-350%	-681%	-970%
	% Undercoverage Reduction	102%	115%	142%	69%	19%	-40%	65%	44%	15%	54%	20%	-9%
	% Increase C.I. Width	108%	95%	87%	80%	33%	3%	102%	22%	9%	89%	-2%	-18%
	% Increase MSE	25%	64%	79%	-12%	13%	46%	-39%	-25%	-8%	-72%	-54%	-46%

Figure 15 displays the coverage results, and it is clear that the approach the simply adjusts for the expected bias does not do as well for this metric as some of the other alternative approaches. Notably, it always provides less improvement in coverage than the shocking of bootstrap observations. Admittedly, Approach 3 – Shocking Bootstrap Observations – alone does not improve coverage as much as some of the other alternative approaches. However, those approaches that do have higher coverage rates also have wider intervals, and sometimes much wider intervals.

Figure 14: Comparison of % Reduction in Bias Results



In summary, it appears that either

- simply adjusting the simple percentile bootstrap variance estimator by its expected bias (i.e., simply multiply results by $\frac{n}{n-1}$), or

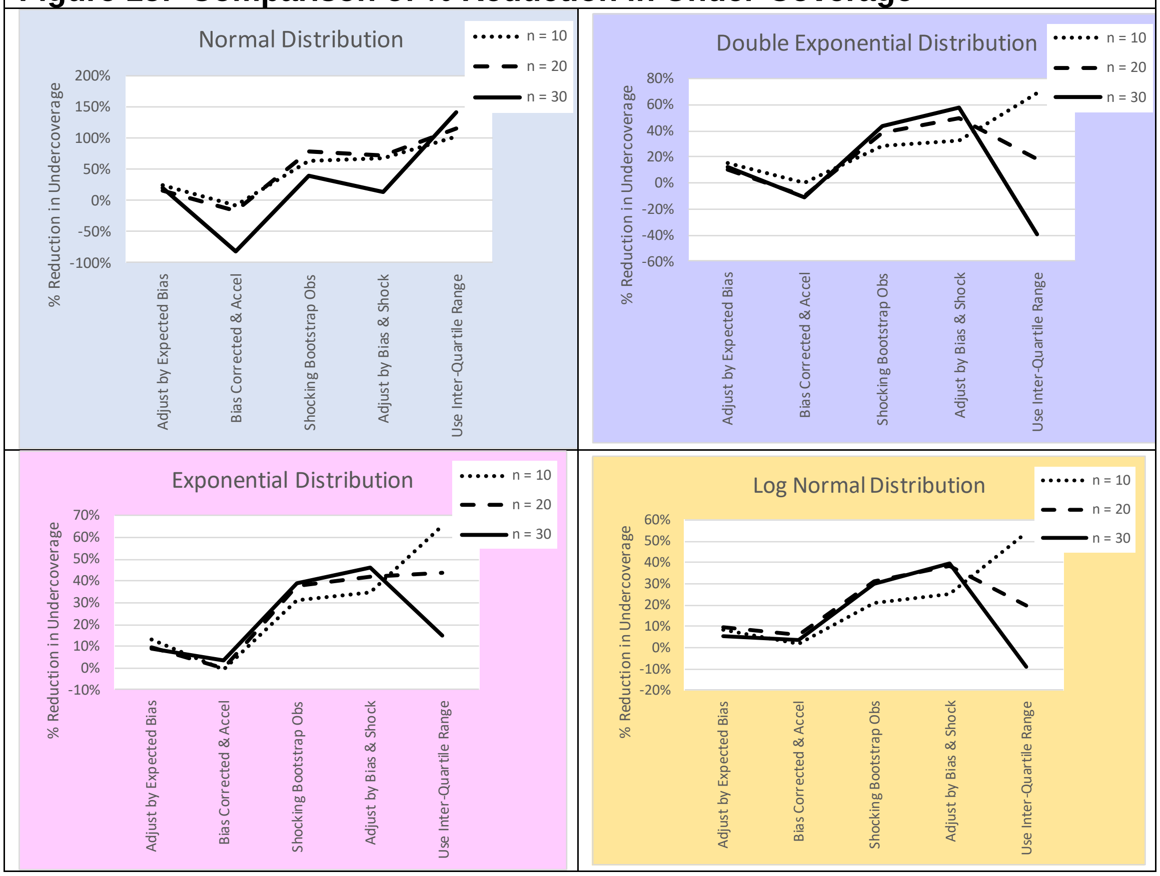
- shocking repeated bootstrap observations using a shock distribution that is normal in shape with a mean of zero and a standard deviation equal to 3/4 of the adjusted original sample IQR (i.e., shocks $\sim Normal(\mu = 0, \sigma = \frac{3}{4}\frac{IQR_{Orig}}{1.35})$)

