Considerations Regarding Standards of Materiality in Estimates of Outstanding Liabilities

Emmanuel T. Bardis, FCAS, MAAA, PhD, MBA
Christina L. Gwilliam, FCAS, MAAA
Stephen P. Lowe, FCAS, MAAA
Atul S. Malhotra, FCAS, MAAA, MBA

Emmanuel T. Bardis
Towers Perrin
111 Huntington Avenue
8th Floor
Boston, MA 02199-7612
(617)638-3807
manolis.bardis@towersperrin.com

Christina L. Gwilliam
Towers Perrin
2107 Wilson Blvd.
Suite 500
Arlington, VA 22201-3062
(703)351-4807
christina.gwilliam@towersperrin.com

Stephen P. Lowe
Towers Perrin
Forestal Centre
175 Powder Forest Drive
Weatogue, CT 06089-9658
(860)843-7057
stephen.lowe@towersperrin.com

Atul S. Malhotra
ACE USA
455 Market Street
Suite 500
San Francisco, CA 94105
(415)547-4587
atul.malhotra@ace-ina.com

Abstract: This paper reports on our research into the issues associated with establishing standards for materiality associated with claim liability estimates. In our research we explored several alternative methods for developing benchmarks for materiality. Rather than restrict ourselves to theoretical considerations, we tested the various methods empirically using public data for individual companies and various lines of business. The empirical test results raise many practical issues that must be considered in such an exercise. This paper is meant to promote discussion on this topic and related issues.

Keywords: reserve variability; uncertainty and ranges; materiality; range of reasonable estimates; range of reasonably probable outcomes; statement of actuarial opinion

Table of Contents

1. EXECUTIVE SUMMARY

As a result of the recent accounting scandals and the stock market boom/bust, there has been an increased desire by shareholders, regulators and rating agencies for transparency in financial statements. Within the non-life / property and casualty insurance sector, the largest liability on an insurer's balance sheet is the loss reserve. There is an increased desire to better understand the uncertainty associated with estimates of unpaid claims underlying the loss reserve. A single point estimate gives no sense of the degree of certainty (or uncertainty) as to the likelihood that actual claim liabilities will ultimately be close to the estimate. Therefore actuaries are increasingly asked to supply a range of reasonably possible outcomes. In the U.S., Appointed Actuaries are required to identify significant risks and uncertainties that could result in material adverse deviation in the loss reserve, and to specify the materiality standard for the specific company. There is little guidance on how to estimate the range of reasonable estimates, or on what this materiality standard should be. This paper seeks to explore ways to measure reserve volatility and to assist the actuary in these areas. In the context of the paper we develop a framework that is designed to answer two distinct questions:

- *By what amount must two estimates of unpaid claim liabilities differ to be considered materially different from each other?*
- *What is the magnitude of the reasonably probable total deviation in actual claim liabilities from the estimate of expected claim liabilities?*

Both of these questions are related to the volatility of the claim generation process characterizing non-life / property and casualty exposures, but they focus on different issues that arise from the uncertainty the volatility creates. Note that materiality in the context of actuarial opinions has a different meaning. For actuarial opinions, materiality is related to an adverse claim liability deviation that would significantly affect the viability of a company. Our use of the term materiality is explained in our Conceptual Framework in Section 2.3.

The first question gives rise to a *Range of Reasonable Estimates,* ideally reflecting uncertainties as to the parameters and model selected to produce estimates of the expected claim liabilities. The second question gives rise to a *Range of Reasonably Probable Outcomes*, incorporating process as well as parameter and model risk. Both ranges depend on standards that must give due consideration to statistical, financial, and solvency perspectives.

In addition to providing a framework for analyzing the two questions posed above, the paper reports on our empirical research, in which we explored several alternative methods for measuring process, parameter, and model risk, and for translating the amount of measured risk into benchmark ranges. The empirical test results raise many practical measurement issues that will require further research to resolve. While the paper presents empirical results for illustration and comparison, the ranges derived are subject to substantive limitations and should therefore not be considered a recommendation.

1.1 Research Approach

We designed our study using the framework of statistical hypothesis testing. We used data from the 2003 Annual Statement of a sample of U.S. insurers for the personal auto liability, homeowners, workers compensation and other liability lines of business. To measure uncertainty in the unpaid claims we used two stochastic methods on individual lines of business: the Bootstrapping methodology of England and Verrall, and the Mack stochastic methodology[1]. The coefficients of variation resulting from our analysis provide a measure of the reserve volatility. As explained in more detail in subsequent sections of this paper, we endeavored to bifurcate total volatility into process and parameter risk. Next, we used two approaches to estimate materiality standards. The two approaches are a percentile/threshold approach and a tail value at risk (TVar) approach. Finally, these monoline results were combined to recognize the risk diversification benefits of multi-line writers. We used a Copula[2] type approach to aggregate the claim liability distributions. We note specifically that we have not used statistical hypothesis testing as our approach. Instead, we use the terminology or the framework associated with hypothesis testing to explain the results of our study for the reader's benefit.

