How Much is Enough? An Empirical Testing of the Relationship between the Variability of Reserve Estimates and the Volume of Data

by Mark W. Littman, FCAS

How Much is Enough?

An Empirical Testing of the Relationship between the Variability of Reserve Estimates and the Volume of Data

ABSTRACT

Actuaries deal with data sets for reserving analysis that can vary substantially in terms of the volume of underlying data. The volume of data is one of several factors that affect the degree of certainty that can be attached to reserve estimates. affecting, in turn, the breadth of a range of reserve estimates that can be considered reasonable. We believe that the "performance" of a reserving method is evaluated, in part, by the variability of the estimates derived from the application of the method to a particular set of data. Variability is partially dependent on the volume of data utilized. We propose that an inverse relationship is present between volume of data and variability.

This paper describes our testing of this hypothesis by quantitative analysis of empirical data. We provide insights, if not absolute answers, regarding:

- how much effect the volume of data has on the variability inherent in reserve estimates in a loss development context, and
- how much data may be required to achieve a certain tolerance level in reserve estimates.

How Much is Enough?

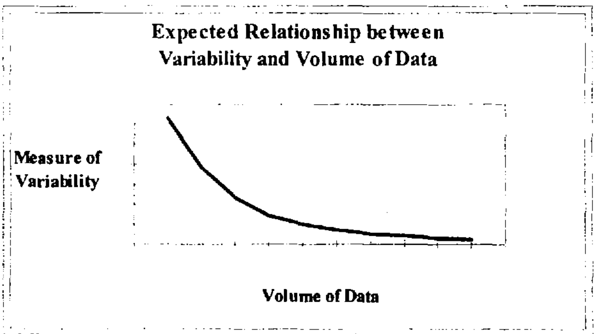
An Empirical Testing of the Relationship between the Variability of Reserve Estimates and the Volume of Data

INTRODUCTION

In the context of the loss reserving process, it is commonly held that the volume of data is one of the major factors influencing the degree of certainty that can be attached to point-estimates of the reserves. For example, if the actuary is presented with an extensive history of claims development data for a large volume of business, the reserve estimates will generally be considered to be fairly reliable. The actuary may consider a range of reasonable estimates to be plus or minus 5% around the point estimate.

If the actuary, on the other hand, is presented with the same "size" development triangle, but where the values therein are much smaller (say, from a small company writing only a fraction of the business of the large company), the degree of certainty attached to the actuarial estimate is likely to be diminished. For this case, the actuary may deem a range of reasonable estimates to be 10% or more around the point-estimate.

To illustrate our point, we would expect an inverse relationship between volume of data and the variability (uncertainty) associated with actuarial estimates of reserves.



The objective of this paper is to attempt to measure the expected inverse relationship between a measure of variability and the volume of underlying data.

BACKGROUND

Several actuarial publications contain discussions of the issues of credibility and reliability of actuarial estimates.

CAS Statement of Principles

The CAS Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves (the Statement) indicates that:

> the uncertainty inherent in the estimation of required provisions for unpaid losses or loss adjustment expenses implies that a range of estimates can be actuarially sound. The true value of the liability for losses or loss adjustment expenses at any accounting date can be known only when all attendant claims have been settled.

The Statement also discusses credibility as one of many items for the actuary to consider when evaluating the loss and loss adjustment expense reserves of an entity. The Statement says that (***bold italics*** added for emphasis):

> A group of claims should be ***large enough*** to be ***statistically reliable***. Obtaining homogeneous groupings requires refinement and partitioning of the total data. There is a point at which partitioning divides data into groups ***too small*** to provide ***credible*** development patterns. Each situation requires a balancing of the homogeneity and amount of data in each grouping. Thus, line and coverage definitions suitable for the establishment of reserves for ***large*** insurers can be in much finer detail than in the case of ***small*** insurers. Where a ***very small*** group of claims is involved, use of external information such as industry aggregates may be necessary.

The Statement did not, however, define the highlighted terms, such as "large enough," "statistically reliable," "too small," "credible," and "very small."

ASB Standard of Practice on Credibility

The Actuarial Standards Board has promulgated Standard of Practice No. 25, Credibility Procedures Applicable to Accident and Health, Group Term Life, and Property/Casualty Coverages (October 1996) (the Standard). Although the Standard appears to be mainly focused on ratemaking and prospective experience rating, the Standard indicates that it is applicable "whenever else credibility procedures are used, including but not limited to reserve analysis, solvency testing, and asset/liability management."

The Standard defines full credibility as the "level at which the subject experience is assigned full predictive value based on a selected confidence interval."

Further, the Standard describes the purpose of credibility procedures as:

- to blend information from subject experience with information from one or more sets of related experience when the subject experience does not have full credibility in order to improve the estimates of expected values, or

- to determine when the subject experience should have full credibility and blending is unnecessary.

These excerpts from the Statement and the Standard regarding credibility and reliability tend to be expressed in qualitative terms, rather than quantitative terms.

Objective of our Testing

How does a reserving actuary determine when the data in a loss development triangle are not statistically reliable for estimating development patterns and therefore, not reliable for estimating ultimates and the unpaid liabilities? In other words, how much data is enough?

An empirical testing of the relationship between volume of data and uncertainty (as approximated by the selected measure discussed below), we believe, will at least partially address these questions and raise the level of dialogue that, up to now, has involved mostly professional judgment.

We acknowledge that the volume of data available for analysis is just one of many factors affecting the variability of the reserve estimates. Other factors include changes in claim practices, external environment, the nature of the coverage afforded, the type of policy and its terms and conditions, growth in the underlying book of business, and changes in the composition of the book of business.

Therefore, while we do not expect our testing to yield "perfect regression statistics" (for instance, R^2-statistics of 1.000), we do expect the relationship to be confirmed as statistically significant. In this way, our testing will begin to give clues for "how much is enough?" and an approach for introducing other factors.

SELECTED MEASURES OF VOLUME AND VARIABILITY

Measure of Volume

Credibility standards are generally expressed in terms of an empirical measure of volume required in order to achieve full credibility for the specified purpose. We considered the relative merits of using each of the following as the measure of volume:

⇒ aggregate earned premiums for the experience period,
⇒ aggregate paid losses for the experience period, and
⇒ aggregate reported losses for the experience period.

We elected to use earned premiums as the measure of volume, as this would serve as a proxy for underlying exposures, and would not be distorted by the relative frequency/severity attributes of the coverages. (Claim count statistics from the publicly available data sources were not considered reliable.)

Variability Measure - Absolute Basis

We recognize that no single measure can be considered to be "best" for evaluating variability or uncertainty. We believe that the performance of a reserving method is, in part, evaluated by the variability of the estimates that can reasonably be generated by the application of the method to a particular data set.

The loss development method (also known as the chain ladder method) is the most widely used method for estimating ultimate claim costs, and the corresponding unpaid claims liabilities. Dr. Thomas Mack (Munich Re, Germany) wrote a paper describing an algorithm for explicitly evaluating the variability of ultimate loss (and reserve) estimates based on the chain ladder model. His paper was titled, "Measuring the Variability of Chain Ladder Reserve Estimates," and was submitted in 1993 for the CAS call program on Variability of Loss Reserves.

Mack described an algorithm that provides for a direct calculation of the estimated standard error (deviation) of the ultimate loss estimate. This was shown to be the same as the estimated standard error of the loss reserve estimate, because the difference (cumulative paid losses) was a known, constant amount. The algorithm generates an estimate of the standard error for each loss period (for example, accident year) as well as the standard error for all loss periods combined.

The estimated standard error represents an absolute measure of variability.

In Appendix I, we provide two numerical examples to give the reader a basic familiarity with how the Mack method works, with the complex mathematical formulae shielded from the reader. Reviewing our examples is not a substitute for a thorough reading of the original paper, which we encourage. In Appendix II, we discuss some of the underlying criteria and conditions that underlie Mack's model.

Variability Measure - Relative Basis

We have chosen the estimated standard error as the measure of variability by which to evaluate the performance of reserving methods. But we can reasonably expect that the absolute value of the estimated standard error will be positively correlated with the volume of the underlying data. For example, assume the following:

Company	Earned Premium	Estimated Reserves	Estimated Standard Error of Loss Estimate: Absolute	Estimated Standard Error of Loss Estimate: Relative (to Reserves)
A	1,000	400	60	15%
B	10,000	4,000	400	10%

The estimated reserves for both companies represent 40% of the experience period earned premium. The absolute size of the estimated standard error of the loss estimate for Company B is larger than that for Company A, due to its larger size (based on earned premium). But, we assert that, with the effect of the law of large numbers coming into play, the standard error expressed as a percentage to estimated reserves for Company B (10%) would be less than that for Company A (15%).