would provide the generally best performing corrections for the low side bias of the bootstrap point estimate and the under-coverage of the corresponding interval estimate. If the analyst is primarily interested in a point estimate of the population variance, the simple adjustment approach seems the best alternative. However, if an interval estimate is more desirable, then the shocking approach appears to be the better choice.

If the analyst desires both a good point estimate, as well as an interval estimate with the desired coverage level, this work suggests

- adjusting the simple bootstrap percentile variance point estimate for its expected bias (again, simply multiply by $\frac{n}{n-1}$, where n = sample size) to get a less biased point estimator, and
- bootstrapping the IQR and using the associated bootstrap percentiles to obtain a confidence interval for the population variance.

Figure 15: Comparison of % Reduction in Under-Coverage



A 95% confidence interval of the form suggested by the last item above would be given as:

$$\left[\left(\frac{IQR^*_{(0.025)}}{1.35}\right)^2, \left(\frac{IQR^*_{(0.975)}}{1.35}\right)^2\right],$$

where $IQR^*_{(p)}$ = the pth percentile of the IQR results among the bootstrap re-samples. This work suggests the divisor of 1.35 is generally appropriate for smaller samples (e.g., n = 10) and certainly if the analysts believes the data may likely come from a normal or nearly normal distribution. However, for larger sample sizes combined with a belief that the data are not normally distributed, using a divisor of one might be more appropriate.

Future Works

There are a number of potential evaluations that could potentially add to the understanding of how to most appropriately manage the low bias and under-coverage inherent in simple percentile bootstrap variance estimation. One immediate follow-up study could be to refine the values of the multiplier, w, to see which one(s) give the highest coverage, refining the exact value at which the turning point in the number of Missed Low and High occurs (which is probably not exactly 3/4). Clearly, additional non-normal distributions could be evaluated, such as contaminated normal distributions. However, for these distributions, it would likely be necessary to specify whether estimation of the actual distribution variance or of the variance of the primary (i.e., non-contaminating distribution) was the desired focus of the analyst.

As noted above, exploration on the adjustment to the IQR estimator of variance (i.e., smaller divisors than 1.35) might prove interesting for approaches involving shocks to bootstrap observations. Finally, consideration of one-sided intervals (correcting just the upper limits) or tests for population variances based on the bootstrap method might be of interest as frequently analysts might be much more concerned with either over-estimation or under-estimation of the variance due to the penalties of one of these issues being much larger than for the other.

References

Brennan, R. L. (2007). Unbiased Estimates of Variance Components with Bootstrap procedures. Educational and Psychological Measurement, 67, 784–803.

Chernick, M. R. and LaBudde, R. A. (2010). Revisiting Qualms about Bootstrap confidence intervals. Am. J. Math. Manag. Sci. 29, 437 – 456.

Chernick, M. R. and LaBudde, R. (2011). An Introduction to the Bootstrap with Applications in R Wiley, Hoboken.

Efron, B. (1979). Bootstrap methods: Another Look at the jackknife. Ann. Stat. 7, 1 – 26.

Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap, Chapman & Hall,

New York.

Hogg, Robert V.; McKean, Joseph W.; and Craig, Allen T. (2013). "Introduction to Mathematical Statistics".

Hollander, M. & Wolfe, D. A. (1999). Nonparametric Statistical Methods (Second Edition), John Wiley and Sons, Inc., New York.