1.2 Results

We derived indicated reserve ranges on two bases: the "range of estimation" basis, which is used to estimate the range of reasonable estimates, and the "range of outcomes" basis,

---

[1] The Bootstrapping and Mack methods are described in subsequent sections of the text as well as in Appendix A.
[2] Copula theory is described in subsequent sections of the text as well as in Appendix C.

which is used to estimate the range of reasonably probable outcomes. As shown in the table below, the outcome standards are higher than the estimation standards by an average of 75%.

Standards of Materiality – Mack

	Range of Estimation		**Range of Outcome**	
Line of Business	**Lower Tail Test**	**Upper Tail Test**	**Lower Tail Test**	**Upper Tail Test**
Personal Auto Liability	-5.8%	6.7%	-10.2%	12.2%
Homeowners	-9.7%	11.4%	-17.5%	21.5%
Workers Compensation	-13.6%	16.4%	-20.8%	26.2%
Other Liability	-16.4%	20.2%	-28.0%	37.7%

One reason for the difference between the two types of ranges is that outcome standards include process and parameter risk whereas estimation standards only include parameter risk.

Finally, we created a fictitious company that writes all four lines of business to see the benefit of risk diversification.

Standards of Materiality – Mack

Type	**Lower Tail**	**Upper Tail**
Range of Estimation	-12.4%	15.4%
Range of Outcomes	-14.4%	18.1%

1.3 Conclusions and Implications

The major conclusions of our studies were as follows:

- Materiality can have different implications when viewed from a statistical, financial or solvency perspective.
- Materiality standards should clearly be different in a Range of Estimation context than in a Range of Outcomes context.

- Standards of materiality should vary by line of business. Lines of business that historically exhibit higher volatility should have higher standards of materiality (i.e., wider ranges).
- Materiality standards can be arrived at using a framework of statistical hypothesis testing and applying techniques such as percentile/threshold and or TVar.
- Any approach to studying or deriving standards of materiality requires the measure of an appetite for adverse outcomes such as benchmark percentile/threshold of adverse deviation or benchmark exceedence ratio. In terms of the hypothesis testing framework, this relates to one's tolerance level for making a "Type I" or a "Type II" error. Specifically, all else being equal, a wide materiality standard range allows a higher probability of accepting the hypothesis that two reserve estimates are not materially different when in fact they are (i.e., it involves a higher probability of a Type II error). Conversely, a lower materiality standard increases the risk of a Type I error (i.e., concluding that two estimates are statistically different when in fact they are not).
- It is our recommendation that these benchmarks be derived based on combined industry data. Then materiality standards can be derived for individual companies using these benchmarks and their own implied volatility.
- The percentile/threshold and the TVar approaches used in this study yield different standards of materiality applied on the same data as they essentially measure volatility differently. The latter is a more conservative approach.
- Diversification for multi-line writers reduces overall volatility of liabilities compared to mono-line writers, requiring lower levels of surplus, and thus multi-line writers should have lowers standards of statistical and financial materiality compared to mono-line writers.
- The results of our analysis showed that financially impaired companies in general should have narrower standards of materiality compared to financially healthy companies.
- Some of the other conclusions that we reached as a by-product of our extensive use of standard stochastic methodologies are as follows:
- Standard volatility-measuring techniques overstate the volatility of the underlying loss exposure (loss generating process) when used on data without any adjustment for exogenous and endogenous factors impacting the company. For example, these methods

are influenced by trends, changes in case reserving levels, changes in claim settlement rates, and other factors. Adjustments should be carried out to scrub the triangles of these factors before these methodologies can be applied.

- The standard Mack and Bootstrapping stochastic methods usually give different measures (answers) for volatility of the underlying loss data. Our research on industry data showed that the Bootstrapping method has a tendency to overreact to sudden changes in data.
- Both the Mack and Bootstrapping stochastic methodologies give different results for volatility when applied to paid and incurred loss data of the same underlying loss exposure. Both methods apply with more confidence to paid loss development data. The results of these stochastic methods when applied to incurred loss development data, where negative development is prevalent, are not very credible
- The standard stochastic methodologies such as Mack and Bootstrapping do not perform well in differentiating between process and parameter risk. Loss data should be adjusted to a stationary basis in order to achieve a credible differentiation between process and parameter risk.

