Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

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

Hou-wen Jeng

September 20, 2010

This paper explores the use of robust location estimators in loss reserving

Frequently Encountered Problems in Loss Reserving:
- Small sample sizes (usually no more than 10 data points);
- Outliers (the criteria for identifying outliers and how to deal with outliers once identified)

Probability of at least one outlier from the right tail in a sample size n

- **Outlier defined as >= 95th percentile**

n=5	n=6	n=7	n=8	n=9	n=10
22.6%	26.5%	30.2%	33.7%	37.0%	40.1%

- **Outlier defined as >= 90th percentile**

n=5	n=6	n=7	n=8	n=9	n=10
41.0%	46.9%	52.2%	57.0%	61.3%	65.1%

Methodology vs. Practice

Current methodologies seem to
- focus mainly on estimating the population means of the underlying distributions;
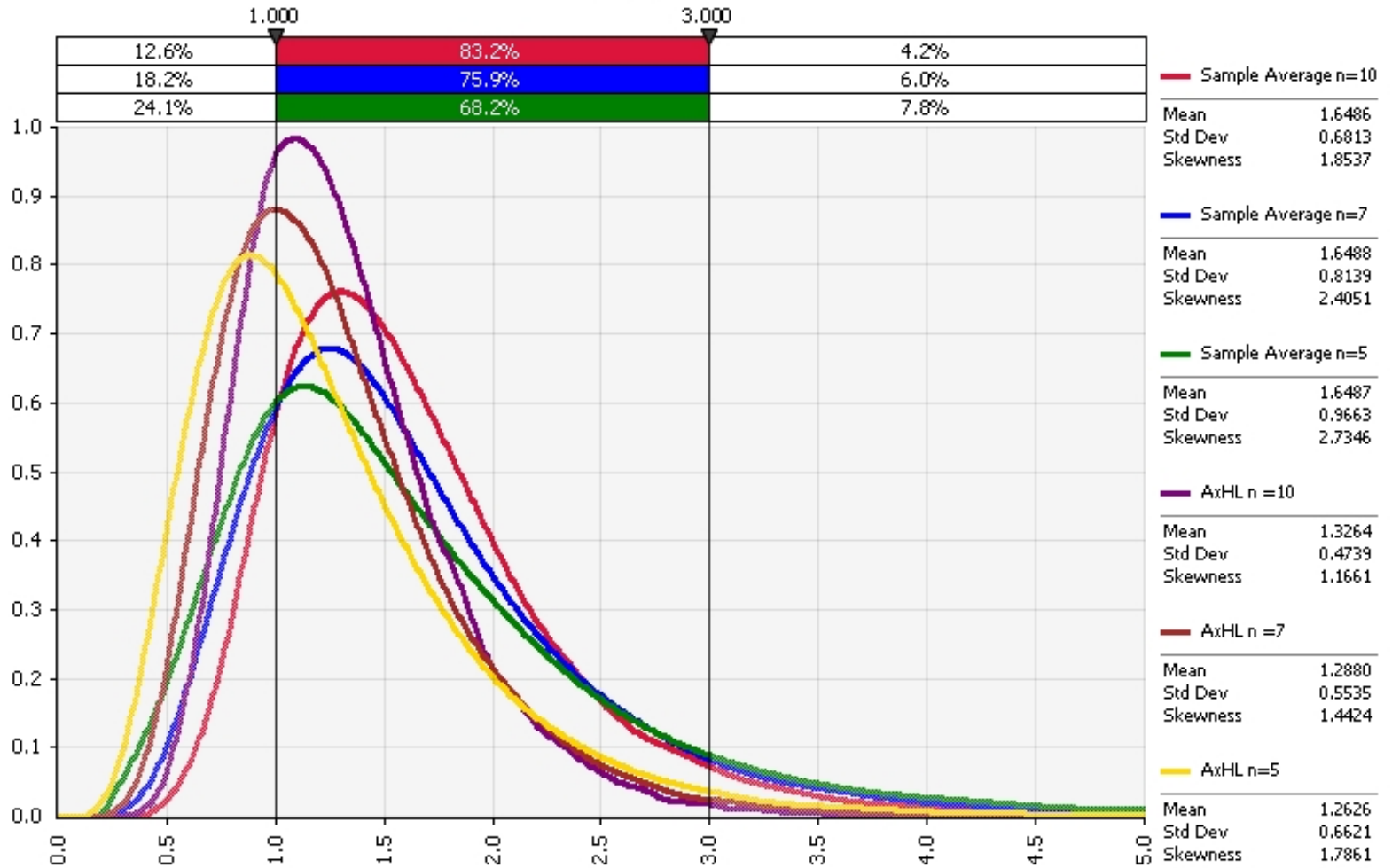- prefer unbiased estimators.

Current practice
- Sample averages are rarely used. Instead, a variety of averaging methods are used.
- relies heavily on judgment;
- lacks consistency.

Average-Excluding-High-and-Low

- A type of trimmed means (robust estimators)
- Biased with respect to the mean for a positively skewed distribution;
- More efficient than the sample mean;
- Usually has lower mean squared error than the sample mean
- Finite sample statistics can be derived;
- Has the same asymptotic mean as the sample mean.

Parent : LogNormal (mean=1.649, sd=2.161)

100,000 iterations

	1.000		3.000	
12.6%		83.2%		4.2%
18.2%		75.9%		6.0%
24.1%		68.2%		7.8%

── Sample Average n=10	
Mean	1.6486
Std Dev	0.6813
Skewness	1.8537

── Sample Average n=7	
Mean	1.6488
Std Dev	0.8139
Skewness	2.4051

── Sample Average n=5	
Mean	1.6487
Std Dev	0.9663
Skewness	2.7346

── AxHL n =10	
Mean	1.3264
Std Dev	0.4739
Skewness	1.1661

── AxHL n =7	
Mean	1.2880
Std Dev	0.5535
Skewness	1.4424

── AxHL n=5	
Mean	1.2626
Std Dev	0.6621
Skewness	1.7861

Robust estimators in general

- Seek to find the "central tendency" of the distribution;
- Efficiency is utmost important while bias is usually not a concern;
- Mean squared error may be a good compromise in selecting the desired estimator;
- The robustness of Average-Excluding-High-and-Low is limited but it may be good enough for small samples.

One robust estimator in particular

- Huber's M-Estimators identify outliers using a mini-max statistical procedure, which accommodates outliers, rather than reject them outright.
- Behave well under many scenarios;
- Relatively easy to calculate;
- More importantly, the modeler needs to exhibit his/her degree of "risk tolerance" by selecting the K factor.
- The selection of the K factor can also enforce consistency.

"Just which robust methods you use is not important – what is important is that you use some." **John Tukey**

- We are not advocating abandoning the sample average as an estimator;
- We only suggest that efficient robust estimators should always be considered along with other unbiased estimators.