We use the relative estimated standard error in our testing, in order to evaluate the expected inverse relationship between variability and volume. The above illustration sets the stage for our empirical testing, by highlighting the dependent and independent variables:

- The volume of data (the independent variable) is the experience period earned premium
- The measure of variability (the dependent variable) is the estimated standard error of the loss estimate for the experience period (all accident years combined), expressed as a percentage of the estimated reserves

DATA AND METHOD FOR TESTING

Data for Empirical Testing

The database used for our empirical testing was populated with data from Schedule P of the statutory-basis Annual Statement for about 125 companies as of December 31, 1995. The data were for accident years 1986 to 1995 at annual valuations and included:

- Net earned premiums (Part 1_)
- Net incurred (ultimate) loss and ALAE development (Part 2_)
- Net paid loss and ALAE development (Part 3_)
- Net loss and ALAE IBNR provisions (Part 4_)

The data were gathered for the following lines of business:

- Commercial Auto Liability
- Commercial Multiple-Peril
- Homeowners
- Private Passenger Auto Liability

The data were obtained from the OneSource CD that contains Schedule P data for virtually all companies and groups filing the statutory annual statement in the U.S. For the purpose of this paper, we use the term "losses" to refer to losses and ALAE. Net reported loss triangles were determined by subtracting the IBNR triangles from the incurred triangles.

(Our evaluation of the claim count data reported in Schedule P, especially at the industry aggregate level, quickly raised doubts over the integrity of these data across the industry, and therefore, we did not gather the claim count statistics. Claim counts may have served as a preferred measure of volume over premium.)

We attempted to gather a broad cross-section of companies in terms of size and market. We generally retained the data for company groups, rather than for the individual subsidiary companies. The data for these companies represented the following proportions of the total industry, as reported by A.M. Best in the 1995 Edition of Best's Aggregates & Averages:

Annual Statement Line	% Earned Premium for 1995
Commercial Auto Liability	70%
Commercial Multi-Peril	73%
Homeowners	86%
Private Passenger Auto Liability	88%

Method of Deriving Actuarial Estimates

For each company and line of business, point-estimates of loss reserves were based on the loss development method applied to the paid and reported loss development histories discussed above. The age-to-age development factors were based on the volume-weighted average of all years' historical age-to-age development factors. The reported loss development tail factor, from 120 months to ultimate, was set at 1.000 for the purpose of our testing. The paid loss tail factor was set at the ratio of the reported losses to paid losses at the 120 month valuation. In this way, the estimate of ultimate losses derived from paid loss data for the 1986 accident year was equal to the estimate based on reported data, lest we create differences attributable to tail factors.

Using the Corporate edition of the Affinity actuarial software, we created an object-oriented database that served as the repository of the large volume of input and output items. A "2-tab" spreadsheet file served as the template for the actuarial calculations. The data items for each company that were saved from the spreadsheet to the database were:

- estimated total reserves
- estimated ultimate losses
- estimated standard error

Two values for each item were stored based on the analysis of paid and reported data.

In some cases, either the paid or reported loss development triangle (or both) contained one or more data points that appeared to cause undue distortions on the evaluation of the estimated reserves and standard errors. In these cases, we deleted or adjusted the suspicious data points, and utilized the estimates derived from the adjusted data in our final results.

Fitting a Curve to Describe the Relationship

As we sought to establish a quantitative relationship among estimated standard errors (expressed as percentages to the estimates of total reserves) to the earned premium measure of volume, we fit the empirical data to the general form of the Weibull curve. This form could permit a flexible, non-linear, negative correlation among the independent and dependent variables.

From the original sample of about 125 companies, there were some companies for certain lines of business for which:

◊ the development history did not extend 10 years.

◊ the estimated relative standard error was negative (due to a negative estimate for the total reserve)

◊ the historical data were just too sparse.

The observations from these companies' data were excluded from the curve fitting phase of the testing.

SUMMARY OF RESULTS

Statistical Significance of the Relationship

Two statistics that evaluate the significance of a relationship between the dependent and independent variables are the R^2 and F- statistics. The R^2 statistic, in simple terms, measures the proportion of variance of the data points from their overall mean that is 'explained' by the curve. The F-statistic, in general terms, evaluates whether the collection of independent variable(s) are relevant in the statistical model. A summary of the R^2 and F- statistics based on our testing is shown in Table 1:

Table 1
Statistical Significance of Relationship between
Earned Premium and Relative Estimated Standard Errors of Reserve Estimates

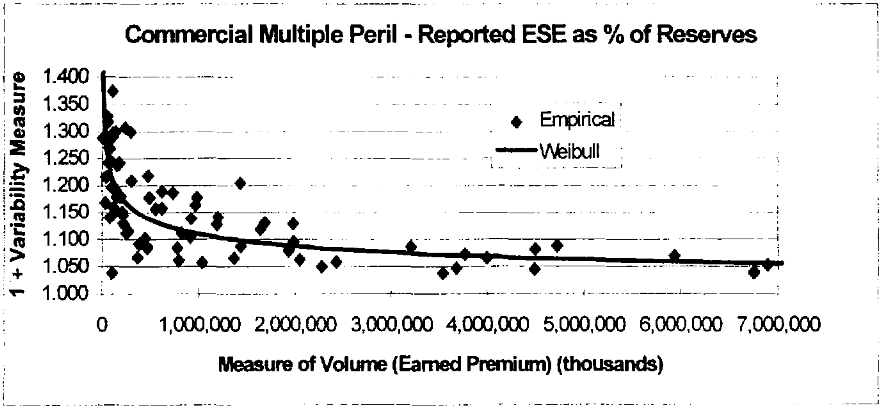
	R^2 statistic		F-statistic	
Line	Paid Basis	Reported Basis	Paid Basis	Reported Basis
Comm. Auto Liability	0.533	0.583	88	108
Comm. Multiple- Peril	0.595	0.630	119	138
Homeowners	0.566	0.542	115	104
Priv. Pass. Auto Liability	0.509	0.551	101	119

The R^2 statistics suggested that the curve fitting explained between one-half and two-thirds of the variation of the data points from their mean. Considering that volume of data is only one of a host of factors affecting the variability of actuarial estimates, we consider these findings to be very reasonable.

The F-statistics were very significant, confirming the hypothesis regarding volume of data as a major factor influencing the variability of reserve estimates.

A graphical presentation of the empirical data and the fitted curve for Commercial Multi-Peril based on the testing of reported loss development data is in Chart 1 below:

Chart 1



In Chart 1, an estimated standard error of 10% of the reserve estimate is shown as 1.100. The addition of 1.000 to the percentages enabled the curve fitting. Also, the term "Reported ESE" stands for the estimated standard error based on the testing of reported loss development data.

The chart illustrates the inverse relationship between the volume of data as measured by earned premium and the measure of variability, as measured by the estimated standard error expressed as a percentage of the total reserves estimate for all years combined. The data series in the chart labeled as "Empirical" represents the data points associated with each of the companies.

Exhibits 1 to 4, Sheets A and B, contain graphical presentations of the empirical data and the fitted curves for all four lines of business based on paid and reported data. All of the charts appear about the same, in terms of the relationship among the fitted curve and the dispersion of the empirical data points.

We offer several observations based on Chart 1:

- The expected inverse relationship between premium volume and relative estimated standard errors is confirmed. (This observation is applicable to all four lines.)
- For companies with approximately $1 billion of aggregate 10-year premiums (or roughly $100 million per year), there is a dispersion of the estimated relative standard errors. The relative errors are generally in the range from 5% to 20%.
- For companies with less than approximately $250 million aggregate premiums (or roughly $25 million per year), the relative standard errors generally range from 15% to 40%.

- Only a few companies appeared to achieve a relative standard error less than 5% (that is, a "1 + Variability Measure" of 1.050 or less).

How Much Volume to Achieve a 5% Tolerance on Reserves?

A common rule of thumb for evaluating the reasonability of a company's recorded reserves is a 5% tolerance around the actuary's point-estimate. Having evaluated the relationship of earned premium with the relative standard errors, we are now ready to begin our discussion of how much data our testing would suggest as being necessary to achieve such a tolerance. Using the curves as fitted from our empirical testing, we determined the premium volumes that are suggested in order to achieve a relative standard error of 5% of reserves for each line. The results are summarized in Table 2:

Table 2
Aggregate Earned Premium Volume Suggested
to Achieve Relative Standard Error of 5% of Reserves

Line	Paid Basis	Reported Basis
Comm. Auto Liability	$ 6 billion	$ 3 billion
Comm. Multiple- Peril	10 billion	9 billion
Homeowners	23 billion	13 billion
Priv. Pass. Auto Liability	5 billion	5 billion

While at first, the reader may gasp at the magnitude of these figures, the reader should keep in mind that these represent the aggregate earned premiums for the experience period for the reserving study. If the experience period contains 10 years, then the suggested annual premium volume would be "only" $500 million for private passenger auto liability.

