On Small Samples and the Use of Robust Estimators in Loss Reserving

Hou-wen Jeng*

Abstract

This paper explores the use of robust location estimators such as Average-Excluding-High-And-Low and Huber's M-estimators in loss reserving. Standard order statistics results are used to investigate the finite-sample properties of Average-Excluding-High-And-Low for positively skewed distributions including bias and efficiency, based on the criterion of mean squared error. The paper concludes that Averages-Excluding-High-And-Low, although biased with respect to the population mean for positively skewed distributions, is more efficient than the sample average in small samples. The paper also shows that the use of Huber's M-estimators can enhance the consistency in loss development factor selections by identifying the implied risk preference.

Keywords: Robust Estimators; Order Statistics; Averages-Excluding-High-And-Low; Huber's M-Estimators; Loss Reserving.

1 Introduction

In practice, actuarial data are usually plagued by two problems: heterogeneity and small sample sizes. Heterogeneity refers to the fact that the underlying exposures consist of policies with vastly different statistical properties, either within a rating period or between different periods. For example, losses from separate policies may follow different probability distributions, or follow the same type of distribution but with different parameters. Actuaries try hard to adjust the data by using trend factors, rate change history, and other cross-section and time series factors. After these adjustments, in many instances, doubts may still linger as to whether more adjustments are needed to make the data homogeneous.

*Thanks are due to David Homer for discussion, and to Elizabeth Smith of the CAS for editorial review. The usual disclaimer applies.

Heterogeneity, which usually renders the data from older years obsolete, may exacerbate the problem of small sample sizes. A typical example is that the insurer changes its underwriting focus and the current policy mix becomes drastically different from those just a few years before. As a result, when it comes to estimating loss development factors or loss ratios, one rarely can have more than a dozen quality data points. That poses difficult problems in parameter estimation, confidence interval calculation, and hypothesis testing.

A recent paper by Blumsohn and Laufer [2] describes in great detail such dilemmas faced by casualty actuaries. The authors asked a group of actuaries to select loss development factors for an umbrella incurred loss triangle. The methods used by the participants were tabulated and the resulting estimated reserves compared. They found that, due largely to the instability of the loss development, the number of approaches and the selected factors varied widely. They concluded that (1) actuaries should keep an open mind and to approach unstable triangles from a variety of perspectives, and (2) if the selected factors or the fitted model differ significantly from the sample average, one must be sure there is a good reason for the discrepancy.

Blumsohn and Laufer also noted that the majority of the participants were using a variety of averaging methods such as loss weighted averages and Average-Excluding-High-And-Low ($\bar{A}_{xHL}$), which calculates the sample average after discarding the sample maximum and sample minimum. Notice that these methods are equivalent to either down-weighting or rejecting outliers. In the case of $\bar{A}_{xHL}$, the sample maximum and minimum are automatically identified as outliers and excluded. $\bar{A}_{xHL}$ is widely used by practicing actuaries in the estimation of loss development factors and loss ratios despite the potential downward bias pointed out by Wu [12], who argues that if the data exhibit a long-tailed property as they do in most of the insurance loss distributions, the use of $\bar{A}_{xHL}$ would lead to downward bias when compared to the sample average.

Wu's argument seems to be consistent with most of the current actuarial methodologies, which focus mainly on estimating the population means of the underlying distributions, with a clear preference for unbiased estimators. Naturally, the most frequent choice is the sample average due to its simplicity and unbiasedness. However, from a modern robust statistics point of view, the sample average is probably the worst estimator for the population mean. The sample average is not robust in the sense that it takes only one outlier to make the sample average arbitrarily large or small. Thus it is not difficult to understand why Averages-Excluding-High-And-Low are popular with actuaries since in many instances (particularly when the sample sizes are small), the necessity of eliminating extreme outliers seems to outweigh the consequences of possible downward bias. But, is $\bar{A}_{xHL}$ just a convenient escape route for actuaries when facing selection dilemmas? Or are there instances where one can justifiably select $\bar{A}_{xHL}$ over unbiased estimators such as the sample average?

Robust statistics studies the construction of statistical methods and estimators that can produce reliable parameter estimates and that are less sensitive to sample outliers (see Maronna et al. [10]). The idea of robust statistics also stems from the fact that the underlying distribution may not always be correctly specified and the existence of outliers may be the result of contaminated data. In these circumstances, robust estimators can often perform better and are more efficient in terms of variance or mean squared error than, say, the sample average.

This paper tries to argue that in the case of small samples and skewed distributions the use of robust estimators is even more valuable and can help the analyst make difficult selections. The goal here is to rationalize the use of $\bar{A}_{xHL}$ and Huber's M-estimator in loss reserving by providing evidence from the statistics literature on theoretical grounds, and constructing examples to show its relative efficiency in the context of small samples and skewed distributions. The main reasons of using Huber's M-estimator are its relative simplicity and ease of calculation. In addition, the analyst's selection of the critical value in Huber's M-estimator may also reveal his or her risk preference in identifying outliers.

Section 2 explores the implications of small sample sizes, while the standard results from order statistics are used in Section 3 to investigate the finite-sample properties of $\bar{A}_{xHL}$. The means and variances of $\bar{A}_{xHL}$ are calculated and compared with those of the sample average for four positively skewed distributions, namely exponential, Weibull, lognormal, and Pareto. It shows by example that the mean squared error of $\bar{A}_{xHL}$ can be smaller than that of the sample average for positively skewed distributions, and thus more efficient than the sample average despite its downward bias with respect to the population mean. Section 4 discusses the general properties of Huber's M-estimator and its use in loss development factor selections. Section 5 uses the incurred loss triangle from Blumsohn and Laufer [2] to illustrate the merit of $\bar{A}_{xHL}$ and Huber's M-estimator when volatility is the main issue. Section 6 provides a summary of the publicly available softwares in Excel and R that calculate Huber's M-estimators. The concluding remarks are in Section 7.

2 Outliers and Small Samples

Since the sample average can be significantly altered by outliers, the positive skewness of the underlying distribution can exacerbate the outlier problem as outliers may be coming further from the right tail. In the case of small samples, the potential influence of outliers on the sample average is even greater than those in large samples as the weight of each observation is larger. One might think that the impact of the outliers from both tails of the distribution on the sample average may cancel out each other. This may be true for symmetric distributions in a relatively large sample. But most of the actuarial applications considered here involve small samples that presumably are drawn from positively skewed, heavy-tailed distributions with non-negative support, such as lognormal or Pareto distributions. Thus outliers from the right tail, if present in the sample, tend to be larger and their effect on the sample average more significant.

Table 1 shows the probability, by sample size, of having at least one outlier from the right tail of an independent sample when outliers are defined as data points greater than either the 95th percentile or the 90th percentile of the underlying distribution.[1] Given the measurable chance for outliers in small samples, the sample average may not be a reliable estimator for the population mean if the underlying distribution is heavy-tailed.

Table 1 : Probability of At Least One Outlier from the Right Tail in A Sample of Size n

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
Outlier defined as $\geq$ 95th percentile	22.6%	26.5%	30.2%	33.7%	37.0%	40.1%
Outlier defined as $\geq$ 90th percentile	41.0%	46.9%	52.2%	57.0%	61.3%	65.1%

There are other problems associated with small samples from skewed distributions. For example, Fleming [5] warns that the sample average of a small sample from a positively skewed distribution is most likely smaller than the population mean. In other words, the skewness of the parent distribution can be carried over to the sampling distribution of $\bar{X}$. The statistics literature provided an elegant explanation of this phenomenon nearly 70 years ago through the Berry-Esseen theorem, [2] [3] which says that the largest difference between the sampling distribution function of $\bar{X}$ and the standard normal distribution (its limiting distribution) is bounded by a ratio of the skewness of the underlying distribution to the square root of the sample size. In short, it simply means that the larger the skewness, the slower the speed of convergence to normality. Thus in order to achieve a certain level of sampling precision, the sample average may require a considerably large

[1]The probability is $(1-(0.95)^n)$ or $(1-(0.90)^n)$.

[2]David Homer pointed out this fact to me.