Schenker, N. (1985). Qualms about Bootstrap Confidence Intervals. J. Am. Stat. Assoc. 80, 360 – 36

Whaley, Dewey L. (2005). The Interquartile Range: Theory and Estimation. Electronic Theses and Dissertations. Paper 1030. http://dc.etsu.edu/etd/1030.

R Codes

```
## Install the "xlsx" package to export the result into Excel
library(xlsx)

#############     BIAS     CORRECTED     AND     ACCELERATED     (BCA)     FUNCTION
####################
## Need to code this manually/override the existing function in R since the bcanon function in
the bootstrap does not give the desired outcome
# ***Ripping out the bcanon function's codes to tweak for what I really want it to do***
    Randy_bcanon = function (x, nboot, theta, ..., alpha = c(0.025, 0.05, 0.1, 0.16,
            .84, 0.9, 0.95, 0.975))
    {
    if (!all(alpha < 1) || !all(alpha > 0))
            stop("All elements of alpha must be in (0,1)")
    alpha_sorted <- sort(alpha)
    if (nboot <= 1/min(alpha_sorted[1], 1 - alpha_sorted[length(alpha_sorted)]))
            warning("nboot is not large enough to estimate your chosen alpha.")
    call <- match.call()
    n <- length(x)
    thetahat <- theta(x, ...)
    bootsam <- matrix(sample(x, size = n * nboot, replace = TRUE), nrow = nboot)
    thetastar <- apply(bootsam, 1, theta, ...)
    z0 <- qnorm(sum(thetastar < thetahat)/nboot)

    u <- rep(0, n)
    for (i in 1:n) {
            u[i] <- theta(x[-i], ...)
    }
    uu <- mean(u) - u
    acc <- sum(uu * uu * uu)/(6 * (sum(uu * uu))^1.5)

    # Need to add the upper zalpha as well #
    Lower_zalpha <- qnorm(alpha)
    Upper_zalpha <- qnorm(1-alpha)

    # Need to modify this code to give upper and lower limit points #
    Lalpha <- pnorm(z0 + (z0 + Lower_zalpha)/(1 - acc * (z0 + Lower_zalpha)))
    Ualpha <- pnorm(z0 + (z0 + Upper_zalpha)/(1 - acc * (z0 + Upper_zalpha)))

    conf_int <- c()

    Lower_Limit <- quantile(x = thetastar, probs = Lalpha, type = 1)
    Upper_Limit <- quantile(x = thetastar, probs = Ualpha, type = 1)

    conf_int <- cbind(Lower_Limit, Upper_Limit)

    return(c(z0, acc, conf_int))
```

```
    }
################################################################################
#######
### Additional analysis is required in Excel ###
original_sample <- data  ######## User input data here!!!!!!!!!!!! ########
original_mean <- mean(original_sample)
original_var <- var(original_sample)
original_sd <- sd(original_sample)
original_IQR <- (quantile(original_sample, c(.75), na.rm=TRUE) - quantile(original_sample, c(.25), na.rm=TRUE))

n <- length(original_sample)
B <- 1000

## Store the bootstrapped variances as outputs from the loop
final_result_boot_var <- c()
## Store the bootstrapped IQRs as outputs from the loop
final_result_boot_IQR <- c()

## Store the results from Remedy 1
final_result_boot_var_remedy1 <- c()

## Store the results from Remedy 2
final_result_boot_var_remedy2 <- c()

## Store the results from Remedy 3
final_result_scenario_38_shocked_IQR <- c()

result_boot_var <- c()
result_boot_IQR <- c()

theta_var <- function(original_sample){var(original_sample)}

## Calculated from Excel
theoretical_IQR <- c(1.34897950039216)

for (i in 1:B){

    ### Percentile Bootstrap ###
    boot_sample <- sample(original_sample, n, replace = TRUE, prob = NULL)
    boot_var <- var(boot_sample)
    boot_IQR <- (quantile(boot_sample, c(.75), na.rm=TRUE) - quantile(boot_sample, c(.25), na.rm=TRUE))

    result_boot_var <- cbind(result_boot_var, boot_var)
    result_boot_IQR <- cbind(result_boot_IQR, boot_IQR)

    ############# REMEDY 3 ###################
```

```
## Bootstrapping with shocks ONLY ON REPEATED DATA POINTS but with DIFFERENT
shock_size's