2. INTRODUCTION

2.1 Background

Actuaries today are being asked by the investment and regulatory communities not only to specify their best estimate of a property and casualty insurer's claim liabilities[3], but also to specify a range of reasonably possible outcomes around their best estimate. These requests in part are being driven by a spate of reserve increases taken by major insurers (particularly those writing U.S. business) in the last few years, which has heightened the issue of "reserve risk." In general there is a movement towards understanding the uncertainty or variability associated with estimates of claim liabilities, as the range around the estimate provides insight as to the solidity of the reserves recorded on the balance sheet (i.e., what percentile within the range of estimates does the carried reserve represent[4]?). Understanding the variability is important to the external stakeholders bearing the risk (shareholders and policyholders), and to the directors of the company who are responsible for managing its risk and capital. A single point estimate gives no sense of the degree of certainty (or lack of certainty) as to the likelihood that the actual claim liabilities will ultimately be close to the estimate.

Additionally, an issue that actuaries and directors of insurance companies often face is how to reconcile differences between alternative estimates of claim liabilities: management's estimate, internal actuarial estimates, and external actuarial estimates. In such instances, directors are faced with the difficult task of choosing a reserve to record based on one of the alternative estimates. How should they make this decision? Are these estimates different enough that one can assume that they are truly differences in opinion, or do they merely reflect differences in methods and assumptions that are within a range of reasonableness?

---

[3] Throughout this paper we refer generically to claim liabilities as being the uncertain amount that will ultimately be paid by the insurer to settle claims arising from insurance coverage that it has provided. The term is meant to be inclusive of defense, adjustment, and other settlement costs in addition to direct payments to the claimant.

[4] In this paper we do not address the issue of how an estimate of liabilities is translated into a reserve on the balance sheet. Generally the literature is vague on this subject, specifying for example that the company should record its "best estimate". While some may interpret this as implying that the reserve should be set equal to the mean estimate, others might interpret it as requiring that the reserve be set at the median, or some other percentile that includes a margin. For purposes of exposition we have therefore assumed that there is a pre-ordained mapping from the selected distribution of claim liability outcomes to an appropriate reserve; the focus of our inquiry is on the selection of the distribution itself.

Given two estimates that are different, are the differences between the two estimates material, and will the booking of reserves based on either of the estimates cause the users of the financial statement to draw different conclusions?

These issues have gained importance lately with changes to the year-end 2005 U.S. opinion process for non-life companies. The Model Law developed by the National Association of Insurance Commissioners, which has been adopted by a few states at this juncture but is expected to be adopted by most states, specifies that opinions should include an Actuarial Opinion Summary that details the opining actuary's own point estimate and range, if one was generated.

2.2 Purpose/Objective of the Paper

This paper is intended to address the following two questions, both of which arise as practical issues in actuarial practice today:

1. *By what amount must two estimates of claim liabilities differ to be considered materially different from each other?* This question often arises in the context of reserve opinions, for example when a reviewing actuary is comparing his or her estimate to management's estimate underlying the held reserve. For sufficiently small differences the conclusion should be that the two estimates are not significantly different. However, at some point the difference between the two estimates becomes sufficiently large that it is significant.

2. *What is the magnitude of the reasonably probable total deviation (adverse or favorable) in actual claim liabilities from the current estimate of expected claim liabilities?* This question arises in the context of solvency, for example when one is stress-testing the balance sheet against the possibility of adverse deviation from the expected level of claim liabilities that would have a significant impact on the company.

Both of these questions are related to the volatility embedded in the claim generation process characterizing non-life / property and casualty exposures, but they focus on different issues that arise from the uncertainty that the volatility creates. In responding to either question, actuaries need benchmark *standards for materiality*, typically expressed as a percentage of the claim liabilities[5], to guide their responses.

This paper reports on our research into the issues associated with establishing standards for materiality associated with claim liability estimates. In our research we explored several alternative methods for developing benchmarks for materiality. Rather than restrict ourselves to theoretical considerations, we tested the various methods empirically using public data for individual companies and various lines of business. The empirical test results raise many practical issues that must be considered in such an exercise.

This paper is meant to promote discussion on this topic and related issues. Our approach is not meant to be definitive, and our empirical results are subject to substantive limitations. The latter are provided for illustration and comparison, and should not be taken as a recommendation. We expect that our approach will continue to evolve with further exploration on the topic.

2.3 Conceptual Framework

The historical loss development data the actuary can use to estimate claim liabilities are a relatively small sample of realizations of the claim generation process. The actual claims generated in each accident or underwriting year are the result of (a) randomness, and (b) differences in environmental influences. These influences are both exogenous (the socio-economic conditions at the time) and endogenous (underwriting and claim handling procedures in place at the time). From the available data, the actuary is asked to discern the expected value of the claim liabilities, and the distribution of possible outcomes around that expectation. With imperfect knowledge, the actuary can only provide an *estimate* of the expected value and the underlying distribution, creating a second level of uncertainty above that inherent in the claim generation process itself.

Within a reserving context, actuaries attempt to estimate the true, but unknown, expected claim liabilities by applying an actuarial model to the available historical data. It helps to think about the uncertainty involved in estimating claim liabilities in terms of the following continuum:

---

[5] In certain contexts, materiality standards might also be expressed as a percentage of net income or capital.