[3]Formally, let $X_1, ..., X_n$ be i.i.d. with $E(X_1)=\mu$, $Var(X_1)=\sigma^2$, and $\beta_3 = E(|X_1-\mu|^3) < \infty$. Then there exists a constant C, independent of n or the distribution of the X_i, such that

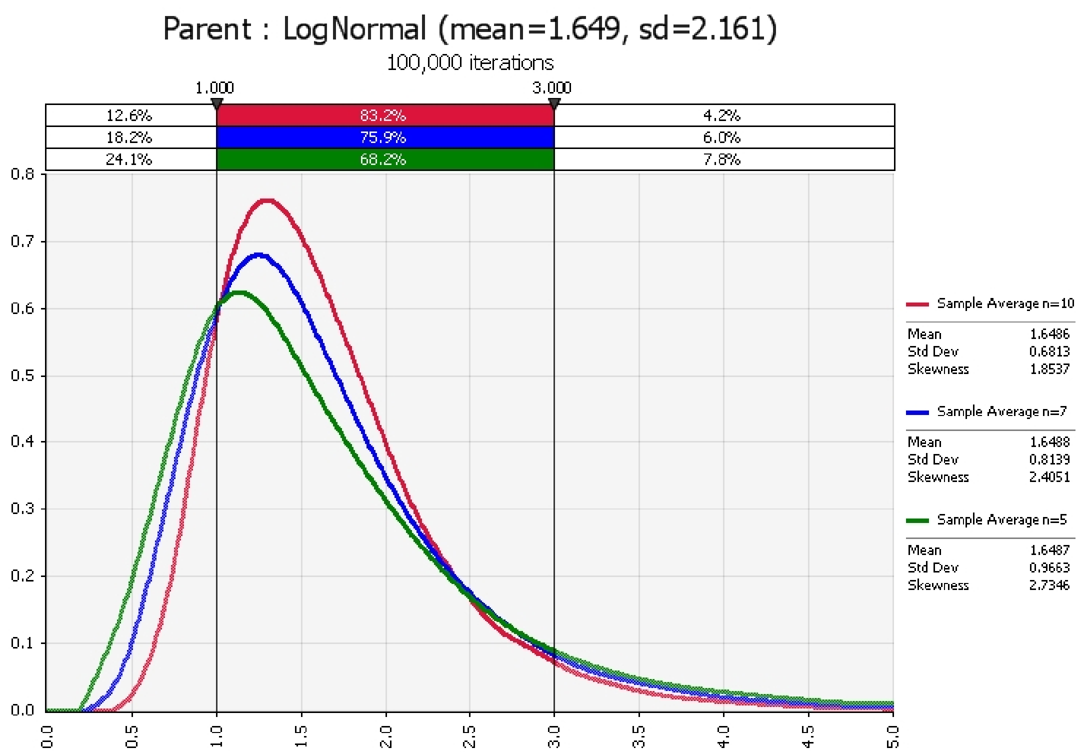
$$\sup_X \left| P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq x) - \Phi(x) \right| \leq \frac{C\beta_3/\sigma^3}{\sqrt{n}},$$

where β_3/σ^3 is the skewness and $C \leq 0.7655$.

sample size to compensate for the skewness of the parent distribution. This issue is of a different nature from the outlier problem. The solution seems to either get a larger sample or use certain transformation methods to get around the skewness problem. From a robust statistics point of view, the outlier problem exists in samples of all sizes. For small samples, however, the choice of the estimator (either robust or non-robust) may have a more significant impact on the final results.

The following graph shows the simulated results for the sampling distributions of $\bar{X}$ from a lognormal parent distribution (mean = 1.649, sd = 2.161, or $\mu = 0$, $\sigma = 1$) with different sample sizes (n=10, 7, and 5). Notice the gradual increase in skewness (thicker tail) when the sample size decreases from 10 to 5.

Graph 1 : Sampling Distributions of $\bar{X}$ (n = 10, 7, and 5)



3 $\bar{A}_{xHL}$: A Robust Estimator

Trimmed means, which are considered robust estimators for location parameters, calculate the sample average after discarding a fixed *number* or a fixed *percentage* of the observations from both ends of an ordered sample. Trimmed means are less sensitive to outliers compared to the sample average $\bar{X}$. Trimmed means come in many varieties, and their statistical properties as well as asymptotic behavior are studied extensively in the statistics literature (see Maronna et al. [10] and Wilcox [11]). The use of Average-Excluding-High-And-Low $\bar{A}_{xHL}$ in actuarial practice is a classical example of trimmed means. It is obvious that $\bar{A}_{xHL}$ is just a special case of trimmed means, where only the sample maximum and minimum are discarded.

3.1 Finite-Sample Statistics of $\bar{A}_{xHL}$

In this section we calculate the finite-sample mean and variance of $\bar{A}_{xHL}$ while the asymptotic properties of $\bar{A}_{xHL}$ are explored in Section 3.3. Let $(X_1, \ldots, X_n)$ be a sample of n independent and identically distributed random variables. Denote the cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$ with mean μ and variance σ^2 (subject to existence).

Let $X_{(i)}$ be the ith order statistic of $(X_1, \ldots, X_n)$. Thus $X_{(1)} = min(X_1, \ldots, X_n)$, $X_{(n)} = max(X_1, \ldots, X_n)$, and $X_{(1)} \leq \ldots \leq X_{(i)} \leq \ldots \leq X_{(n)}$ for $1 \leq i \leq n$. Average-Excluding-High-And-Low $\bar{A}_{xHL}$ is defined as

$$\bar{A}_{xHL} = \frac{\sum_{i=1}^{n} X_i - X_{(1)} - X_{(n)}}{n-2}.$$

The mean and the variance of $\bar{A}_{xHL}$ when the sample size is n are

$$E(\bar{A}_{xHL}) = \frac{n\mu - E(X_{(1)}) - E(X_{(n)})}{n-2} \tag{1}$$

and

$$Var(\bar{A}_{xHL}) = Var\left\{\frac{\sum_{i=2}^{n-1} X_{(i)}}{n-2}\right\} = \frac{\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} Cov(X_{(i)}, X_{(j)})}{(n-2)^2}, \tag{2}$$

respectively. Note that although all observations are i.i.d., the order statistics (i.e., $X_{(i)}$) of an independent sample are correlated with one another.

3.2 Bias and Relative Efficiency of $\bar{A}_{xHL}$

Next, define $Bias$ of $\bar{A}_{xHL}$ with respect to the population mean μ as

$$Bias(\bar{A}_{xHL}; \mu) = \frac{E(\bar{A}_{xHL}) - \mu}{\mu} = \frac{2\mu - E(X_{(1)}) - E(X_{(n)})}{(n-2)\mu}.$$

It can be shown that $2\mu = (E(X_{(1)}) + E(X_{(n)}))$ for symmetric distributions and $2\mu < (E(X_{(1)}) + E(X_{(n)}))$ for positively skewed distributions. For the latter case, it implies $Bias(\bar{A}_{xHL}) < 0$.

Define MSE with respect to the population mean μ as

$$\begin{aligned} MSE(\bar{A}_{xHL};\mu) &= Var(\bar{A}_{xHL}) + \left\{Bias(\bar{A}_{xHL};\mu) * \mu\right\}^2 \\ &= Var(\bar{A}_{xHL}) + \left\{\frac{2\mu - E(X_{(1)}) - E(X_{(n)})}{(n-2)}\right\}^2 \end{aligned}$$

and the relative efficiency between $\bar{X}$ and $\bar{A}_{xHL}$ with respect to the population mean μ as

$$REff(\bar{X}, \bar{A}_{xHL};\mu) = \frac{MSE(\bar{X};\mu)}{MSE(\bar{A}_{xHL};\mu)} = \frac{Var(\bar{X})}{Var(\bar{A}_{xHL}) + \left\{\frac{2\mu - E(X_{(1)}) - E(X_{(n)})}{(n-2)}\right\}^2}.$$

$Bias$ is a common way to quantify the distance between an estimator and a parameter while MSE is a widely accepted measure of accuracy for estimators with respect to a parameter. Traditionally, the efficiency measure is a ratio between the Cramér-Rao lower bound and the variance of an unbiased estimator. Here, however, $REff(\bar{X}, \bar{A}_{xHL};\mu)$ is narrowly defined to compare the MSEs of the sample average $\bar{X}$ and $\bar{A}_{xHL}$. Note that if the underlying distribution is skewed, $\bar{A}_{xHL}$ is always biased with respect to the population mean. As such, MSE may be a more appropriate measure in comparing $\bar{X}$ and $\bar{A}_{xHL}$ as it penalizes the estimator for its deviation from the parameter μ.