## extract unique elements
uniques <- boot_sample[!duplicated(boot_sample)]
## extract duplicate elements
repeats <- boot_sample[duplicated(boot_sample)]

#38 Repeated-Normal(0, 3/4*Orig.IQR.)
scenario_38_shock_size <- (3/4)*sqrt((original_IQR/theoretical_IQR)^2)
scenario_38_noise_repeats_shocked<-repeats+rnorm(length(repeats),0,
scenario_38_shock_size)
scenario_38_shocked_sample <- c(uniques, scenario_38_noise_repeats_shocked)
scenario_38_shocked_IQR <- var(scenario_38_shocked_sample)
result_scenario_38_shocked_IQR<-cbind(result_scenario_38_shocked_IQR,
scenario_38_shocked_IQR)


scenario_38_shocked_min_IQR <- min(result_scenario_38_shocked_IQR)
scenario_38_shocked_2.5tile_IQR <- quantile(result_scenario_38_shocked_IQR, c(.025),
na.rm=TRUE)
scenario_38_shocked_25tile_IQR <- quantile(result_scenario_38_shocked_IQR, c(.25),
na.rm=TRUE)
scenario_38_shocked_mean_IQR <- mean(result_scenario_38_shocked_IQR)
scenario_38_shocked_median_IQR <- median(result_scenario_38_shocked_IQR)
scenario_38_shocked_75tile_IQR <- quantile(result_scenario_38_shocked_IQR, c(.75),
na.rm=TRUE)
scenario_38_shocked_97.5tile_IQR<-quantile(result_scenario_38_shocked_IQR, c(.975),
na.rm=TRUE)
scenario_38_shocked_max_IQR<- max(result_scenario_38_shocked_IQR)
scenario_38_shocked_less_IQR<-
sum(length(result_scenario_38_shocked_IQR[result_scenario_38_shocked_IQR<
original_var]))
scenario_38_shocked_vector_IQR<-c(scenario_38_shocked_min_IQR,
scenario_38_shocked_2.5tile_IQR,scenario_38_shocked_25tile_IQR,
scenario_38_shocked_mean_IQR,
        scenario_38_shocked_median_IQR,scenario_38_shocked_75tile_IQR,
scenario_38_shocked_97.5tile_IQR,scenario_38_shocked_max_IQR,
scenario_38_shocked_less_IQR)
final_result_scenario_38_shocked_IQR <- rbind(final_result_scenario_38_shocked_IQR,
scenario_38_shocked_vector_IQR)
}

result_scenario_38_shocked_IQR <- c()

## Calculating relevant summary statistics for variances from original bootstrap
boot_sample_min_var <- min(result_boot_var)
boot_sample_2.5tile_var <- quantile(result_boot_var, c(.025), na.rm=TRUE)
boot_sample_25tile_var <- quantile(result_boot_var, c(.25), na.rm=TRUE)
boot_sample_mean_var <- mean(result_boot_var)
```

```
    boot_sample_median_var <- median(result_boot_var)
    boot_sample_75tile_var <- quantile(result_boot_var, c(.75), na.rm=TRUE)
    boot_sample_97.5tile_var <- quantile(result_boot_var, c(.975), na.rm=TRUE)
    boot_sample_max_var <- max(result_boot_var)
    boot_sample_less_var <- sum(length(result_boot_var[result_boot_var < original_var]))
    boot_sample_vector_var<-c(boot_sample_min_var,boot_sample_2.5tile_var,
boot_sample_25tile_var, boot_sample_mean_var,
            boot_sample_median_var,  boot_sample_75tile_var,  boot_sample_97.5tile_var,
boot_sample_max_var, boot_sample_less_var)
    final_result_boot_var <- rbind(final_result_boot_var, boot_sample_vector_var)

    ## Calculating relevant summary statistics for IQR from original bootstrap
    ## to be used in REMEDY 4 ####
    boot_sample_min_IQR <- min(result_boot_IQR)
    boot_sample_2.5tile_IQR <- quantile(result_boot_IQR, c(.025), na.rm=TRUE)
    boot_sample_25tile_IQR <- quantile(result_boot_IQR, c(.25), na.rm=TRUE)
    boot_sample_mean_IQR <- mean(result_boot_IQR)
    boot_sample_median_IQR <- median(result_boot_IQR)
    boot_sample_75tile_IQR <- quantile(result_boot_IQR, c(.75), na.rm=TRUE)
    boot_sample_97.5tile_IQR <- quantile(result_boot_IQR, c(.975), na.rm=TRUE)
    boot_sample_max_IQR <- max(result_boot_IQR)
    boot_sample_less_IQR <- sum(length(result_boot_IQR[result_boot_IQR < original_IQR]))
    boot_sample_vector_IQR<-c(boot_sample_min_IQR,boot_sample_2.5tile_IQR,