We also observe that the reported development histories achieved the 5% tolerance (in relation to reserves) at a premium volume less than that for the paid development histories. This lends some support to a common view that the inclusion of case reserves in reserving analysis provides additional "value." Another view on this finding is that reported losses are almost always larger than paid losses, and therefore, reported loss development data may demonstrate lesser variability than paid loss development data.

How Stable are Industry Aggregate Development Data?

Before discussing the results further, we share the results of applying the Mack algorithm to the data for the aggregate property/casualty industry as reported in Best's Aggregates and Averages. The results are summarized in Table 3.

Table 3
Summary of Testing on Industry Data
Estimated Standard Error as Percentage of Est. Total Reserves

Line	Earned Premium	Paid Basis	Reported Basis
Comm. Auto Liability	$ 116B	2.7%	1.6%
Comm. Multi- Peril	169B	2.7%	3.1%
Homeowners	197B	9.3%	4.6%
Priv. Pass. Auto Liability	483B	2.7%	2.4%

In relation to estimated reserves, the estimated standard errors were generally in the 2% to 3% range, except for Homeowners. We offer further discussion of the Homeowners results in the section *Homeowners - Revisited* later in this paper.

Keeping in mind the apparently high volumes of earned premium suggested in Table 2 to achieve a 5% relative standard error on reserves, Table 3 shows that even the industry experience, with its tremendous volume of data, still maintains a degree of variability of reserve estimates. In fact, for Homeowners, on a paid basis, the experience of the industry did not achieve a 5% relative standard error, while the reported loss development experience achieved no better than a 4.6% relative standard error.

The suggested premium volume requirements (Table 2) to achieve a 5% relative standard error on reserves seem quite onerous. In our results database, we also retained the estimates of ultimate losses. In the next section, we shift the focus of our testing and presentation of results to our evaluation of volume requirements to achieve a 5% standard error relative to estimated ultimate losses.

Shifting the Testing Toward a 5% Tolerance in Relation to Ultimates

The statistical significance of the curve fitting of the estimated standard errors expressed as percentages of ultimates against earned premiums was not quite as good as for relative standard errors to reserves, but the statistics still suggested a meaningful relationship. Based on the sample of companies in our testing, the indicated earned premiums to achieve a 5% estimated standard error in relation to the ultimate loss estimate are summarized in Table 4 below:

Table 4
Aggregate Earned Premium Volume Suggested
to Achieve Relative Standard Error of 5% of Ultimate

Line	Paid Basis	Reported Basis
Comm. Auto Liability	$ 6 million	$ 4 million
Comm. Multiple- Peril	27 million	35 million
Homeowners	300 thousand	450 thousand
Priv. Pass. Auto Liability	15 million	5 million

These suggested premium requirements are substantially lower than those suggested for a tolerance of 5% of reserves. In fact, the suggestion for Homeowners looks too low. Let us emphasize that we are not presenting these as "the answers" to this difficult question. The main point to take from these results is that the volume requirement to achieve a 5% tolerance in relation to ultimates is substantially lower than the volume required to achieve a similar tolerance in relation to reserves. (See the *Areas for Further Investigation* section for thoughts on additional work that may be needed to get "the answers.")

We again refer to the results of applying the Mack algorithm to the aggregate industry data. This time, however, we express the estimated standard errors in relation to the ultimate loss estimates. The results are summarized in Table 5.

Table 5
Summary of Testing on Industry Data
Estimated Standard Error as Percentage of Estimated Ultimate

Line	Earned Premium	Paid Basis	Reported Basis
Comm. Auto Liability	$ 116B	0.7%	0.4%
Comm. Multi- Peril	169B	0.6%	0.7%
Homeowners	197B	0.6%	0.3%
Priv. Pass. Auto Liability	483B	0.5%	0.4%

The results demonstrate the "law of large numbers" very much in effect for estimating ultimate losses based on industry data. The aggregate industry data suggest very small estimated standard errors in relation to estimated ultimate losses. The smallest relative standard error was on a reported basis for Homeowners (0.3%); this "makes sense," since Homeowners would likely be ranked as the line with the quickest loss reporting among the four. Recall, however, that Homeowners was the line with the largest estimated standard error in relation to reserves.

We are left in a quandary. How do we explain the differences in apparent volume requirements to achieve 5% tolerances for reserves and for ultimates? And, how can Homeowners be the most "certain" with respect to ultimates but apparently the "least certain" with respect to reserves?

A Different View as to Why Small Variability of Ultimates can be Equivalent to Large Variability of Reserves

We offer a different way of looking at this situation, which will help address the questions raised above and provide another way to evaluate the overall reasonability of the results our testing has generated.

In Table 6 on the next page, we present this alternate view for Homeowners and Private Passenger Auto Liability. The basic premise is to assume the number of years that an experience period must include in order to meet a 5% of ultimate tolerance, and then assess the implications of that 5% of ultimate tolerance expressed in relation to the unpaid portion of the estimated ultimate losses.

For Homeowners, assume that only one accident year is needed to meet the 5% of ultimate standard. Referring to Table 6, assume ultimate losses for that year of 100 (row 1, column 2). An industry loss payment pattern (column 4) suggests that a majority (69.2%) of the ultimate losses have been paid. Therefore, 30.8% of the ultimate is in an unpaid (reserve) status; the implied unpaid for the first accident year is, therefore, 30.8% of 100, for 31 (rounded), shown in column 7. A 5% tolerance of ultimate (or 5 out of 100), represents 16% of the unpaid losses (column 9).

Continuing with the example for Homeowners, assume that more years are required to meet the 5% of ultimate standard. An increasing relative variability of reserves is implied, to the point where, if 10 years were required, the 5% of ultimate standard implies (or is equivalent to) an 80% of reserves standard. The larger relative variability to reserves arises from adding more (older) data to the experience period, where the additional experience is mostly paid, with little increase to the reserve level.

The illustration of this alternate view applied to Private Passenger Auto Liability is presented in the lower portion of Table 6. The same 5% of ultimate standard implies a variability of reserves of 8% to 32% (depending on the number of accident years in the experience period). These implied variability measures in relation to reserves for Private Passenger Auto Liability are less than the implied values for Homeowners, as a greater proportion of the ultimate losses for Private Passenger Auto Liability are in an unpaid status.

Exhibit 5 presents a similar analysis for Commercial Auto Liability and Commercial Multi-Peril.

The implications follow from Mack's conclusion cited earlier that the estimated standard error of ultimate estimates is the same as that for reserve estimates. Since ultimates are larger than reserves, the ratio of a standard error to ultimates is less (sometimes substantially less) than the ratio of the same standard error to reserves. Our numerical examples in Table 6 and Exhibit 5 provide insights for the potential magnitude of the difference in the ratios.

Table 6
A Different Way to Assess the Implications of Tolerances to Ultimate
on Tolerances to Reserves

# Years Needed	Estimated Ultimate Each Year	Combined Estimated Ultimate	Payout Pattern	Paid To-Date Each Year	Combined Paid	Implied UnPaid	5% of Combined Ultimate	Implied Percentage to UnPaid
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	assumed	sum of (2)	assumed	(2) x (4)	sum of (5)	(3) - (6)	5% x (3)	(8) / (7)
growth	5%							
				Homeowners				
1	100	100	69.2%	69	69	31	5	16%
2	95	195	91.0%	87	156	39	10	25%
3	91	286	94.6%	86	242	44	14	32%
4	86	372	96.7%	84	325	47	19	39%
5	82	455	98.0%	81	406	49	23	46%
6	78	533	98.8%	77	483	50	27	53%
7	75	608	99.3%	74	557	50	30	60%
8	71	679	99.6%	71	628	51	34	67%
9	68	746	99.7%	67	695	51	37	73%
10	64	811	99.8%	64	760	51	41	80%
				Private Passenger Auto Liability				
1	100	100	34.6%	35	35	65	5	8%
2	95	195	67.1%	64	99	97	10	10%
3	91	286	82.6%	75	173	112	14	13%
4	86	372	90.9%	79	252	120	19	15%
5	82	455	95.3%	78	330	124	23	18%
6	78	533	97.5%	76	407	126	27	21%
7	75	608	98.6%	74	480	127	30	24%
8	71	679	99.2%	70	551	128	34	27%
9	68	746	99.5%	67	618	128	37	29%
10	64	811	99.6%	64	682	128	41	32%

A summary of the results of our evaluation from this alternate perspective for the four lines in our study is presented in Table 7:

Table 7
Implied Tolerances for Reserves
Based on a Constant 5% of Ultimate Tolerance
Based on a Different View

Line	Number of Accident Years Needed to Meet Data Requirement				
	1	2	3	5	10
Comm. Auto Liab	6%	7%	9%	12%	20%
Comm. Multi-Peril	8%	9%	11%	$\boxed{14\%}$	22%
Homeowners	16%	25%	32%	46%	80%
Priv. Pass. Auto Liab	8%	10%	13%	18%	32%

The interpretation of the highlighted value is: for Commercial Multi-Peril, if five accident years are needed to meet the volume requirement to achieve an expected 5% tolerance in relation to ultimate, using an industry-based loss payment pattern, the implied tolerance in relation to reserves is 14%.