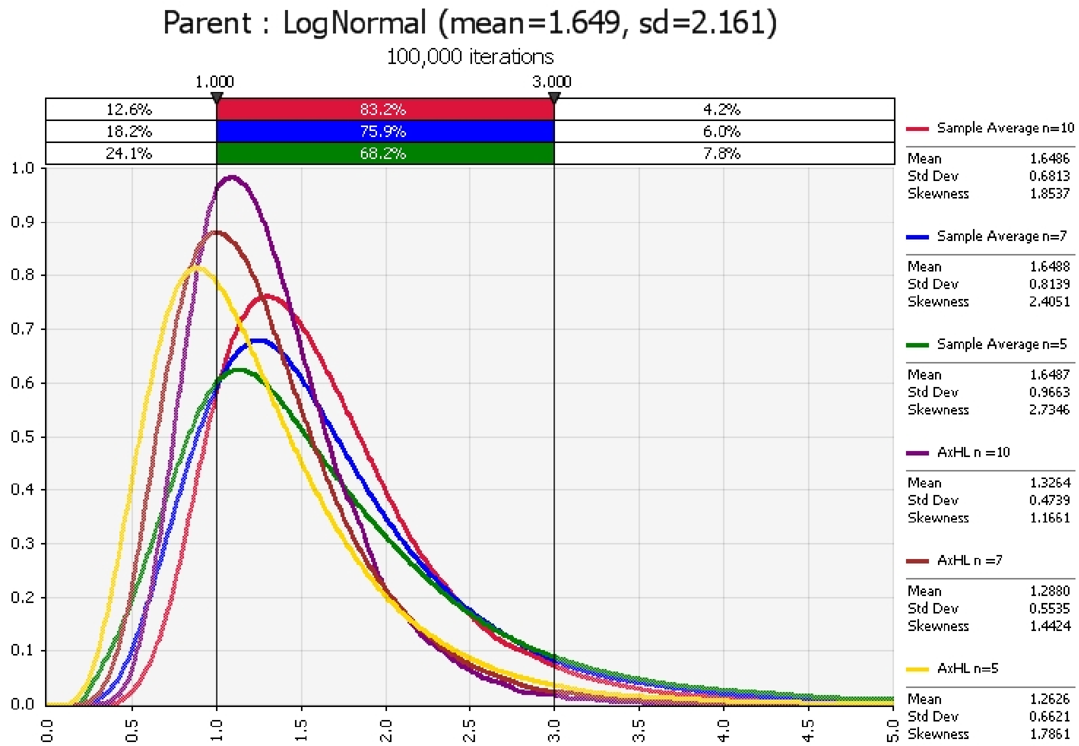
In the appendix, the mean, variance, $Bias$, MSE, $Asym$, and $REff$ of $\bar{A}_{xHL}$ from five distributions are calculated for sample sizes from five to ten as shown in Table 2. The distributions range from symmetric (standard normal), light-tailed (exponential) to positively skewed and heavy-tailed (lognormal, Pareto) distribution. The selections of the parameter values are subjective as the goals are to illustrate the influence of sample size on $E(\bar{A}_{xHL})$ and $Var(\bar{A}_{xHL})$ and to contrast their differences with $E(\bar{X})$ and $Var(\bar{X})$, respectively.

Table 2 : Means, Variances, Coefficients of Variation and Skewness of Selected Distributions

Distribution	Mean	Variance	Coeff. Vari.	Skewness
Standard Normal	0	1	N.A.	0
Exponential ($\theta = 1$)	1	1	100%	2
Pareto ($\theta = 1$, $\alpha = 4$)	1.333	0.222	35%	7.071
LogNormal ($\mu = 0, \sigma^2 = 1$)	1.649	4.671	131%	6.185
Weibull (θ=1, τ=0.5)	2	20	224%	6.618

Overlaying on Graph 1, Graph 2 shows the simulated results for the sampling distributions of $\bar{A}_{xHL}$ from a lognormal parent distribution with sample sizes of 10, 7, and 5. Note the differences in skewness (thicker tail) and standard deviation between the corresponding distributions of $\bar{X}$ and $\bar{A}_{xHL}$ with the same sample size.

Graph 2 : Sampling Distributions of $\bar{X}$ and $\bar{A}_{xHL}$ (n = 10, 7, and 5)



As indicated earlier, $\bar{A}_{xHL}$ are biased downward in positively skewed distributions. The degree of the bias depends on the shape parameter and the sample size. The larger the sample size, the smaller the bias. Table 3 summarizes the results from the calculations using order statistics in the appendix. Note that $\bar{A}_{xHL}$ are unbiased if the underlying distribution is symmetric.

Table 3 : $Bias(\bar{A}_{xHL}; \mu)$ by Sample Size n

Sample Size	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
Standard Normal	0 %	0%	0%	0%	0%	0%
Exponential ($\theta = 1$)	-16%	-15%	-15%	-14%	-13%	-13%
Pareto ($\theta = 1$, $\alpha = 4$)	-6%	-6%	-6%	-6%		
LogNormal ($\mu = 0, \sigma^2 = 1$)	-23%	-23%	-22%	-21%	-20%	-20%
Weibull (θ=1, τ=0.5)	-46%	-44%	-43%	-41%	-40%	-38%

While $\bar{A}_{xHL}$ is biased with respect to the population mean for positively skewed distributions, they are more efficient than the sample average in terms of mean squared error. The efficiency advantage is consistent across the sample sizes as shown in Table 4, which summarizes the results from the appendix. Note that given a normal distribution, the sample average is universally more efficient than

$\bar{A}_{xHL}$ regardless of the sample size. For an exponential distribution, $\bar{A}_{xHL}$ is almost as efficient as the sample average. However, for the Pareto, LogNormal, and Weibull distributions, $\bar{A}_{xHL}$ are much more efficient than $\bar{X}$.

Table 4 : $REff(\bar{X}, \bar{A}_{xHL}; \mu)$ **by Sample Size** n

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
Standard Normal	88%	91%	92%	93%	94%	95%
Exponential ($\theta=1$)	97%	97%	97%	97%	97%	97%
Pareto ($\theta=1$, $\alpha=4$)	151%	148%	145%	142%		
LogNormal ($\mu=0, \sigma^2=1$)	159%	159%	159%	159%	159%	159%
Weibull (θ=1, τ=0.5)	185%	185%	185%	185%	185%	185%

3.3 Asymptotic Properties of $\bar{A}_{xHL}$

The asymptotic properties of trimmed means depend on the nature of trimming in relation to the sample size n. If the number of the trimmed observations is fixed, the trimming is considered *light*, such as $\bar{A}_{xHL}$. All other cases are considered either *intermediate* or *heavy* trimming, where the number of the trimmed observations may be infinite as n goes to infinity. For example a 25% trimmed mean is calculated by trimming 25% of the observations from both ends of the ordered sample regardless of the sample size.

Light Trimming - Note that the value of $\bar{A}_{xHL}$ approaches $\bar{X}$ as n becomes large. Kesten [8] also shows that the convergence in distribution of *lightly* trimmed means and sample average are equivalent. In other words, both $\bar{A}_{xHL}$ and the sample average are asymptotically normal with the same normalizing factors (i.e., the asymptotic mean and standard deviation). Thus the asymptotic mean of $\bar{A}_{xHL}$ is the population mean μ and it is in this sense that $\bar{A}_{xHL}$ is asymptotically unbiased. However, as shown in Section 3.2 and the appendix, depending on the type of the distribution, the finite-sample properties of $\bar{A}_{xHL}$ and the sample average can be very different.

Heavy Trimming - In the case of heavy trimming, where a fixed *percentage* of the sample points are trimmed from both ends of the ordered sample, Csörgő et al. [3] have shown that a normalized trimmed mean so defined converges in distribution to a standard normal random variable, and the asymptotic mean is the expected value of a truncated parent distribution with the upper and lower truncation points at the same fixed percentiles as in the sample. For example, if a trimmed mean is obtained by trimming 20% of the sample from both ends, the support of the truncated distribution is from the 20th percentile to the 80th percentile of the parent distribution.

Wu [12] indicates that $\bar{A}_{xHL}$ would underestimate the population mean of a positively skewed distribution. He first defines the asymptotic means of $\bar{A}_{xHL}$ ([12]

p. 717, Exhibit 1) as

$$Asym(\bar{A}_{xHL}) = \frac{1}{1-2/n} \int_{F^{-1}(\frac{1}{n})}^{F^{-1}(1-\frac{1}{n})} xf(x)dx, \tag{3}$$

which[4] is equivalent to the asymptotic mean for *heavily* trimmed means when the trimming percentage is fixed at $1/n$. For example, if the sample size is five, $Asym(\bar{A}_{xHL})$ is the expected value of a truncated parent distribution with the upper and lower truncation points at the 80th and 20th percentiles of the parent distribution, respectively. As such, $F^{-1}(1-\frac{1}{n}) = F^{-1}(0.8)$, $F^{-1}(\frac{1}{n}) = F^{-1}(0.2)$, and

$$Asym(\bar{A}_{xHL}) = \frac{1}{1-2/5} \int_{F^{-1}(0.2)}^{F^{-1}(0.8)} xf(x)dx.$$

The magnitude of the truncation is based on the size of the sample. As a result, $Asym(\bar{A}_{xHL})$ can be different when the sample size varies. Wu [12] then estimates the bias of $\bar{A}_{xHL}$ by comparing $Asym(\bar{A}_{xHL})$ with the population mean, and argues that the sampling bias can be corrected by using a ratio of the population mean and $Asym(\bar{A}_{xHL})$.