boot_sample_25tile_IQR, boot_sample_mean_IQR,
            boot_sample_median_IQR, boot_sample_75tile_IQR, boot_sample_97.5tile_IQR,
boot_sample_max_IQR, boot_sample_less_IQR)
    final_result_boot_IQR <- rbind(final_result_boot_IQR, boot_sample_vector_IQR)

    ############# n/(n-1) REMEDY 1 ####################
    ## Adjustment by n/(n-1) factor
    result_boot_var_remedy1 <- (n/(n-1))*result_boot_var

    boot_sample_min_var_remedy1 <- min(result_boot_var_remedy1)
    boot_sample_2.5tile_var_remedy1<-quantile(result_boot_var_remedy1,c(.025),
na.rm=TRUE)
    boot_sample_25tile_var_remedy1<-quantile(result_boot_var_remedy1,c(.25),
na.rm=TRUE)
    boot_sample_mean_var_remedy1 <- mean(result_boot_var_remedy1)
    boot_sample_median_var_remedy1 <- median(result_boot_var_remedy1)
    boot_sample_75tile_var_remedy1<-quantile(result_boot_var_remedy1,c(.75),
na.rm=TRUE)
    boot_sample_97.5tile_var_remedy1<-quantile(result_boot_var_remedy1,c(.975),
na.rm=TRUE)
    boot_sample_max_var_remedy1 <- max(result_boot_var_remedy1)
    boot_sample_less_var_remedy1<-
sum(length(result_boot_var_remedy1[result_boot_var_remedy1 < original_var]))
    boot_sample_vector_var_remedy1<-c(boot_sample_min_var_remedy1,
boot_sample_2.5tile_var_remedy1,boot_sample_25tile_var_remedy1,
boot_sample_mean_var_remedy1,
```

```
            boot_sample_median_var_remedy1,boot_sample_75tile_var_remedy1,
boot_sample_97.5tile_var_remedy1,boot_sample_max_var_remedy1,
boot_sample_less_var_remedy1)
    final_result_boot_var_remedy1<-rbind(final_result_boot_var_remedy1,
boot_sample_vector_var_remedy1)
    #####################################################

    ############# BCa REMEDY 2 ########################
    # Use "rbind"/"cbind" functions to combine these vectors from Randy_bcanon function into
a matrix
    boot_BCa_var <- Randy_bcanon(original_sample, nboot = 1000, theta = theta_var, alpha
= 0.05)

    final_result_boot_var_remedy2 <- rbind(final_result_boot_var_remedy2, boot_BCa_var)
    ######################################################




rownames(final_result_boot_var) <- c()
rownames(final_result_boot_IQR) <- c()

## Export the result into an Excel file
write.xlsx(final_result_boot_var, "D:/Master Thesis/MasterThesisBootVAR.xlsx")
#write.xlsx(final_result_boot_var, "E:/Master Thesis/MasterThesisBootVAR.xlsx")

write.xlsx(final_result_boot_IQR, "D:/Master Thesis/MasterThesisBootIQR.xlsx")
#write.xlsx(final_result_boot_IQR, "E:/Master Thesis/MasterThesisBootIQR.xlsx")

##### REMEDY 1 RESULTS FOR EVALUATIONS ######################
rownames(final_result_boot_var_remedy1) <- c()
## Export the result into an Excel file
write.xlsx(final_result_boot_var_remedy1, "D:/Master Thesis/PercentileBootRemedy1.xlsx")
#write.xlsx(final_result_boot_var_remedy1, "E:/Master Thesis/PercentileBootRemedy1.xlsx")
###############################################################

##### REMEDY 2 RESULTS FOR EVALUATIONS ######################
rownames(final_result_boot_var_remedy2) <- c()
## Export the result into an Excel file
write.xlsx(final_result_boot_var_remedy2, "D:/Master Thesis/PercentileBootRemedy2.xlsx")
#write.xlsx(final_result_boot_var_remedy2, "E:/Master Thesis/PercentileBootRemedy2.xlsx")
###############################################################

##### REMEDY 3 RESULTS FOR EVALUATIONS ######################
rownames(final_result_scenario_38_shocked_IQR) <- c()
## Export the result into an Excel file
write.xlsx(final_result_scenario_38_shocked_IQR,"D:/Master
Thesis/PercentileBootRemedy3scenario38.xlsx")
#write.xlsx(final_result_scenario_38_shocked_IQR,"E:/Master
Thesis/PercentileBootRemedy3scenario38.xlsx")
```