Actual Claim Liabilities ⟷ True Expected Claim Liabilities ⟷ Model Estimate of Expected Claim Liabilities

The true expected claim liabilities could be considered as the indication from the "perfect" actuarial model where:

— there is no uncertainty associated with the models inputs; and

— all the assumptions employed by the actuarial model are correct.

The potential differences between the actual claim liabilities and the true expected claim liabilities are due to process risk while the potential differences between the true expected claim liabilities and the actuary's model estimate are due to parameter risk and model risk. A detailed description of all the risks associated with the measurement of claim liabilities follows.

- Process risk represents the fundamental uncertainty due to the presence of randomness when losses are generated. Even when an actuary can achieve a "perfect" model, the random nature within which losses are generated would prohibit that actuary from calculating the actual claim liability amount.
- Parameter risk is the uncertainty associated with the unknown parameters of statistical models, even if the selection of the model is correct (i.e., we might know with certainty that the link ratios at a certain maturity follow a log-normal distribution, but we are not sure about the correct parameters associated with that distribution); and
- Model risk is the risk associated with the uncertainty that the loss generating process is not represented correctly by the particular model selected.

Some actuarial literature separates that risk between model risk and specification risk; the former relates to the question if the selected model is correct while the latter relates to the question if the distributions employed by the model are correct). Model risk is the most difficult type of risk to measure since every stochastic model is based on the premise that its fundamental assumptions are correct. Traditional stochastic reserving models, including Mack and Bootstrapping, ignore model risk. One way of approximating model risk is hindcast testing. With hindcast testing a model employs a subset of the historical data to project losses for the remainder of the historical period and compare the actual and projected results. The resulting residuals provide a proxy for model risk.

In the context of uncertain claim liabilities, materiality must be examined from several different perspectives.

- The statistical perspective on materiality reflects the fact that one is estimating the shape and parameters of an unknown claim liability distribution.
- The financial perspective on materiality relates to the question: Would users of the financial statements draw different conclusions if the figures presented were different? This perspective draws on the other elements of the balance sheet, and the income statement.
- The solvency perspective on materiality links the uncertainties associated with the claim liabilities to the capital and claims-paying capacity of the enterprise.

Materiality questions arise most commonly in the context of alternative actuarial estimates, relating to the first question posed at the outset of our paper: Given the uncertainty in the estimation process, is the difference between one actuarial estimate of the claim liabilities and another actuarial estimate significant? In the context of this question we are concerned with the uncertainty of the *expected* liabilities (and not random variations between actual and expected, i.e., process risk); only parameter and model risk are relevant. In other words, the relevant distribution is the distribution of the estimated mean.

> The *Range of Reasonable Estimates* is the range within which alternative estimates of the expected claim liabilities would be deemed to be immaterial, in the sense that (a) the difference between the estimates is not statistically significant, and (b) the difference in the resulting reserves is not financially material. Within this range one could not say that one estimate was actuarially "better" than the other. An actuary reviewing the reserves of a company would accept the reserves if his or her own estimate were within this range.

Materiality can also arise in the context of solvency and risk management, in which one should consider the total risk embedded in the claim liability estimation process, including parameter, model and process risks. In this case we are interested in the actual liability outcomes, so we need to measure all types of risk that could have an adverse effect on a company's surplus.

> The *Range of Reasonably Probable Outcomes* is the range within which the alternative actual claim outcomes are expected to fall with reasonable confidence, in the sense that (a) the outcomes outside of the range which, while possible, have low statistical probability, and (b) for a reasonably well capitalized company, outcomes within the range would not threaten the solvency of the company.

In this paper we focus on materiality standards for the range of reasonable estimates and the range of reasonably probable outcomes. In the first context we refer to the relevant materiality as *estimation materiality*, which ideally will reflect only model and parameter risk. In developing estimation materiality standards we considered only the statistical perspective. We did not consider the financial or solvency perspective; however, as a refinement it might be appropriate to consider the financial perspective.

The latter range relates principally to the financial and capital management (or solvency) perspective on materiality and links the uncertainty associated with the actual claim liability distribution to the finances of the company. All types of risk (model, parameter, process) that could have an adverse effect on the income and capital needs of the company should be measured here. In this context we refer to the relevant materiality standard as *outcome materiality*. When measuring outcome materiality we considered the statistical and solvency perspective, but not the financial perspective.

We note that there is not a clear distinction between the concepts of Range of Reasonable Estimates and Range of Reasonably Probable Outcomes. The underlying precept of our analysis is reserve volatility, which is captured in the definition of Range of Reasonable Estimates. The Range of Reasonably Probable Outcomes is a slightly broader concept in that it tries to incorporate reserve volatility in conjunction with management input and the financial condition of the company (i.e. surplus).