Wu's approach to the problem raises two issues. First, we know through the statistics literature (e.g., Kesten [8]), when the trimming is light, such as $\bar{A}_{xHL}$, the asymptotic mean is the same as the underlying population mean regardless of the sample size. Second, although $\bar{A}_{xHL}$ and $\bar{X}$ have the same asymptotic mean, the finite-sample expected values for $\bar{A}_{xHL}$ can be very different, depending on the sample size. The sample sizes under consideration in actuarial practice are usually quite small. Despite the fact that the exact distribution of $\bar{A}_{xHL}$ is often intractable, the means, variances, and covariances of $\bar{A}_{xHL}$ for small samples can often be derived explicitly or numerically approximated. Therefore, it is not necessary to use the asymptotic mean to calculate the theoretical bias in small samples since doing so would actually overstate the size of the bias.

Table 5 : $Bias(Asym(\bar{A}_{xHL});\mu)$ **by Sample Size** n

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
Standard Normal	0 %	0%	0%	0%	0%	0%
Exponential ($\theta=1$)	-24%	-22%	-21%	-19%	-18%	-17%
Pareto ($\theta=1$, $\alpha=4$)	-11%	-10%	-10%	-10%		
LogNormal ($\mu=0, \sigma^2=1$)	-33%	-31%	-29%	-27%	-26%	-25%
Weibull (θ=1, τ=0.5)	-64%	-60%	-57%	-54%	-52%	-50%

[4]Since we are only interested in continuous distributions, $F^{-1}(u)$ here is assumed to be uniquely determined for each u in $[0,1]$.

The appendix compares the estimated bias resulting from using $Asym(\bar{A}_{xHL})$ instead of $E(\bar{A}_{xHL})$ for the five distributions. Table 5 summarizes the results by sample size and indicates that the **overstatement** exists across the sample sizes of five to ten and can be as much as 50% for some positively skewed distributions, compared to $Bias(\bar{A}_{xHL};\mu)$ in Table 3.

One may also argue on philosophical grounds that the correction, for either small or large samples, is not necessary. That is, from a robust statistics point of view, the examination and treatment of outliers are of fundamental importance[5] while the unbiasedness with respect to the population mean is never an objective nor a concern. In general, the goal of robust location estimators is to measure the **central tendency** of the distribution, *not* the population mean. Thus the question is not whether the outliers should be eliminated or not, but how to lessen their impact if outliers exert undue influence on the estimation.

Unbiasedness seems to have been fully embraced in the casualty literature as the most important property for an estimator, but in practice unbiased estimators, such as the sample average are rarely used as selections. Instead, it is always some type of modified average depending on the circumstance, the data, and the analyst. Moving away from the "first moment only" mentality can help us achieve a shorter confidence interval and gain efficiency in terms of mean squared error, which considers both the first and second moments. Here it should be emphasized that we are not advocating abandoning the sample average as an estimator. Rather, we suggest that efficient robust estimators should always be considered along with other unbiased estimators.

4 Huber's M-Estimators

To calculate $\bar{A}_{xHL}$, automatically trimmed are the sample maximum and sample minimum, which may or may not be outliers relative to the rest of the sample. Thus it makes sense if the trimming can be limited to the outliers identified during the calculating process. Huber's M-estimator does exactly that. The theory of Huber (See Huber and Ronchetti [7]) is to solve the following problem given n i.i.d. observations $(x_1, \ldots, x_n)$:

$$\min_t \left(\sum_{i=1}^{n} \xi(x_i - t)\right)$$

[5] The usual benchmarks for robustness measurement are breakdown point and influence function (see Maronna et al. [10]).

with ξ a suitable function. Or equivalently, $\sum_{i=1}^{n} \Psi(x_i - t) = 0$ where Ψ is the derivative of ξ. Specifically, Huber's Ψ is defined as follows:

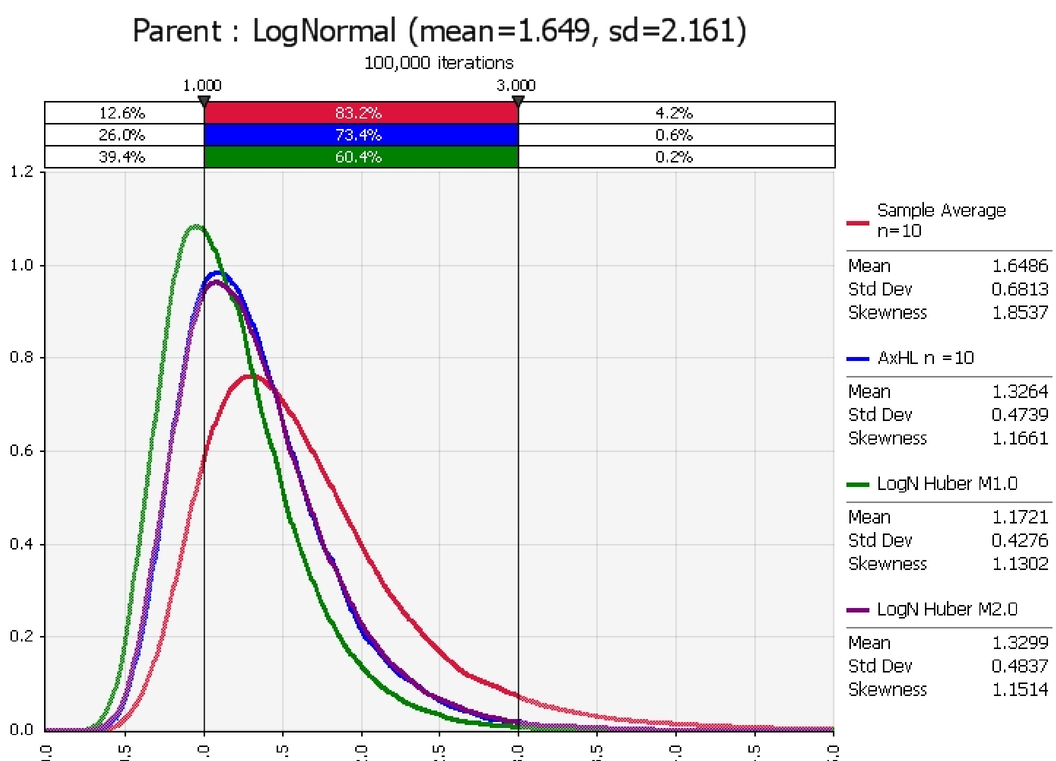
$$\Psi(x) = \begin{cases} K & \text{if } x > K, \\ x & \text{if } |x| \leq K, \\ -K & \text{if } x < -K, \end{cases}$$

where $K > 0$ is a factor selected by the analyst. In practice, the following form of Ψ is used:

$$\sum_{i=1}^{n} \Psi\left(\frac{x_i - t}{\tau}\right) = 0 \tag{4}$$

where τ is a scale measure added to ensure that the resulting solution $t = M$ is scale equivariant. The intuition here is that instead of trimming a fixed number or percentage of observations, only those observations with the adjusted values of $(x - M)/\tau$ outside the range of $[-K, K]$ are replaced by either $(M - \tau K)$ or $(M + \tau K)$. Note that the presumed outliers are not trimmed but replaced.

Graph 3 : Sampling Distributions of $\bar{X}, \bar{A}_{xHL}$, and Huber's M-Estimators (n=10)



If $K = \infty$, $\sum_{i=1}^{n}(\Psi(x_i - t)) = \sum_{i=1}^{n}(x_i - t) = 0$ and the solution to the optimization is the sample average $\bar{X}$. And if $K = 0$, the sample median is the solution. If K is between 0 and infinity, no closed-form solutions exist and a numerical approximation using the Newton-Raphson algorithm is usually employed to derive the solution. Note that when K is between 0 and infinity, the solution is not necessarily between the median and $\bar{X}$ due to the non-linearity of the problem.