Appendix

Tables of Bias, Coverage, Interval Width, and MSE for All Approaches

Approach 1 – Adjusting by Expected Bias

Table 1	Approach 1 Bias (Mean)		
Distribution	n		
	10	20	30
Normal	0.00630	0.00025	0.00699
DoubleExp	-0.04960	0.00789	0.01827
Exponential	-0.00176	-0.00379	-0.00762
LogNormal	-0.01147	-0.01146	-0.01219

Table 2	Approach 1 Coverage		
Distribution	**n**		
	10	20	30
Normal	819	873	908
DoubleExp	701	791	831
Exponential	629	719	762
LogNormal	547	663	703

Table 3	Approach 1 Conf. Int. Width		
Distribution	n		
	10	20	30
Normal	1.49758	1.14479	0.95787
DoubleExp	1.69090	1.48488	1.31678
Exponential	1.81923	1.60975	1.44337
LogNormal	1.93023	1.73693	1.61403

Table 4	Approach 1 MSE ($bias^2 + var_B$)		
Distribution	n		
	10	20	30
Normal	0.24704	0.10440	0.07091
DoubleExp	0.47484	0.26531	0.18743
Exponential	0.74771	0.39942	0.27114
LogNormal	1.97640	0.80625	0.58624

Approach 2: Bias Corrected and Accelerated Method (BCa)

Table 5	Approach 2 Bias (Mean)		
Distribution	n		
	10	20	30
Normal	0.20967	0.18712	0.13343
DoubleExp	0.19558	0.26119	0.24588
Exponential	0.20363	0.27847	0.26780
LogNormal	0.20813	0.32312	0.33935

Table 6	Approach 2 Coverage		
Distribution	n		
	10	20	30
Normal	761	843	854
DoubleExp	656	755	799
Exponential	577	694	750
LogNormal	519	652	698

Table 7	Approach 2 Conf. Int. Width		
Distribution	n		
	10	20	30
Normal	1.30215	1.05087	0.86344
DoubleExp	1.55035	1.42212	1.27206
Exponential	1.64091	1.60699	1.47308
LogNormal	1.77523	1.80491	1.73642

Table 8	Approach 2 MSE ($bias^2 + var_B$)		
Distribution	n		
	10	20	30
Normal	0.40605	0.18029	0.11198
DoubleExp	0.85432	0.56338	0.40106
Exponential	1.20558	0.92814	0.71915
LogNormal	3.33809	2.20214	1.99457