Using the fitted curves from our empirical testing, we calculated the relative standard error of reserve estimates at the premium volume suggested to achieve a 5% of ultimate tolerance. The results are summarized in Table 8:

Table 8
Implied Tolerance for Reserves
Based on Premium Volume to Achieve 5% Tolerance for Ultimates
Based on the Empirical Testing

Line	Paid Basis	Reported Basis
Comm. Auto Liab.	39%	45%
Comm. Multi-Peril	27%	28%
Homeowners	76%	68%
Priv. Pass. Auto Liab.	29%	33%

The results from our evaluation using industry patterns and varying the number of accident years needed to achieve the data requirement for a 5% of ultimate standard (Table 7) and the results from our empirical testing (Table 8) are not inconsistent.

Homeowners - Revisited

For Homeowners, the estimated standard errors based on aggregate industry development data (Table 3) were 4.6% on a reported basis and 9.3% on a paid basis – the highest of the four lines tested. While these results may at first seem counter-intuitive, consider that the line's high proportion of property claims tend to be settled (at their ultimate value) quickly. The relatively smaller proportion of longer-settling claims contain a degree of variability, which, in relation to outstanding liabilities, is relatively greater than for other lines where the level of unpaid reserves remains "higher" for a longer time.

FURTHER DISCUSSION

We can begin to form a general observation that the volume of data necessary to achieve a 5% standard error in relation to ultimate estimates may be associated with standard errors that are quite significant in relation to the estimates of unpaid liabilities.

Yet, the volume of data necessary to achieve 5% tolerance for reserves may be so high that complete reliance on a company's own loss development data would be a rare situation. In the practical arena, a blending of an individual company's experience with external benchmark experience may need to be a more common element of actuarial reserving analysis.

Or, a 5% tolerance for reserves may be considered too stringent, and therefore a broader tolerance may be suggested for practical purposes. (Let us be clear that we are not proposing any particular tolerance as being "acceptable," either from an actuarial view or a financial reporting view.)

Reference to Credibility Standards for Ratemaking

Credibility mainly relates to the data that is the subject of the analysis, while reliability relates to the estimates derived by the actuarial analysis of the subject data. Therefore, the concepts of credibility and reliability are very much related. The syllabus for Part 6 of the CAS Exam contains a number of papers that describe full and partial credibility standards for ratemaking. The standards that are discussed tend to share a common trait of being expressed (directly or indirectly) in terms of number of claims.

Some of the historic credibility standards were rooted in "simple" terms based on parameters and assumptions for loss distributions. These credibility standards often took the form of:

$$\text{Standard} = (\text{ t-statistic } / \text{ tolerance })\,\hat{}\,2$$

The historical parameters of this formula for General Liability were a t-statistic of 1.96 (for a 95% confidence interval) and a tolerance of 7.5%, which combined to generate a full credibility standard of 683 claims. For auto liability, the parameters were 1.645 (for a 90% confidence interval) and 5%, for a standard of 1,082 claims.

These historic standards mainly reflected variability of claim frequency with an assumed poisson distribution, and did not reflect variability of claim severity. More recent credibility standards have been enhanced for more advanced treatment of frequency and severity distributions.

Nevertheless, the credibility standards in ratemaking speak to the volume of data needed to achieve a certain tolerance in relation to ultimate. These cannot be used for reserving without, at a minimum, assessing the implications of such tolerances in relation to reserves.

Using the Historic Formula to Assess Implications for Reserving

The historic formula that utilized the t-statistic and the selected tolerance level in relation to ultimates can provide insights as to the potential magnitude of volume requirements to achieve the same tolerance level in relation to reserves.

Suppose that, for General Liability, a 7.5% tolerance in relation to reserves was equivalent to a 2% tolerance for ultimates. Reducing the tolerance value to 2% from 7.5% in the formula would increase the required volume by a factor of 14.

In the case of auto liability, assume that the 5% tolerance in relation to reserves was equivalent to a 1% tolerance for ultimates. Reducing the tolerance factor to 1% would increase the volume requirement by a factor of 25.

Volume of Data and Homogeneity

In many cases, actuaries are faced with decisions as to the segmentation of their loss development data, with an eye toward enhancing the homogeneity of individual segments while preserving some "credible" volume in each segment. For instance, the actuary must decide whether to analyze development data for auto liability combined, or separately for bodily injury, property damage, and personal injury protection.

Our discussion of the volume of data that may be needed to achieve certain tolerances in relation to reserves should not automatically lead actuaries to combine data sets that have been analyzed separately for the sole purpose of achieving a greater volume of data in a combined development triangle. The goal of achieving credible volumes of data should not replace the goal of maintaining homogeneous segments. While the two goals are in natural conflict with each other, the actuary must strive for a balance. The Mack approach is one way to evaluate whether a proposed segmentation of a loss development data set would improve the relative variability of the estimates for each segment due to enhanced homogeneity or whether it would hurt the relative variability due to smaller volumes of data in each segment.

AREAS FOR FURTHER INVESTIGATION

We recognize that the results of our testing cannot be considered conclusive, and, possibly, may raise more questions than are answered in this paper. We offer the following ideas for areas for further investigation:

Increased Sample

Naturally, an increased sample of companies is one area for further testing.

"Scrubbing" Procedure for Unusual Data Points

A procedure that would identify and "scrub" unusual data points in the historical data triangles may serve to minimize the degree of variability associated with the estimates. How "deviant" must a data point be in order to be considered "unusual?"

Confidence Intervals

Credibility standards in ratemaking are generally expressed as the volume of data in order that the estimates be within x% of the expected value at a confidence level of n%. The Mack algorithm describes an approach for measuring standard errors (deviations); the actuary will in many cases determine a point-estimate, which may serve as the mean of the distribution of reserve estimates. But the form of the distribution is not addressed. Mack's paper discusses the arithmetic, if one assumes an underlying normal distribution. Extension using the log-normal is also reasonably practical. But, nevertheless, this is not an area where the research is complete.

Aggregating Results across Lines

Our testing was done on a line by line basis. Many companies underwrite multiple lines, which may serve to reduce the total variability associated with aggregated reserve estimates. As evidence of the industry's acknowledgment of some degree of independence among lines of business, the NAIC Risk-Based Capital (RBC) formula provides for a credit to the overall RBC calculated for reserves risk to companies that underwrite multiple lines of business. (We note that the RBC formula does not explicitly provide for a "credit" based on volume of business alone.)

In our study, we did not attempt to evaluate independence among various lines of business. Clearly, the expectation would be that an analysis of aggregate variability for many lines on a combined basis would yield smaller overall premium requirements for a specified tolerance. The reduced, aggregate requirements, however, would need to be "distributed" to the individual lines, as actuaries generally prefer to analyze loss development by line of business.

What External Information to Incorporate?

The CAS Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves suggests that for "very small" groups of claims, the use of external information such as industry aggregates may be necessary.

We suggest that external information should be viewed, first, as that from a peer group of companies operating in the same market, or second, as that from the entire market for the geographic area under consideration (state or region). These references are, by definition, more refined than the information derived from Best's Aggregates & Averages. The latter source represents an aggregation of the development data from "all" companies (large, medium, and small) and for all states.

A study of the dispersion of loss payment and reporting patterns, for various peer groups of companies, in relation to benchmark patterns based on aggregate industry data, would be interesting. Of particular interest would be the impact of using an aggregate industry benchmark versus a peer group benchmark on the estimates for an individual "small" company.

Extension of Ratemaking Credibility Standards

We have demonstrated an empirical testing approach for evaluating volume requirements for achieving a selected tolerance in relation to reserves.

An alternate approach may be to utilize the basic analysis underlying the credibility standards for ultimate, and to supplement it with an analysis of variability of loss payment patterns. In this way, the analysis of the expected variability of estimates of ultimates would be extended to the expected variability of estimates of unpaid liabilities.

CONCLUSION

This paper presented an empirical testing of the relationship between a measure of variability of reserve estimates (being the relative estimated standard error based of Mack's algorithm) with the volume of underlying data (earned premium). We found the relationship to be statistically significant based on a sample of companies, although premium volume is not the sole predictor variable for the relative variability measure.

We observed that a relatively small tolerance in relation to an ultimate loss estimate can be equivalent to a relatively large tolerance in relation to a reserve estimate. We concluded that the volume requirement to achieve a certain tolerance for reserves is substantially greater than that to achieve the same tolerance expressed in relation to ultimate.