The finite-sample properties of Huber's M-estimator can be obtained through simulation. Graph 3 shows the simulation results for the sampling distributions of $\bar{X}$, $\bar{A}_{xHL}$, and Huber's M-estimators with $K = 1.0$ and 2.0 from a lognormal parent distribution when the sample size is 10. The distribution of $\bar{A}_{xHL}$ is almost indistinguishable from that of Huber's M-estimators with $K = 2.0$ while $\bar{X}$ has a thicker tail and a larger standard deviation than the two robust estimators. Note that Huber's M-estimator with $K = 1.0$ has a smaller standard deviation but a larger bias than Huber's M-estimator with $K = 2.0$.

In theory, the selection of the K value is to balance between efficiency (asymptotic variance at the normal distribution) and robustness (resistance against outliers from heavy-tailed distributions). For example, compared with Huber's M-estimator with $K = 1.0$, Huber's M-estimator with $K = 2.0$ has a lower asymptotic variance at the normal distribution. Huber's M-estimator with $K = 1.0$, on the other hand, is more robust in terms of guarding against the impact of outliers.

Using the standard normal approximation may provide another perspective on what the K value implies in the calculation of Huber's M-estimator. Given 1.64 is the 95th percentile of the standard normal distribution, a range of [-1.64, 1.64] for the adjusted value $(x - t)/\tau$ may be interpreted as covering 90% of the underlying distribution.[6] With a higher K value, the range for admissible observations is getting larger and thus fewer observations are classified as outliers. If we define risk as the influence of outliers on the measure of the distribution center, then the selection of the K value may have an added benefit in reflecting the risk preference of the analyst. In other words, the more risk averse the analyst is, the lower the K value may be selected.

5 An LDF Example Using Robust Estimators

In this section, we illustrate the calculation of $\bar{A}_{xHL}$ and Huber's M-estimators using the data from Blumsohn and Laufer [2]. For completeness, the age-to-age development factors of the incurred loss triangle from Blumsohn and Laufer [2] (p. 22) are reproduced in Table 6 below along with the medians, $\bar{A}_{xHL}$, and means of the respective age-to-age factors.[7]

[6] No references can be found for this interpretation, which may be regarded as speculative.

[7] Here we assume that the age-to-age factors in each column are i.i.d.

Table 6: Age-to-Age Loss Development Factors

Year	2-1	3-2	4-3	5-4	6-5	7-6	8-7	9-8	10-9	11-10	12-11
1991	1.68	2.31	1.47	1.22	1.14	0.97	1.05	0.97	1.04	1.01	0.99
1992	6.54	1.26	1.62	1.57	0.87	1.11	1.03	1.01	0.99	0.99	
1993	1.75	2.78	1.32	1.24	1.08	1.01	0.98	0.99	1.01		
1994	3.88	1.83	0.86	0.96	1.20	1.01	1.05	1.01			
1995	2.69	1.81	0.91	1.19	1.00	1.57	1.00				
1996	1.11	1.42	1.12	1.14	1.23	1.01					
1997	1.98	1.41	0.96	1.17	1.02						
1998	3.91	1.10	1.53	1.02							
1999	1.45	0.97	1.44								
2000	1.44	1.13									
2001	1.23										
Med.	1.75	1.41	1.32	1.18	1.08	1.01	1.03	1.00	1.01	1.00	0.99
$\bar{A}_{xHL}$	2.23	1.53	1.25	1.16	1.09	1.03	1.03	1.00	1.01	1.00	0.99
Avg	2.52	1.60	1.25	1.19	1.08	1.11	1.02	0.99	1.01	1.00	0.99

Table 7: Implied Loss Reserves and M-Estimates of LDF for Various K Values

K	Prob	Implied Reserve	2-1	3-2	4-3	5-4	6-5	7-6	8-7	9-8	10-9
0.06	5%	22,017k	1.75	1.41	1.32	1.18	1.08	1.01	1.03	1.00	1.01
0.13	10%	22,089k	1.76	1.41	1.32	1.18	1.08	1.01	1.03	1.00	1.01
0.25	20%	22,382k	1.80	1.41	1.32	1.18	1.08	1.01	1.03	1.00	1.01
0.39	30%	22,741k	1.80	1.43	1.32	1.18	1.08	1.02	1.03	1.00	1.01
0.52	40%	22,673k	1.82	1.46	1.30	1.18	1.08	1.02	1.03	1.00	1.01
0.67	50%	22,854k	1.87	1.48	1.28	1.18	1.08	1.02	1.03	1.00	1.01
0.84	60%	23,415k	1.92	1.49	1.27	1.18	1.09	1.02	1.03	1.00	1.01
1.04	70%	23,503k	1.97	1.51	1.25	1.18	1.09	1.02	1.02	1.00	1.01
1.15	75%	23,650k	2.00	1.52	1.25	1.17	1.09	1.03	1.02	1.00	1.01
1.28	80%	23,758k	2.04	1.54	1.25	1.17	1.09	1.03	1.02	1.00	1.01
1.64	90%	24,227k	2.14	1.57	1.25	1.17	1.08	1.03	1.02	1.00	1.01
1.96	95%	24,908k	2.23	1.59	1.25	1.16	1.08	1.04	1.02	0.99	1.01
2.58	99%	25,799k	2.31	1.60	1.25	1.16	1.08	1.04	1.02	0.99	1.01
Med.		22,017k	1.75	1.41	1.32	1.18	1.08	1.01	1.03	1.00	1.01
$\bar{A}_{xHL}$		25,502k	2.23	1.53	1.25	1.16	1.09	1.03	1.03	1.00	1.01
Avg		33,349k	2.52	1.60	1.25	1.19	1.08	1.11	1.02	0.99	1.01

Using the data from Table 6, Table 7 shows the resulting estimated loss reserves with various K values in Huber's M-estimators. For each K value, the implied loss reserve is calculated by assuming that the same K value is used for each of the columns in the age-to-age factor selection. The K value ranges from 0.06 to 2.58, which correspond to the 5% and 99% pseudo-probability ranges, respectively (i.e., as if approximated by a standard normal distribution). Using a K value of 0.06 implies that the analyst classifies any observation x as an outlier if its adjusted value $(x - M)/\tau$ is outside the range of [-0.06, 0.06].

In the example, at the 5% pseudo-probability level, all observations are deemed outliers and the Huber's M-estimate is the sample median for all ages. With higher K values, the M-estimates change gradually from the sample median to the sample average. Finally at the 99% pseudo-probability level, only a handful of observations are deemed outliers and the Huber's M-estimate is the sample average for most of the age-to-age factors. The range of the implied reserves is between 22.0 million and 25.8 million, corresponding to K=0.06 and K=2.58, respectively.

Table 8 below shows the data points in the 2-1 age-to-age factors that are deemed outliers for various K values in the calculation of Huber's M-estimates. At K = 0.06, all points are outliers except 1.75, which happens to be the sample median. As K becomes larger, fewer data points are declared outliers. At K = 2.58, the only outlier is 6.54.

Table 8: Implied Outliers For 2-1 Age-To-Age Factors By K Value

K	Out1	Out2	Out3	Out4	Out5	Out6	Out7	Out8	Out9	Out10
0.06	6.54	3.91	3.88	2.69	1.98	1.68	1.45	1.44	1.23	1.11
0.13	6.54	3.91	3.88	2.69	1.98		1.45	1.44	1.23	1.11
0.25	6.54	3.91	3.88	2.69			1.45	1.44	1.23	1.11
0.39	6.54	3.91	3.88	2.69			1.45	1.44	1.23	1.11
0.52	6.54	3.91	3.88	2.69					1.23	1.11
0.67	6.54	3.91	3.88	2.69					1.23	1.11
0.84	6.54	3.91	3.88	2.69					1.23	1.11
1.04	6.54	3.91	3.88							1.11
1.28	6.54	3.91	3.88							
1.64	6.54	3.91	3.88							
1.96	6.54	3.91	3.88							
2.58	6.54									

A few comments on the age-to-age selection methods may be in order:

- One potential flaw or inconsistency of $\bar{A}_{xHL}$ when applied to the setting of age-to-age factor calculation is that $\bar{A}_{xHL}$ trims a different percentage of data for each of the columns. For example, for the 2-1 factor, two out of

11 observations or 18.2% of the data are trimmed while for the 8-7 factors, 40% of the data (two out of five) are trimmed. The inconsistency stems from the fact that $\bar{A}_{xHL}$ trims less percentage of the data when the data are more volatile (e.g., the 2-1 factors) and more percentage of the data when the data are relatively stable (e.g., the 8-7 factors). As indicated in Section 3.2, the finite-sample properties of $\bar{A}_{xHL}$ are dependent on the sample size and can be very different between trimming 18.2% and 40% of the data. On the other hand, using Huber's M-estimators and selecting "appropriate" K values by age may avoid this problem and maintain some level of consistency in the age-to-age factor calculation.