Additionally, in setting materiality standards we did not consider other sources of risk, such as market, credit, operational or insurance underwriting risk.

In summary, for a given set of claim liabilities, the objective is to develop:

a) an appropriate standard for a *range of reasonable estimates*, reflecting appropriate criteria for *estimation materiality*; and

b) an appropriate standard for a *range of reasonably probable outcomes*, reflecting appropriate criteria for *outcome materiality*.

To develop these ranges, it is necessary to estimate the claim liability distribution and to separate process from parameter and model risk. As we discuss later in the Methodology section, the claim liability distributions in this paper are estimated with stochastic reserving methods, which provide distributions for both the actual claim liabilities and the estimate of the expected claim liabilities.

Once appropriate claim liability distributions have been produced, the two ranges embodying our materiality standards can be obtained from them. In the case of each distribution, this requires the selection or derivation of a threshold [5]. The threshold can be based either on a specified percentile of the distribution (generally, a VaR approach), or on a specified expected exceedence value (generally, a TVaR approach)

The percentile threshold approach is a point measure in the sense that it measures the probability of an outcome being worse than a given monetary threshold (e.g., probability of ruin). While the percentile threshold approach measures the probability that a particular value will be exceeded say once every 100 years, the expected "exceedence" threshold approach measures the expected value of the exceeded amount (every 100 years) when the threshold is exceeded. The expected exceedence threshold approach provides values higher than the percentile threshold approach, as it is influenced by the outcomes of remote loss outcomes. In the chart shown below the percentile threshold approach focuses on finding the shaded region, whereas the expected exceedence threshold approach focuses on estimating the expected value of losses exceeding the threshold, as a percentage of expected liabilities. Essentially, these two paradigms measure "tail" risk differently.

We formulate the problem of analyzing estimation materiality in the framework of statistical hypothesis testing. Although we do not actually perform hypothesis testing, this framework has the advantage of helping to explain the variables required to calculate materiality and analyze the results obtained from our analysis. The only divergence between a true statistical hypothesis testing and the methodology employed in this paper is that, while statistical hypothesis testing compares the distributions of two estimates of the mean, in this paper we compare the distribution of expected claim liabilities to an alternative point estimate of the mean that is considered to be certain. In that respect our employed approach resembles the measurement of a statistical confidence level.

Consider a distribution of expected claim liabilities where:

$C\alpha$ = the mean of the distribution

m_1 = the upper bound of the range of reasonable estimates

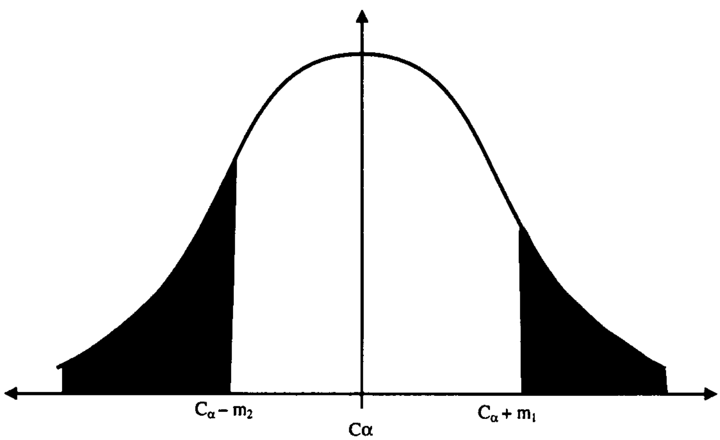
m_2 = the lower bound of the range of reasonable estimates

We can set up the problem in this framework as follows:

H_0 (Null Hypothesis): The two estimates of the expected claim liabilities are the same

H_1 (Alternate Hypothesis): The two estimates are not the same

A formulation of the problem pictorially is as follows:



Reserves in the range $C_\alpha - m_2$ to $C_\alpha + m_1$ are not considered significantly different

The Type I error in statistical hypothesis testing measures the probability of rejecting the null hypothesis when the null hypothesis is true. Typically, m1 and m2, defining the range of reasonable estimates, are determined by selecting a significance level, reflecting an acceptably low probability of a Type I error. The significance level is measured by "r" (in our paper), shown by the shaded region in the chart above. Note that as the stringency of the significance level is tightened, the range of reasonable estimates expands.

If the alternative estimate of the expected claim liabilities falls outside of the range of reasonable estimates (in other words if the alternative estimate amount falls in the shaded region in the chart above) then we can reject the null hypothesis that the original estimate underlying the reserve and the alternative estimate are essentially the same.

The formulation for analyzing outcome materiality follows a more traditional confidence level construct. However, the picture is essentially the same as that shown above. We seek to define a range of reasonably probable outcomes, such that the likelihood of actual claim liabilities being outside of that range is reasonably small. However, rather than defining the range purely from a statistical perspective, we define it with reference to a solvency perspective as well. The benchmark level of outcome materiality is based on an empirical analysis of the typical relationship of reserves to risk-based capital, and the level of adverse deviation that would cause the insurer to "ruin" by failing the risk-based capital adequacy test.