The implications on the reserving practices of actuaries are significant, and yet have only begun to be explored. We propose that actuaries and other reserving specialists must, as a common element of reserves analysis, take steps to:

- evaluate the variability of the underlying data (The Mack approach is one of several that can be used to assess the variability present in a particular loss development data set.)
- assess the implications for the tolerances that can realistically be expected from the use of the subject company's data alone, and
- consider the need for blending external benchmark loss development information with the company-based experience.

So, how much is enough? Alot.

Appendix I

In order to assist the reader in understanding the method used in our testing to calculate the estimated standard errors that serve as the measure of variability, we present two examples in this appendix.

EXAMPLE 1

We created a loss development triangle consisting of six accident years valued at annual intervals. We established the illustrative data to introduce variability in development factors up to the 36 month maturity, with no variation in development factors for later intervals. The data are shown below:

Paid Amounts						
Accident Yr	12	24	36	48	60	72
1991	760	1,140	1,311	1,377	1,390	1,390
1992	872	1,221	1,465	1,538	1,554	
1993	960	1,536	1,690	1,774		
1994	877	1,272	1,494			
1995	763	1,183				
1996	790					

The age-to-age development factors are:

Age-to-Age Factors						
Accident Yr	12 - 24	24 - 36	36 - 48	48 - 60	60 - 72	To Ult
1991	1.500	1.150	1.050	1.010	1.000	
1992	1.400	1.200	1.050	1.010		
1993	1.600	1.100	1.050			
1994	1.450	1.175				
1995	1.550					
1996						

The reader can see how development factors for the 12-24 and 24-36 month intervals demonstrate a degree of volatility, while the factors for the later intervals are the same for all accident years, demonstrating no apparent variation.

The volume-weighted average of the age-to-age factors for all years are:

Averages of Age-to-Age Factors						
	12 - 24	24 - 36	36 - 48	48 - 60	60 - 72	To Ult
Volume Wtd All Yrs	1.501	1.153	1.050	1.010	1.000	

Using a tail factor of 1.00 for development beyond 72 months maturity, the derivation of estimated ultimate losses and the estimated reserves is shown below:

Estimated Ultimate, Reserves, and Standard Errors						
Accident Yr	Paid Amounts	Factor to Ultimate	Estimate of Ultimate	Estimate of Total Reserve	Estimated Standard Error	Std Error as % of Total Reserve
1991	1,390	1.000	1,390	0		
1992	1,554	1.000	1,554	0	0	
1993	1,774	1.010	1,792	18	0	0%
1994	1,494	1.061	1,585	90	0	0%
1995	1,183	1.223	1,446	264	65	24%
1996	790	1.835	1,450	660	110	17%
Total	8,185		9,216	1,032	133	13%

The estimated standard errors for each accident year reflect the variability associated with the remaining development intervals. Therefore, for accident years 1994, 1993, and 1992, which are at or beyond the 36 month maturity, the estimated variability for remaining development was nil. That is, the estimated standard errors for accident years 1992 to 1994 and prior were 0.

We observe also that no standard error was calculated for accident year 1991, the oldest year in the experience period. The empirical data in the loss development triangle provided no points or "clues" about variability associated with development beyond 72 months. Therefore, one condition that the data should satisfy for the Mack approach is that the development history extend long enough so that no further development is expected.

For accident year 1995, the historical development experience suggested variability in the age-to-age factor for the 24-36 month interval. The algorithm evaluated the variability of the historical factors around the selected factor, and estimated the standard error to be 65. In relation to the estimated total reserve of 264, the relative standard error was 24% for that year.

For accident year 1996, the most immature year in the experience period, the variability observed for the 12-24 and 24-36 month intervals was quantified by the algorithm to be 110, or 17% of the estimated reserve of 660.

For evaluating the estimated standard error of the reserve estimate for all accident years combined, the Mack method assumes independence among the accident years. Therefore, the estimated standard error for all years combined is less than the sum of the estimated standard errors for each of the years. In the example, the overall estimated standard error was 133, or 13% of the total reserve estimate of 1,032.

EXAMPLE 2

We provide a second example that reflects variability throughout the entire loss development triangle, bringing the example closer to the types of data that were present in our empirical testing. The development data are:

Paid Amounts						
Accident Yr	12	24	36	48	60	72
1991	760	1,445	1,769	1,996	2,050	2,076
1992	872	1,540	1,878	2,115	2,198	
1993	960	1,691	2,050	2,298		
1994	877	1,536	1,976			
1995	763	1,540				
1996	790					

The historical age-to-age development factors are shown below:

Age-to-Age Factors						
Accident Yr	12 - 24	24 - 36	36 - 48	48 - 60	60 - 72	To Ult
1991	1.901	1.224	1.128	1.027	1.013	
1992	1.766	1.219	1.126	1.039		
1993	1.761	1.212	1.121			
1994	1.751	1.286				
1995	2.018					
1996						

The volume-weighted average factors are shown below:

Averages of Age-to-Age Factors						
	12 - 24	24 - 36	36 - 48	48 - 60	60 - 72	To Ult
Volume Wtd All Yrs	1.832	1.235	1.125	1.033	1.013	

Using a tail factor of 1.00 for development beyond 72 months maturity, the derivation of estimated ultimate losses and the estimated reserves is shown below:

Estimated Ultimate, Reserves, and Standard Errors						
Accident Yr	Paid Amounts	Factor to Ultimate	Estimate of Ultimate	Estimate of Total Reserve	Estimated Standard Error	Std Error as % of Total Reserve
1991	2,076	1.000	2,076	0		
1992	2,198	1.013	2,226	28	11	40%
1993	2,298	1.046	2,405	107	27	25%
1994	1,976	1.177	2,326	350	27	8%
1995	1,540	1.454	2,239	699	75	11%
1996	790	2.664	2,104	1,314	165	13%
Total	10,878		13,376	2,498	199	8%

The estimated standard errors progressed from being small for the oldest year (e.g., 11 for accident year 1992), then increasing as the maturities of the experience decreased, with the largest estimated standard error calculated for the most recent year (165 for the 1996 year). We also observe a similar progression of the estimated total reserves by accident year. The relationship of the estimated standard error to the estimated total reserve, however, depends on the proportion of ultimate losses that have been paid to-date.

For all accident years combined, the estimated standard error was 199, or 8% of the total reserve estimate.

Appendix II

ASSUMPTIONS AND LIMITATIONS OF THE MACK METHOD

The paper written by Dr. Mack identified several criteria that underlie the valid application of the chain ladder method to a set of loss development data. Two of these criteria are:

1) the absence of calendar period effects (along the diagonals of a left-oriented (loss period by age) development triangle, and

2) the independence among pairs of link ratios (age-to-age development factors) for neighboring development intervals

Our understanding of the algorithm suggests an additional, inherent assumption: that the entire triangle is "relevant" for evaluating ultimate costs. If the link ratios for certain intervals "long ago" are erratic and not representative of current experience, the reserving specialist will likely "ignore" such outlying observations when making selections for factors used to estimate ultimate losses. Alternatively, if the claim settlement practices or case reserving practices of the company have changed, the more recent factors would generally be considered more relevant than the factors from periods before the changes. The algorithm described by Mack uses the entire triangle, and evaluates the variability of all historical observations around the selected factors, in order to derive the estimates of standard error of the ultimate loss estimates.

In addition, we observe that the algorithm does not evaluate the variability associated with the tail development factor (since there were no "tail period" data in the development triangle by which to evaluate such variability). This feature was one of our considerations in our election of the four lines (commercial auto liability, commercial multi-peril, homeowners, and private passenger auto liability), where development at 120 months is virtually complete.