- The loss reserve estimates based on the M-estimates are in the middle-to-lower range of the reserves estimated by the participants in the Blumsohn and Laufer study. The primary reason is that many participants downweight or ignore the negative development in the age-to-age factor selection. For the earlier development ages, their age-to-age factor selections seem to largely fall within the range of the M-estimates with the K values between 0.06 and 2.58, except for the age 7-6 factors, where 1.566 is a prominent outlier and causes a great deal of variations in the participants' selection.

- One interesting observation regarding the age-to-age factor selections by the participants of the Blumsohn and Laufer study is that the implied K values across ages are not consistent. For example, one may select 1.75 for the 2-1 factor with an implied K value of 0.06 while selecting 1.60 for the 3-2 factor with an implied K value of 2.58 (see Table 7). This lack of consistency in terms of the K value may be due to the fact that different averaging methods were used for different ages in selections while the statistical implications of the methods are not obvious.

- When Huber's M-estimator is used with a specific K value, the confidence interval for the loss reserves can be obtained by bootstrapping individual age-to-age factors. One potential problem of this approach is that the τ value can easily become zero in equation (4) for the bootstrap samples when the sample size is small. Note that Huber's M-estimator is not well-defined when $\tau = 0$.

6 Software Implementation

Software in Excel VBA -

- Written by this author and included in the appendix are two Excel/VBA functions (*HuberM* and *MADN*) for calculating Huber's M-estimators. To implement the functions, copy the code for both functions into a Visual Basic module of the desired Excel file. The first required input for *HuberM* is a numeric range/vector while the second required input is the selected K value. Note that *HuberM* is not well-defined when τ from Equation (4) is zero. When this occurs, Excel will exhibit a warning message.

- The Royal Society of Chemistry has made available an Excel Add-in for Huber's M-estimator, *RobStat.xla*.[8] All the installation instructions are in the *ReadMe.txt* file, as well as in the full help system. The Add-in has two Excel functions, *A15_MEAN* and *H15_MEAN*, which calculate two types of Huber's M-estimators. The difference is that the former uses a fixed MADN for τ from Equation (4) in the iteration process while the latter continues to update the τ in each iteration.

  Despite the Add-in's strength in error handling and help system, this author was not able to reconcile the calculation results from *A15_MEAN* (or *H15_MEAN*) with the results from any **R**-based functions including *huberM* in the **R** package and *mest* in Wilcox's collection.

- The function *TRIMMEAN(array,α%)* supplied by Excel calculates the $\alpha\%$ trimmed mean for the array specified in the first argument of the function. For example, the 20% trimmed mean *TRIMMEAN(array,20%)* for a sample size five is the same as $\bar{A}_{xHL}$.

Software in **R** -

- Two **R** packages ("**robust**" and "**robustbase**") are available on the **R** website to calculate a variety of robust estimators. The function *huberM* in "**robustbase**" calculates Huber's M-estimator, which requires a numeric vector and a K value as inputs.

- Wilcox [11] maintains a significant collection of **R** functions in robust statistics.[9] *mest* is the function that calculates Huber's M-estimator.

- Interested readers can also find other collections of related **R** or **S-Plus** functions in *http://www.statistik.tuwien.ac.at/rsr/index.html.*

---

[8]http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software /RobustStatistics.asp

[9]http://www-rcf.usc.edu/ rwilcox/

7 Conclusion

Modern robust statistics has made it well-known that outliers can have unbounded influence on classical estimators such as the sample average, resulting in: (1) inaccurate parameter estimates/inference, (2) large standard errors, and (3) wide confidence intervals.

The purpose of this paper is to provide some theoretical facts and examples regarding average-excluding-high-and-low and more broadly, some robust estimators, which may not have been given proper credit in our literature. We have shown by example that $\bar{A}_{xHL}$ is more efficient than the sample average. It also shows that using Huber's M-estimators with selected K values may have some more benefit than using $\bar{A}_{xHL}$.

Although these two estimators only represent a tiny portion of the large number of robust estimators in the statistics literature, one of the major advantages of $\bar{A}_{xHL}$ and Huber's M-estimators is that they can be easily implemented through simple software (see Section 6). The famed Princeton Study on robust estimators (see Andrews et al. [1]) also shows that (1) some trimmed means (similar to $\bar{A}_{xHL}$) and Huber's M-estimators behave rather well under many scenarios in comparison with other robust estimators, and (2) no single robust estimator is more efficient for all distributions.

John Tukey, an early pioneer of the modern robust statistics, once said "*ust which robust/resistant methods you use is not important – what is important is that you use some.*" It is this author's belief that the use of $\bar{A}_{xHL}$ and Huber's M-estimators may be beneficial to actuaries in tackling the day-to-day selection problems.

Appendix

A.0 Basic Formulas in Order Statistics

Let $(X_1, \ldots, X_n)$ be a sample of n independent and identically distributed random variables. Denote the cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$ with mean μ and variance σ^2 (subject to existence).

Let $X_{(i)}$ be the ith order statistic of $(X_1, \ldots, X_n)$. Thus $X_{(1)} = \min(X_1, \ldots, X_n)$, $X_{(n)} = \max(X_1, \ldots, X_n)$, and $X_{(1)} \leq \ldots \leq X_{(i)} \leq \ldots \leq X_{(n)}$ for $1 \leq i \leq n$. If $F(x)$ is absolutely continuous, the expected value and the variance of $X_{(i)}$, and the expected value of $X_{(i)}$ and $X_{(j)}$ for $1 \leq i < j \leq n$ can be expressed as (see David and Nagaraja [4])

$$E(X_{(i)}) = \binom{n}{i} \int_{-\infty}^{\infty} x f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i} dx$$

$$Var(X_{(i)}) = \binom{n}{i} \int_{-\infty}^{\infty} x^2 f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i} dx - [E(X_{(i)})]^2$$

and

$$\begin{aligned} E(X_{(i)}, X_{(j)}) &= Cov(X_{(i)}, X_{(j)}) + E(X_{(i)})E(X_{(j)}) \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times \\ &\quad \int_{-\infty}^{\infty} \int_{-\infty}^{y} xyf(x)f(y)[F(x)]^{i-1}[F(y) - F(x)]^{j-i-1}[1 - F(y)]^{n-j} dxdy, \end{aligned}$$

respectively.

The closed-form solutions to $E(X_{(i)}), Var(X_{(i)})$, and $Cov(X_{(i)}, X_{(j)})$ can be derived explicitly for the exponential, Weibull, and Pareto distributions. For the lognormal distribution, numerical approximation is needed to calculate these statistics. In the order statistics literature, extensive studies (see David and Nagraja [4]) were performed in the 1950s and 1960s on the calculations of the moments of order statistics for various distributions by sample size. Harter and Balakrishnan [6] have summarized and tabulated the numerical results of those studies in their 1996 Handbook.

In this section, we calculate and tabulate the numerical values of the means and variances of $\bar{A}_{xHL}$ for the standard normal distribution and four distributions with nonnegative supports, namely the exponential, lognormal, Pareto and Weibull distributions.[10] In other words, We assume the underlying distribution is known and there is no model misspecification or data contamination. We then employ these results to calculate the exact values of the bias and the relative efficiency of the sample average and $\bar{A}_{xHL}$ for the sample sizes between five and ten. Although $\bar{A}_{xHL}$ may be significantly downward biased for a positively skewed distribution, $MSE(\bar{A}_{xHL})$ usually is smaller than $MSE(\bar{X})$, which is just $Var(\bar{X}) = \sigma^2/n$. That is, even considering the penalty for bias, the average distance as defined by MSE between $\bar{A}_{xHL}$ and μ may still be shorter than that between $\bar{X}$ and μ.

A.1 The Standard Normal Distribution

Since the standard normal is symmetric, $\bar{A}_{xHL}$ is unbiased. Note that $\bar{X}$ is more efficient than $\bar{A}_{xHL}$ as $Var(\bar{X}) \leq Var(\bar{A}_{xHL})$ for all sample sizes. In fact, the standard normal distribution has the rare property that $\bar{X}$ is more efficient than most robust location estimators.