In the context of outcome materiality, a higher probability of ruin corresponds to a smaller range of reasonably probable outcomes.

For both types of ranges, we develop empirical measures of m_1 and m_2 in this paper. They may be interpreted as an explicit function of three primary variables amongst others:

$m = f(\sigma, r, \alpha)$ where:

> σ is the implied volatility of the claim liabilities for line of business under consideration, or the uncertainty of the estimated mean;
>
> r is the selected threshold. The corresponding factor in statistical hypothesis testing is the probability of Type I error; and
>
> α is the implied percentile of the carried reserves in relationship to the expected claim liabilities.

m (defining the upper or lower bound of the range) is directly proportional to σ. A more volatile book of business will require a larger allocation of surplus and thus will have a higher m. In other words, the more volatile a book of business, the greater the uncertainty associated with the claim liability estimates. As a result, the corresponding m should be greater for this line to consider the greater uncertainty of the loss process. In Step 1 of the Methodology section we outline how we calculated the implied volatility of each line of business.

α, the implied percentile of the carried reserves, is another important factor. If the carried reserves are booked at a higher percentile of the claim liability distribution then a lower standard of outcome materiality is acceptable.

As r increases m should decrease, a higher r (i.e., a larger shaded area) reflects a higher level of conservatism. A higher r also implies a higher probability of ruin (i.e., it is easier for actual claim liabilities to fall in the shaded region). A higher r also implies that it is easier to conclude that the alternative estimate of claim liabilities is different from the original estimate underlying the reserve.

Other factors that should be considered in selecting the thresholds that define the materiality standards may be the following:

- type of exposures involved
- primary / reinsurance limits
- size of reserves / expected loss / no. of exposures or claims
- average age of reserves
- expectation of parameter risk associated with the particular LOB
- probable maximum loss
- asset variability
- net income variability

In our study, we have not specifically analyzed the impact of these issues in the calculation of the standards of materiality. Generally, the impact of the above factors on the standards will depend on whether they add to or decrease the volatility of the claim liabilities, or increase or decrease the uncertainty associated with the financial and solvency status of the company.

Both of the stochastic reserving models employed in our analysis measure process and parameter risk but neither of them measures explicitly model risk. Further research is needed in the area of the measurement of model risk.

2.4 Acknowledgements

The authors would like to thank Rick Shaw, Amy Bouska, Tom Ghezzi, Jeanne Hollister, Ron Kozlowski, Kahshin Leow and Ollie Sherman for their thoughtful insights on the topic.

3. METHODOLOGY

3.1 Overview

The overall approach is as follows:

Step 1 - Obtain sample balance sheet and historical claim development data for selected companies and lines of business.

Step 2 - Apply stochastic methods to the historical claim development data to measure the distribution of the actual claim liabilities, and the distribution of the estimated expected claim liabilities for each company and each line of business.

Step 3 - Select estimation and outcome thresholds. For outcome materiality thresholds, base selections on typical balance sheet solvency impacts for selected companies.

Step 4 - Develop ranges embodying the materiality standards, based on both percentile thresholds and expected exceedence ratio thresholds.

Step 5 - Recognize risk diversification benefits among multiple lines by incorporating correlation and aggregating the individual line of business distributions to build an aggregate distribution to arrive at ranges embodying the overall materiality standards at a legal entity level.

The following sections will elaborate on each step.

3.2 Step 1 – Data and Data Limitations

U.S. insurers are required to file Annual Statements with state regulatory authorities. The required format includes income statements, balance sheets, cash flows and schedules focusing on aspects such as historical claim development (Schedule P), reinsurance recoverables (Schedule F) and investment (Schedule D). As noted previously, a Statement of Actuarial Opinion must accompany each Annual Statement. Annual Statements and Statements of Actuarial Opinion are in the public domain and can be viewed at each state's

Department of Insurance. We used an internal Annual Statement database, based on data obtained annually from A.M. Best.

We used data from various sections of the Annual Statement. The claim liability development triangles and premiums were obtained from the Schedule P for each company. Measurement of capital came from the "Five-Year Historical Data" exhibit.

We analyzed four lines of business: Personal Auto Liability, Homeowners, Workers Compensation and Other Liability-Occurrence policy forms. These lines were selected to reflect the spectrum from short-tail to long-tail, and the spread in volatility.

Within the U.S. non-life insurance sector, it is common for an insurer to operate through multiple legal entities under common management, often referred to as a group. Multiple entities within a group offer flexibility in terms of capitalization, pricing and regulatory domain. An insurer must file an Annual Statement and a Statement of Actuarial Opinion for each legal entity. Therefore our analysis is done at the legal entity, not group level.