Approach 3: Shocking Bootstrap Resampled Observations

Table 9	Approach 3 Bias (Mean)		
Distribution	n		
	10	20	30
Normal	0.07748	0.16223	0.16594
DoubleExp	-0.04428	0.06928	0.09688
Exponential	0.02572	0.07544	0.09131
LogNormal	-0.01128	0.04610	0.05976

Table 10	Approach 3 Coverage		
Distribution	n		
	10	20	30
Normal	887	931	918
DoubleExp	739	842	874
Exponential	695	791	823
LogNormal	602	731	767

Table 11	Approach 3 Conf. Int. Width		
Distribution	n		
	10	20	30
Normal	1.77056	1.41517	1.17078
DoubleExp	1.75823	1.56459	1.38999
Exponential	1.94674	1.72676	1.54421
LogNormal	1.97956	1.82107	1.68925

Table 12	Approach 3 MSE ($bias^2 + var_B$)		
Distribution	n		
	10	20	30
Normal	0.30359	0.16808	0.12607
DoubleExp	0.44393	0.27938	0.21074
Exponential	0.71412	0.41499	0.29872
LogNormal	1.69473	0.77579	0.57530

Approach 4: Using Inter-Quartile Range Estimate

Table 13	Approach 4 Bias (Mean)		
Distribution	n		
	10	20	30
Normal	-0.21055	-0.09314	-0.07264
DoubleExp	-0.45599	-0.42711	-0.43590
Exponential	-0.39171	-0.37300	-0.36688
LogNormal	-0.49595	-0.47565	-0.48263

Table 14	Approach 4 Coverage		
Distribution	n		
	10	20	30
Normal	954	964	972
DoubleExp	859	806	760
Exponential	821	806	774
LogNormal	748	695	666

Table 15	Approach 4 Conf. Int. Width		
Distribution	n		
	10	20	30
Normal	2.80318	2.12050	1.72747
DoubleExp	2.74021	1.87026	1.30622
Exponential	3.30498	1.87026	1.51938
LogNormal	3.28921	1.60958	1.28532

Table 16	Approach 4 MSE ($bias^2 + var_B$)		
Distribution	n		
	10	20	30
Normal	0.26070	0.15814	0.12001
DoubleExp	0.35503	0.27208	0.25619
Exponential	0.37748	0.27330	0.23458
LogNormal	0.45274	0.33585	0.29523

Approaches 1 and 3 Combined

Table 17	Approach 1 & 3 Bias (Mean)		
Distribution	n		
	10	20	30
Normal	0.13419	0.22340	0.22730
DoubleExp	0.00602	0.12556	0.15461
Exponential	0.07971	0.13204	0.14875
LogNormal	0.04076	0.10116	0.11553

Table 18	Approach 1 & 3 Coverage		
Distribution	n		
	10	20	30
Normal	893	924	904
DoubleExp	750	862	893
Exponential	709	802	838
LogNormal	622	754	792

Table 19	Approach 1 & 3 Conf. Int. Width		
Distribution	n		
	10	20	30
Normal	1.86375	1.48966	1.23240
DoubleExp	1.85077	1.64693	1.46315
Exponential	2.04920	1.81764	1.62549
LogNormal	2.08374	1.91692	1.77815

Table 20	Approach 1 & 3 MSE ($bias^2 + var_B$)		
Distribution	n		
	10	20	30
Normal	0.34774	0.20699	0.16085
DoubleExp	0.48975	0.32001	0.24701
Exponential	0.79689	0.47095	0.34388
LogNormal	1.87933	0.86748	0.64684

Vita

After completing his work at the Duncanville High School, Duncanville, Texas, in 2013, Nghia (Randy) Nguyen entered Stephen F. Austin State University at Nacogdoches, Texas. He received the degree of Bachelor of Science from Stephen F. Austin State University in May 2016. Immediate after completion of his undergraduate degree, Randy entered the Graduate School of Stephen F. Austin State University, and received the degree of Master of Science in May 2018.

Permanent Address: 5728 Goldfinch Way, Dallas, TX 75249

Style manual designation: Chicago Manual of Style

This thesis was typed by Nghia T. Nguyen