Table A.1 : Efficiency of $\bar{A}_{xHL}$ for Standard Normal

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
$E(\bar{X})$	0	0	0	0	0	0
$E(\bar{A}_{xHL})$	0	0	0	0	0	0
$Bias(\bar{A}_{xHL};\mu)$	0%	0%	0%	0%	0%	0%
$Asym(\bar{A}_{xHL})$	0	0	0	0	0	0
$Bias(Asym(\bar{A}_{xHL});\mu)$	0%	0%	0%	0%	0%	0%
$Var(\bar{X})$	0.20000	0.16667	0.14286	0.12500	0.11111	0.10000
$Var(\bar{A}_{xHL})$	0.22706	0.18403	0.15494	0.13387	0.11790	0.10535
$MSE(\bar{A}_{xHL};\mu)$	0.22706	0.18403	0.15494	0.13387	0.11790	0.10535
$REff(\bar{X},\bar{A}_{xHL};\mu)$	88%	91%	92%	93%	94%	95%

A.2 The Exponential Distribution

The pdf and cdf of the exponential distribution with scale parameter θ are

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}, \qquad x \geq 0, \theta > 0,$$

$$F(x;\theta) = 1 - e^{-x/\theta},$$

respectively. Given that the rth moment of X is $E(x^r) = \theta^r r!$, the exponential distribution has a fixed skewness of 2, independent of θ as shown below

$$Skewness(x) = \frac{E(x^3) - 3\theta E(x^2) + 2\theta^3}{\theta^3} = 2.$$

[10]The distribution forms and the corresponding statistics are based on Klugman et al. [9].

The closed-form solutions exist for the mean and variance of $X_{(i)}$, which are

$$E(X_{(i)};\theta) = \theta \sum_{j=1}^{i} \frac{1}{n-j+1},$$

and

$$Var(X_{(i)};\theta) = \theta^2 \sum_{j=1}^{i} \frac{1}{(n-j+1)^2},$$

respectively. For $i < j$, the covariance of $X_{(i)}$ and $X_{(j)}$ is the same as $Var(X_{(i)};\theta)$. For a sample of five,

$$Bias(\bar{A}_{xHL}) = \frac{5\theta - \theta(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}) - \theta(\frac{1}{5})}{(5-2)\theta} - 1 = -16.1\%.$$

Given the fixed skewness of the exponential distribution, it is not surprising that $Bias(\bar{A}_{xHL})$ is dependent on the sample size n only and independent of the parameter θ.

The calculation of $Asym(\bar{A}_{xHL})$ depends on the sample size n and θ.

$$Asym(\bar{A}_{xHL}) = \frac{\theta}{1-2/n}\left\{\Gamma(2;-ln(\frac{1}{n})) - \Gamma(2;-ln(1-\frac{1}{n}))\right\}$$

where $\Gamma(.,.)$ is the incomplete Gamma function. The following table shows the statistics for the exponential distribution with $\theta = 1$. Note that $E(x) = 1$ and $V(x) = 1$ when $\theta = 1$. As expected, when the sample size gets larger the bias is getting smaller. On the other hand, $\bar{A}_{xHL}$ is almost as efficient as $\bar{X}$ for sample sizes 5 to 10.

Table A.2 : Bias and Efficiency of $\bar{A}_{xHL}$ for Exponential (θ=1)

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
$E(\bar{X})$	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
$E(\bar{A}_{xHL})$	0.83889	0.84584	0.85286	0.85952	0.86570	0.87138
$Bias(\bar{A}_{xHL};\mu)$	-16.1%	-15.4%	-14.7%	-14.0%	-13.4%	-12.9%
$Asym(\bar{A}_{xHL})$	0.76085	0.77970	0.79547	0.80914	0.82076	0.83058
$Bias(Asym(\bar{A}_{xHL});\mu)$	-23.9%	-22.0%	-20.5%	-19.1%	-17.9%	-16.9%
$Var(\bar{X})$	0.20000	0.16667	0.14286	0.12500	0.11111	0.10000
$Var(\bar{A}_{xHL})$	0.17966	0.14634	0.12407	0.10801	0.09585	0.08628
$MSE(\bar{A}_{xHL};\mu)$	0.20562	0.17011	0.14572	0.12775	0.11389	0.10282
$REff(\bar{X},\bar{A}_{xHL};\mu)$	97%	97%	97%	97%	97%	97%

A.3 The Pareto Distribution

The pdf and cdf of the Pareto distribution with scale parameter θ and shape parameter α are

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \qquad F(x) = 1 - (\frac{\theta}{x})^{\alpha}, \qquad x \geq \theta, \alpha > 0,$$

respectively. The mean, variance, and skewness are

$$E(x) = \frac{\alpha\theta}{\alpha - 1}, \qquad Var(x) = \frac{\theta^2\alpha}{(\alpha-2)(\alpha-1)^2},$$

and

$$Skewness(x) = \frac{2(1+\alpha)}{\alpha - 3}\sqrt{\frac{\alpha-2}{\alpha}},$$

respectively. The calculation of $Asym(\bar{A}_{xHL})$ depends on the sample size n,

$$Asym(\bar{A}_{xHL}) = \frac{1}{(1-2/n)}\frac{\alpha\theta}{(\alpha-1)}\left\{(1-\frac{1}{n})^{1-1/\alpha} - (\frac{1}{n})^{1-1/\alpha}\right\}.$$

Using Tables C13.1 and C13.2 in Harter and Balakrishnan [6] and Eqs. (1)-(2) in Section 3.1, the means and variances of $\bar{A}_{xHL}$ with the underlying Pareto ($\alpha = 4, \theta = 1$) are shown in the following table. Note that $E(x) = 1.33333$, $Var(x) = 0.22225$, and $Skewness(x) = 7.07106$ for the Pareto distribution with $\alpha = 4$ and $\theta = 1$.

Table A.3 : Bias and Efficiency of $\bar{A}_{xHL}$ for Pareto (θ=1, α=4)

Sample Size	$n = 5$	$n = 6$	$n = 7$	$n = 8$
$E(\bar{X})$	1.33334	1.33332	1.33333	1.33333
$E(\bar{A}_{xHL})$	1.24920	1.25220	1.25530	1.25823
$Bias(\bar{A}_{xHL}; \mu)$	-6.3%	-6.1%	-5.9%	-5.6%
$Asym(\bar{A}_{xHL})$	1.18859	1.19496	1.20037	1.20502
$Bias(Asym(\bar{A}_{xHL}); \mu)$	-10.9%	-10.4%	-10.0%	-9.6%
$Var(\bar{X})$	0.04445	0.03703	0.03174	0.02777
$Var(\bar{A}_{xHL})$	0.02234	0.01839	0.01578	0.01391
$MSE(\bar{A}_{xHL}; \mu)$	0.02942	0.02497	0.02187	0.01954
$REff(\bar{X}, \bar{A}_{xHL}; \mu)$	151%	148%	145%	142%

A.4 The LogNormal Distribution

The pdf of the standard lognormal distribution with location parameter[11] μ and shape parameter σ^2 is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(ln(x)-\mu)^2/2\sigma^2} \qquad\qquad 0 < x < \infty.$$

The mean, variance, and skewness are

$$E(x) = e^{\mu+\sigma^2/2}, \qquad Var(x) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2},$$

and

$$Skewness(x) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1},$$

respectively.

No closed-form solutions exist for the cdf. So numerical approximation has to be performed for the means and variance of $\bar{X}$, $\bar{A}_{xHL}$. The calculation of $Asym(\bar{A}_{xHL})$ depends on the sample size n, μ, and σ^2.

$$Asym(\bar{A}_{xHL}) = \frac{e^{\mu+\sigma^2/2}}{1-2/n} \left\{ \theta(\theta^{-1}(1 - \frac{1}{n}) - \sigma) - \theta(\theta^{-1}(\frac{1}{n}) - \sigma) \right\},$$

where $\theta()$ is the cdf of the standard normal distribution. Using Tables C6.1 and C6.2 in Harter and Balakrishnan [6] and Eqs. (1)-(2) in Section 3.1, the means and variances of $\bar{A}_{xHL}$ with the underlying lognormal ($\mu = 0, \sigma = 1$) are shown in the following table. Note that $E(x) = 1.64872$, $Var(x) = 4.67075$, and $Skewness(x) = 6.1849$ for the lognormal distribution with μ=0 and σ^2=1.