There are often inter-company pooling arrangements whereby an insurer allocates results to entities which may or may not have written the business. The pooling percentages may vary by line and year. The pooling applies to each aspect of the Annual Statement, including the Schedule P data triangles we use, that is, the analyzed triangles may represent a percentage of a larger triangle. Therefore, when we consider the relative size of the sample entities, we need to adjust for pooling. All figures presented are adjusted to reflect the effect of pooling.

We included insurers that cover the spectrum from small single-state or regional to large national companies. For our purposes, we define size in relation to the premium earned from 1994-2003 for each line. Companies with premium below $3 billion are considered small for that line, companies above $10 billion are defined as big, and the rest are medium. For example, a large national writer such as Hartford Financial includes a legal entity, Harford Fire Insurance Company, which we consider small, medium and large depending on the line of business under consideration. (See the table below; figures in $000's)

Line of Business	Net Earned Premium for Line	Size Categorization for Line
Workers Compensation	$ 11,818,872	Big
Personal Auto Liability	$ 10,463,646	Big
Homeowners	$ 4,913,186	Medium
Other Liability	$ 2,914,352	Small

The segmentation by size is intended to group companies with expected similar reserve volatility. All things being equal, we expect claim liabilities on larger volumes of business to be less volatile than smaller books of business.

We took the data "as is," meaning that extensive cleansing of the data was not undertaken. In several instances we adjusted anomalous data, with care not to sanitize the data. Even with these adjustments, some of the data appears to be implausible; companies with implausible data were excluded from our analysis.

While we restricted our analysis to publicly available Annual Statement data, it should be noted that insurers have additional information available internally. Companies often segment their business into more homogenous groups than Annual Statement line of business. The concepts applied here on a line of business basis are illustrative and can also be utilized for different segmentation.

3.3 Step 2 – Use stochastic methods to measure volatility of unpaid claim liabilities

We used the Bootstrapping methodology as described by England and Verall [6] and the Mack Stochastic methodology [12] to estimate the volatility of claim liabilities. For a brief description of these methods please refer to Appendix A. The CV (Coefficient of Variation) is our chosen measure of volatility. The Mack method generates the first two moments of the claim liability distribution, the mean and the standard deviation, while the Bootstrapping method produces an empirical distribution of claim liabilities, so CVs are easily calculated in both cases. The input historical claim development triangles used for both methods are paid loss development triangles (including only allocated loss adjustment expense). In addition, we augmented the Bootstrapping method described in the paper to recognize development beyond the maturity of the triangle (i.e., in the tail). We have assumed a tail that extends to

10-12 years for Homeowners and Personal Auto Liability, 15-20 years for Occurrence Liability and 40-50 years for Workers Compensation. The tail was estimated by fitting an inverse power curve to the development factors for ages of 48 months and beyond, based on Richard E. Sherman's [14] approach as outlined in "Extrapolating, Smoothing, and Interpolating Development Factors." A uniform tail was selected to apply to all accident years within a company. Additionally, for the Mack method we selected standard errors associated with the tail volatility. The selection was essentially based on the empirical results of the Bootstrapping method. We compared the CVs produced by the Bootstrapping method for each line of business in our sample database with the inclusion of a tail factor and exclusion of the tail factor. The difference in the CVs including and excluding the tail factor was then selected as a measure of the standard error associated with the tail factor.

As stated above, the Bootstrapping method provides more than just the mean and variance of the claim liability distribution; it generates the entire distribution. In almost all cases the mean of the distribution generated from the Mack and Bootstrapping methods was different from the carried reserve amount, therefore we performed a linear transformation to force the mean of the distribution to be equal to the carried reserves, while preserving the CV of the distribution. When we describe the "percentile/carried reserve," we are assuming the carried reserve is the best estimate. This is an assumption, not an assertion. Readers are directed to "Management's Best Estimates of Loss Reserves" [10] by Rodney Kreps that notes the mean of the distribution is "probably not a good estimate, as it is almost surely low."

We note that the use of paid claim development data in our analysis is essentially dictated by the inherent limitation of the Bootstrapping and Mack stochastic reserving methodologies used in our analysis. These methodologies do not respond well to reported loss (case reserves + paid losses) data. Indeed both methods produce unreasonable results when used on reported loss triangles which occasionally have age-to-age loss development factors below 1.0 followed by positive development (age-to-age development factors above 1.0). Both of these methods require a somewhat smooth progression of age-to-age loss development factors from immature to mature valuations, declining from high loss development factors for immature data to low development factors for mature data.