Exhibits

Description		**Exhibit**
Chart of Empirical Data and Fitted Curve		
Commercial Auto Liability		1
Commercial Multi-Peril		2
Homeowners		3
Private Passenger Auto Liability		4
For Exhibits 1 - 4	Sheet	
Paid Basis	A	
Reported Basis	B	
Implications of a Selected Ultimate Tolerance Level on Reserves Using Payout Patterns based on Industry Data		
Commercial Auto Liability		5
Commercial Multi-Peril		5

Exhibit 1
Sheet A

Commercial Auto Liability - Paid ESE as % of Reserves

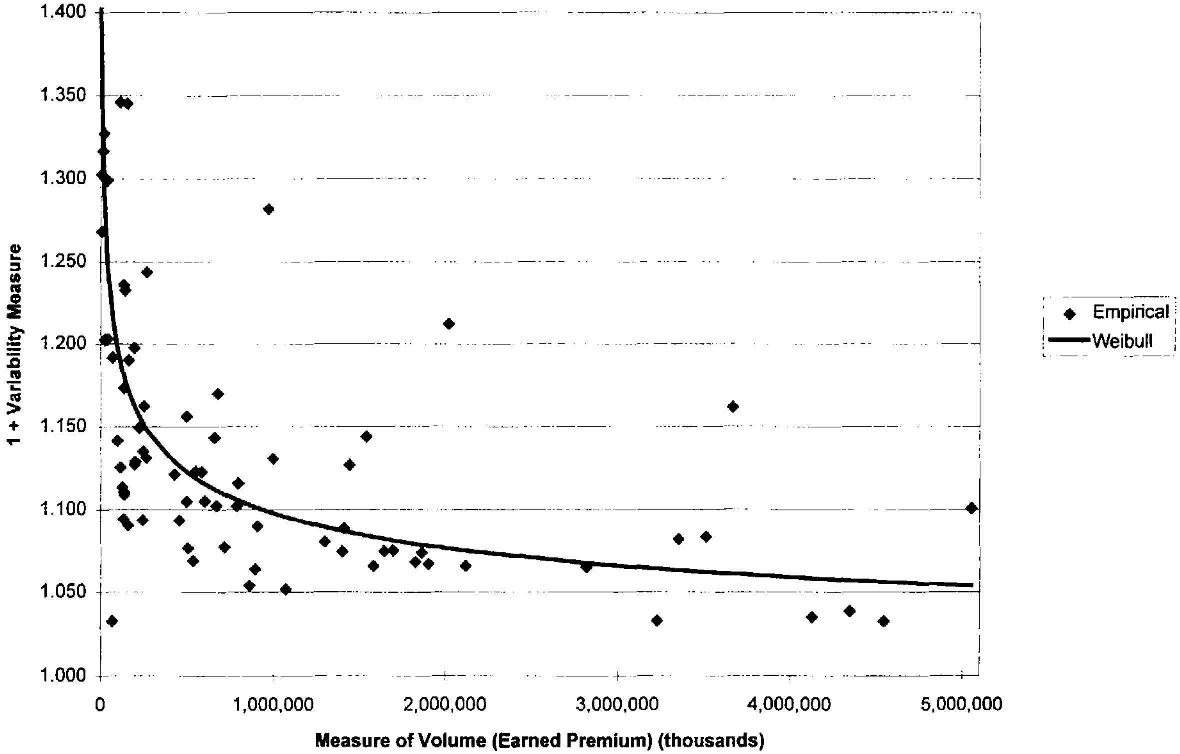


Exhibit 1
Sheet B

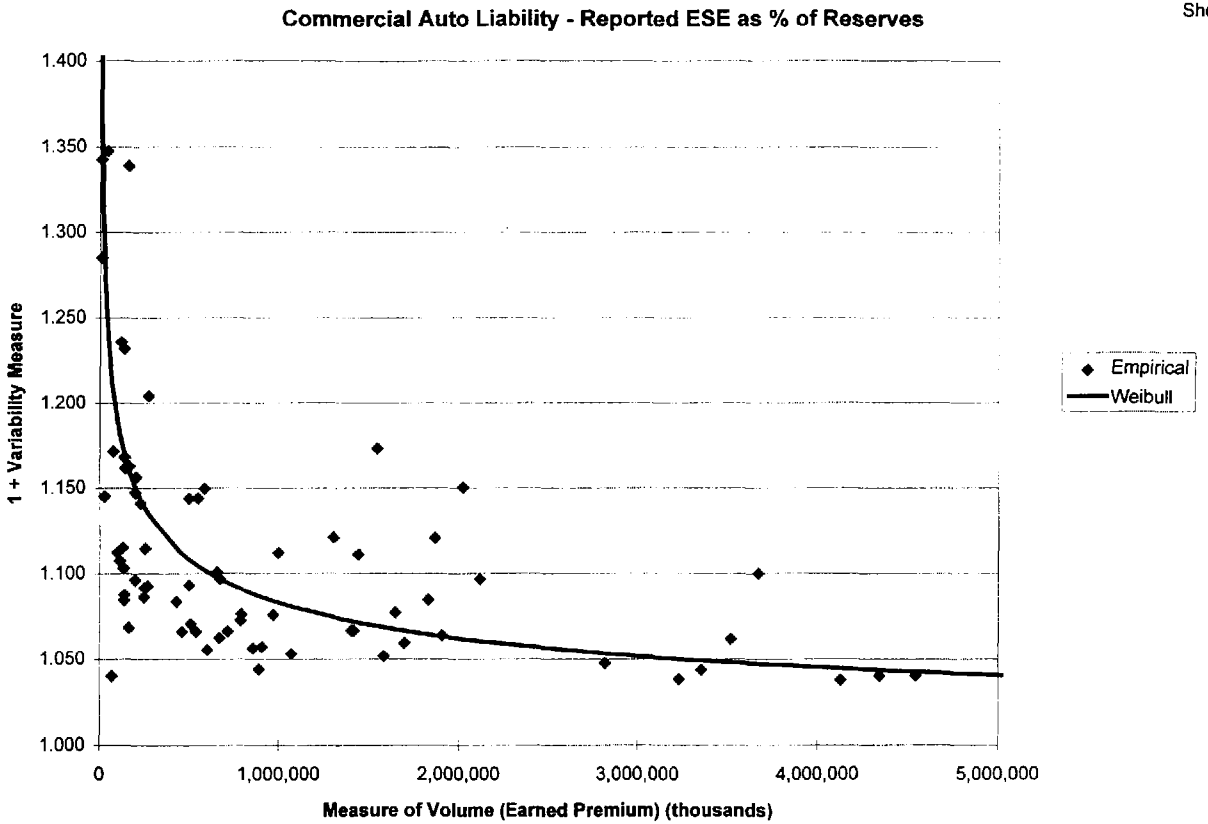
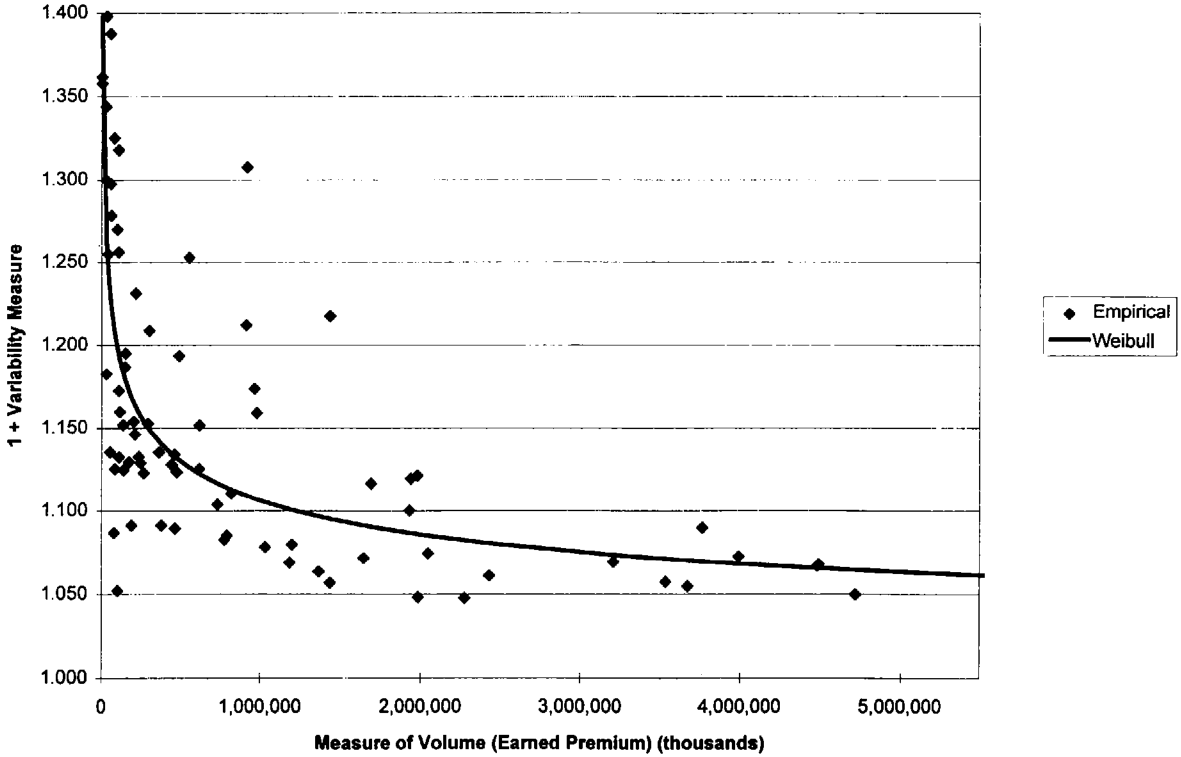