Table A.4 : Bias and Efficiency of $\bar{A}_{xHL}$ for LogNormal (μ=0, σ^2=1)

Sample Size	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$E(\bar{X})$	1.64872	1.64872	1.64873	1.64872	1.64872	1.64872
$E(\bar{A}_{xHL})$	1.26269	1.27623	1.29000	1.30314	1.31542	1.32679
$Bias(\bar{A}_{xHL}; \mu)$	-23.4%	-22.6%	-21.8%	-21.0%	-20.2%	-19.5%
$Asym(\bar{A}_{xHL})$	1.11100	1.14365	1.17164	1.19585	1.21702	1.23571
$Bias(Asym(\bar{A}_{xHL}); \mu)$	-32.6%	-30.6%	-28.9%	-27.5%	-26.2%	-25.1%
$Var(\bar{X})$	0.93415	0.77846	0.66725	0.58385	0.51898	0.46708
$Var(\bar{A}_{xHL})$	0.43857	0.36178	0.31139	0.27534	0.24803	0.22646
$MSE(\bar{A}_{xHL}; \mu)$	0.58759	0.50053	0.44008	0.39477	0.35912	0.33011
$REff(\bar{X}, \bar{A}_{xHL}; \mu)$	159%	159%	159%	159%	159%	159%

[11] This is not the same μ as in $Bias(\bar{A}_{xHL}; \mu)$

A.5 The Weibull Distribution

The pdf and cdf of the two-parameter Weibull distribution with scale parameter θ and shape parameter τ are

$$f(x;\theta,\tau) = \frac{\tau}{\theta}\left\{\frac{x}{\theta}\right\}^{\tau-1} e^{-(x/\theta)^{\tau}} \qquad x \geq 0, \theta > 0, \tau > 0$$

and

$$F(x;\theta,\tau) = 1 - e^{-(x/\theta)^{\tau}},$$

respectively. The mean, variance, and Skewness are

$$E(x) = \theta\Gamma(1+\frac{1}{\tau}), \qquad Var(x) = \theta^2\Gamma(1+\frac{2}{\tau}) - (\theta\Gamma(1+\frac{1}{\tau}))^2,$$

and

$$Skewness(x) = \frac{\Gamma(1+\frac{3}{\tau}) - 3\Gamma(1+\frac{2}{\tau})\Gamma(1+\frac{1}{\tau}) + 2[\Gamma(1+\frac{1}{\tau})]^3}{[\Gamma(1+\frac{2}{\tau}) - [\Gamma(1+\frac{1}{\tau})]^2]^{3/2}},$$

respectively.

Various closed-form solutions exist for the means and variances for $X_{(i)}$ (see Harter and Balakrishnan [6]). The calculation of $Asym(\bar{A}_{xHL})$ depends on the sample size n and parameters θ and τ,

$$\begin{aligned} Asym(\bar{A}_{xHL}) &= \frac{\theta\Gamma(1+\frac{1}{\tau})}{1-2/n}\left\{\Gamma(1+\frac{1}{\tau};[\frac{F^{-1}(1-\frac{1}{n})}{\theta}]^{\tau}) - \Gamma(1+\frac{1}{\tau};[\frac{F^{-1}(\frac{1}{n})}{\theta}]^{\tau})\right\} \\ &= \frac{\theta\Gamma(1+\frac{1}{\tau})}{1-2/n}\left\{\Gamma(1+\frac{1}{\tau};-ln(\frac{1}{n})) - \Gamma(1+\frac{1}{\tau};-ln(1-\frac{1}{n}))\right\}. \end{aligned}$$

Note that $E(x) = 2$, $Var(x) = 20$, and $Skewness(x) = 6.618$ for the Weibull distribution with $\tau = 0.5$ and $\theta = 1$. Using Tables C3.1 and C3.2 in Harter and Balakrishnan [6] and Eqs. (1)-(2) in Section 3.1, the means and variances of $\bar{A}_{xHL}$ with the underlying Weibull distribution (θ=1,τ=0.5) are shown in the following table.

Table A.5 : Bias and Efficiency of $\bar{A}_{xHL}$ for Weibull (θ=1, τ=0.5)

Sample Size	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
$E(\bar{X})$	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000
$E(\bar{A}_{xHL})$	1.08093	1.11264	1.14489	1.17576	1.20464	1.23142
$Bias(\bar{A}_{xHL};\mu)$	-46.0%	-44.4%	-42.8%	-41.2%	-39.8%	-38.4%
$Asym(\bar{A}_{xHL})$	0.72468	0.79828	0.86211	0.91824	0.96747	1.01078
$Bias(Asym(\bar{A}_{xHL});\mu)$	-63.8%	-60.1%	-56.9%	-54.1%	-51.6%	-49.5%
$Var(\bar{X})$	4.00000	3.33333	2.85714	2.50000	2.22222	2.00000
$Var(\bar{A}_{xHL})$	1.31714	1.09927	0.96081	0.86327	0.78969	0.73149
$MSE(\bar{A}_{xHL};\mu)$	2.16184	1.88668	1.69202	1.54263	1.42229	1.32221
$REff(\bar{X},\bar{A}_{xHL};\mu)$	185%	185%	185%	185%	185%	185%

A.6 Excel VBA Functions for Huber's M-Estimator

```
Function HuberM(x As Range, KValue As Double) As Double
 Dim vRange1 As Variant
 Dim dTemp, dHuberSum, dTempHuberM, dMADN As Double
 Dim h, i, j, iRowCount, iColumnCount, iHuberCount As Integer
 iRowCount = x.Rows.Count
 iColumnCount = x.Columns.Count
 vRange1 = x.Cells.Value
 dMADN = MADN(x)
 dTempHuberM = WorksheetFunction.Median(x)
 dTemp = 0
 For h = 1 To 20     '20 is arbitrary
    dHuberSum = 0
    iHuberCount = 0
    For i = 1 To iRowCount
       For j = 1 To iColumnCount
            dTemp = (vRange1(i, j) - dTempHuberM) / dMADN
            If Abs(dTemp) < KValue Then
              dHuberSum = dHuberSum + dTemp
              iHuberCount = iHuberCount + 1
            ElseIf dTemp > KValue Then
              dHuberSum = dHuberSum + KValue
            Else
               dHuberSum = dHuberSum - KValue
            End If
       Next j
    Next i
    If iHuberCount = 0 Then
        dTemp = dTempHuberM
    Else
        dTemp = dTempHuberM + dMADN * dHuberSum / iHuberCount
    End If
    If Abs(dTemp - dTempHuberM) < 0.0001 Then
        HuberM = dTemp
        Exit Function
    Else
        dTempHuberM = dTemp
    End If
Next h
End Function
```

```
Function MADN(x As Range) As Double
 Dim vRange1 As Variant
 Dim dMedian As Double
 Dim i, j, iRowCount, iColumnCount As Integer
 iRowCount = x.Rows.Count
 iColumnCount = x.Columns.Count
 vRange1 = x.Cells.Value
 dMedian = WorksheetFunction.Median(x)
 For i = 1 To iRowCount
      For j = 1 To iColumnCount
        vRange1(i, j) = Abs(vRange1(i, j) - dMedian)
      Next j
 Next i
 MADN = WorksheetFunction.Median(vRange1) / 0.6745
End Function
```

References

[1] Andrews, D., et al., *Robust Estimates of Location*, (Princeton, NJ: Princeton University Press, 1972).

[2] Blumsohn, G. and M. Laufer, "Unstable Loss Development Factors," Casualty Actuarial Society *E-Forum*, Spring 2009, pp. 1-38.

[3] Csörgö, S., E. Haeusler, and D.M. Mason, "A Probabilistic Approach to The Asymptotic Distribution of Independent, Identically Distributed Random Variables," *Advances In Applied Mathematics* 9, 1988, pp. 259-333.

[4] David, H A., and H.N. Nagaraja, *Order Statistics*, 3rd ed., (Hoboken, NJ: John Wiley & Sons, 2003).

[5] Fleming, K., "Yep, We're Skewed," *Variance* 2:2, pp. 179-183, 2009.

[6] Harter, A. and N. Balakrishnan, *CRC Handbook of Tables for the Use of Order Statistics in Estimation* (Boca Raton, FL: CRC Press, 1996).

[7] Huber, P., and E. Ronchetti, *Robust Statistics*, 2nd ed. (Hoboken, NJ: John Wiley & Sons, 2009).

[8] Kesten, H., "Convergence in Distribution of Lightly Trimmed and Untrimmed Sums are Equivalent," *Mathematical Proceedings* 113, Cambridge Philosophical Society, 1993, pp. 615-638.

[9] Klugman, S., H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, 3rd ed. (Hoboken, NJ: John Wiley & Sons, 2008).

[10] Maronna, R., R. Martin, and V. Yohai, *Robust Statistics: Theory and Methods* (Hoboken, NJ: John Wiley & Sons, 2006).

[11] Wilcox, R., *Introduction to Robust Estimation and Hypothesis Testing* (Burlington, MA: Academic Press, 2004).

[12] Wu, C.P., "Downward Bias of Using High-Low Averages for Loss Development Factors," *PCAS* LXXXVI, 1999, pp. 699-735.