Exhibit 2
Sheet A

Commercial Multiple Peril - Paid ESE as % of Reserves



Commercial Multiple Peril - Reported ESE as % of Reserves

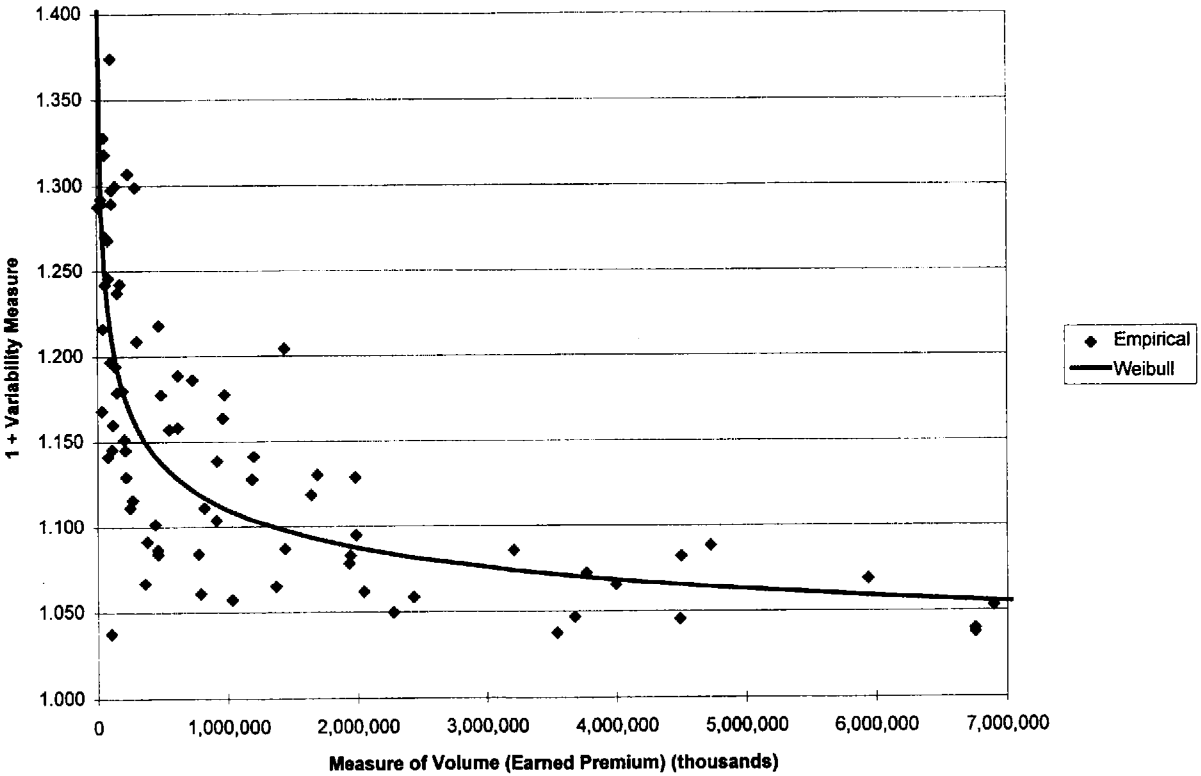
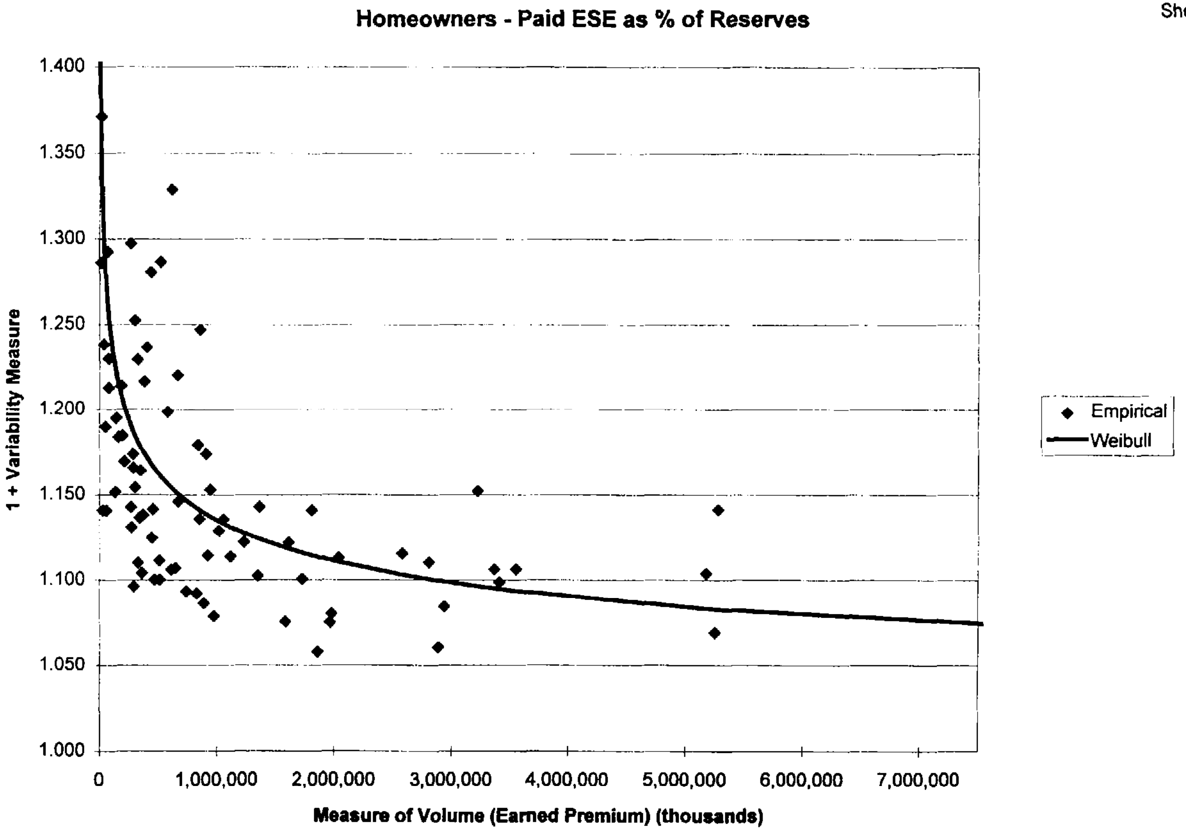


Exhibit 3
Sheet A



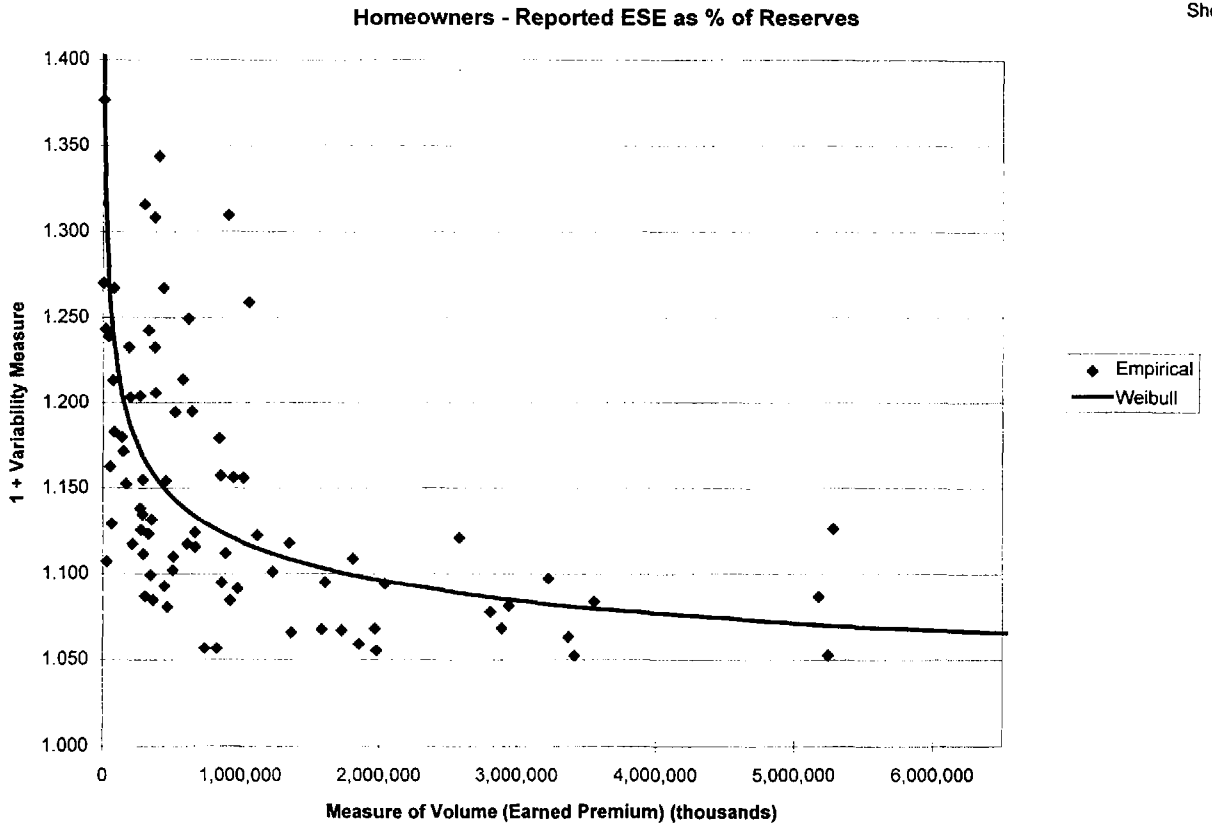

Homeowners - Reported ESE as % of Reserves
1.400
1.350
1.300
1.250
1.200
1.150
1.100
1.050
1.000
1 + Variability Measure
0
1,000,000
2,000,000
3,000,000
4,000,000
5,000,000
6,000,000
Measure of Volume (Earned Premium) (thousands)
Empirical
Weibull

Exhibit 4
Sheet A

Private Passenger Auto Liability - Paid ESE as % of Reserves

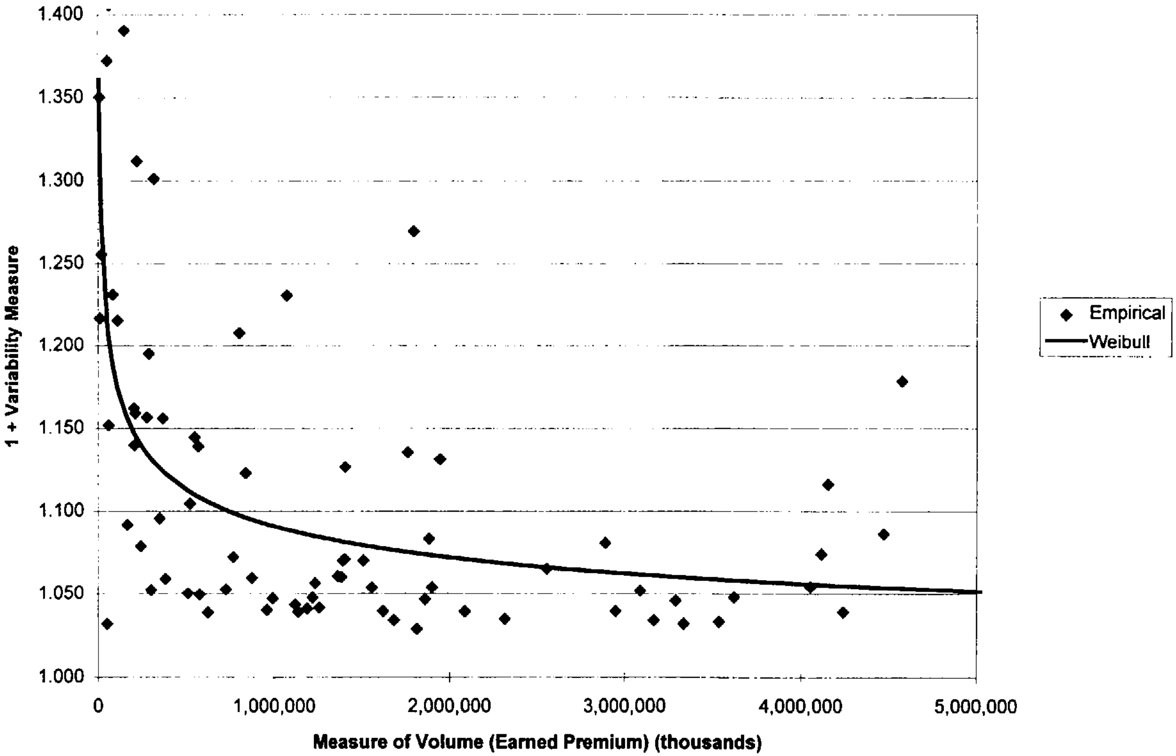


Exhibit 4
Sheet B

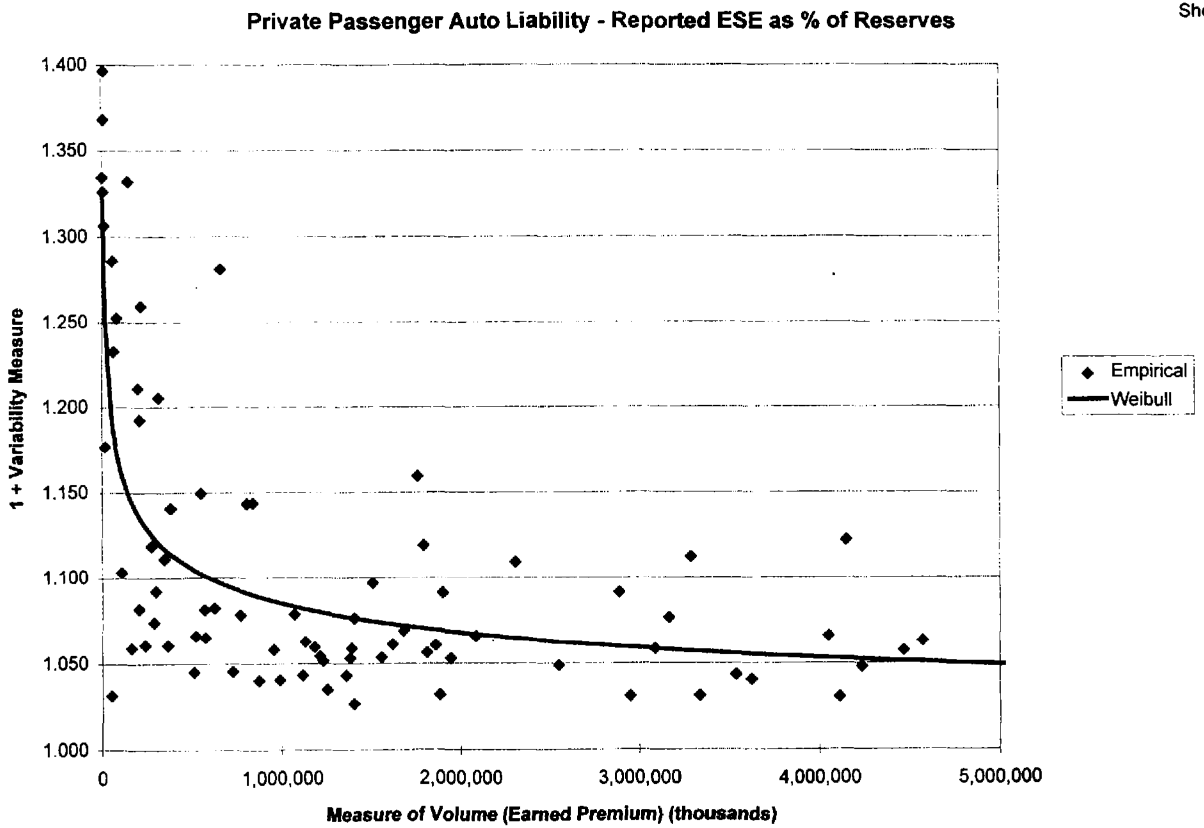


Exhibit 5

Implications of Credibility Standards to Ultimate for Reserves
Based on Estimated Payout Patterns from Industry Data

# Years Needed	Estimated Ultimate Each Year	Combined Estimated Ultimate	Payout Pattern	Paid To-Date Each Year	Combined Paid	Implied UnPaid	5% of Combined Ultimate	Implied Percentage to UnPaid
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	assumed	sum of (2)	assumed	(2) x (4)	sum of (5)	(3) - (6)	5% x (3)	(8) / (7)
growth	5%							
			Commercial Auto Liability					
1	100	100	19.8%	20	20	80	5	6%
2	95	195	45.5%	43	63	132	10	7%
3	91	286	65.9%	60	123	163	14	9%
4	86	372	80.0%	69	192	180	19	10%
5	82	455	88.5%	73	265	190	23	12%
6	78	533	93.2%	73	338	195	27	14%
7	75	608	95.8%	72	409	198	30	15%
8	71	679	97.2%	69	478	200	34	17%
9	68	746	98.0%	66	545	202	37	19%
10	64	811	98.4%	63	608	203	41	20%
			Commercial Multi-Peril					
1	100	100	34.5%	35	35	65	5	8%
2	95	195	56.9%	54	89	107	10	9%
3	91	286	68.3%	62	151	135	14	11%
4	86	372	77.7%	67	218	155	19	12%
5	82	455	85.0%	70	288	167	23	14%
6	78	533	90.0%	71	358	175	27	15%
7	75	608	93.6%	70	428	180	30	17%
8	71	679	95.8%	68	496	183	34	19%
9	68	746	97.3%	66	562	184	37	20%
10	64	811	98.2%	63	625	186	41	22%