Another limitation is that both methodologies assume an essentially stationary process, i.e., that there are no endogenous and exogenous influences on the loss generating process such as company-specific changes in operations, claim settlement rates, premium/exposure growth or changes, large settlements, evolving interpretations of liabilities in the court system, hurricanes, and so forth. Realistically, the loss development data reflected in the triangles of a company are hardly ever stationary, as they include both exogenous and endogenous influences, which cause additional volatility in the loss development triangle. As a result, stochastic reserving methodologies that rely on the volatility inherent in the loss triangle almost always overstate the volatility of the underlying loss generating process. In order to adjust for this distortion, we adjusted the volatility estimates arrived from the use of these stochastic reserving methodologies downward. The adjustment factors were calculated using industry-wide paid loss triangles from A.M. Best (27 to 30 company composite, depending on the line of business), adjusting the triangles for industry exposure changes and frequency trend and other exogenous influences. The frequency trend is applied to adjust for observed declines in claim frequency due to safer workplaces, safer cars, and so forth over the years. We then postulate that this process should create stationary triangles, absent of any exogenous and endogenous factors described above. Thus the volatility present in these stationary triangles will be the true volatility produced by the loss generating process. We measured this volatility in the industry-wide triangles, by line of business, and used it to adjust downwards our overall volatility results produced by the stochastic methods employed in this paper. While we believe these adjustments are reasonable, refinements in the techniques used to better achieve the desired stationarity can certainly improve upon them.

We also independently tested the assumption that the process risk implied by the stochastic methods employed in our analysis is overstated, by comparing the claim volatility calculated by the stochastic methods to the claim volatility obtained via hindcast testing. We employed an independent historical data set for 20 companies and measured the performance of deterministic reserving techniques as they tried to estimate the claim liabilities for these companies. We first estimated the claim liabilities using information that was available at a given point in time and then looked at the available run-off information to see what the actual claim liabilities amounts were with hindsight. The observed estimation error over time and across all companies provides a proxy for the total risk associated with the evaluation of claim liabilities. For the workers compensation line, the hindcast tests results indicate a CV of total risk equal to 8.1%. By comparison, we obtained a parameter-only risk CV of 11.0% from the Mack method. Most of the companies in the hind-cast testing were rather large, with reserves in excess of $100 million, so the associated process risk for these companies should be quite low. The fact that the total risk CV from the hind-cast testing is lower when compared to the parameter risk CV from the Mack method suggests that the process risk and parameter risk implied by our stochastic methods could be overstated. The inability of traditional stochastic reserving methods to separate the variability due to changes in endogenous and exogenous influences, from the true claim volatility due to the claim generating process, is the main reason for this presumed overstatement of process and parameter risk.

3.4 Step 3 – Select Significance Threshold Levels

For outcome materiality, we calculated threshold significance levels for both financially healthy companies and financially impaired ones. Financially impaired companies should get an earlier warning flag when something is wrong with their reserves compared to financially healthy companies, since the underlying assumption is that an adverse claim liability deviation causes a greater financial "hurt" to financially impaired companies.

We employed the "bright line test," which we understand is utilized by the NAIC, in the measurement of outcome benchmark significance levels. The bright line test measures the difference between the surplus as regards to the policyholders and the RBC (Risk Based Capital) capital amount, proposed by the NAIC, that would downgrade the company to the next lower RBC level. If the claim liabilities of a company sustain an adverse deviation greater or equal to the capital level difference mentioned above, that company would be downgraded to the next lower RBC level. That capital level difference to the next lower RBC level, given a distribution around carried reserves, provides a maximum standard of materiality for the company (i.e., the officers of that company would, at least, want to know under what adverse claim liability deviation the company would be downgraded to the next lower RBC level). They might want though to set up an earlier warning flag, based on their experience with the company's financial results, so the adverse deviation from the bright line test can serve, at least, as a maximum standard of materiality.

An assumption in the above analysis is that these companies did not experience significant changes in their distribution of exposures, by line, during the historical period of the analysis. The implied volatility from the claim liabilities for each company was calculated on an all lines combined basis considering both process and parameter risk. An outcome materiality significance level threshold was calculated for an upper tail test within the percentile threshold context and an exceedence ratio threshold was calculated within the

TVar context, given a level of volatility associated with the carried reserve. The outcome materiality significance level threshold for a lower tail test for the percentile threshold approach was calculated judgmentally based on the assumption that the magnitude of the standards of materiality should be higher for an upper tail test when compared to a lower tail test.

For the corresponding estimation materiality significance level threshold we employed a 7.5% rule of thumb benchmark for the upper tail test. That 7.5% represents the average of the 5% to 10% significance levels usually employed in statistical hypothesis testing. Interestingly, we tested the validity of this assumption (7.5%) by estimating the benchmark significance level threshold using parameter risk only from the outputs of the Mack method and found that the resulting estimation materiality benchmark significance level threshold was, on average, similar to the 7.5% that we assumed.

The resulting thresholds for financially healthy companies were as follows:

	Percentile Threshold **Benchmark Significance Levels**		**Tail Value at Risk** **Benchmark Exceedence Ratio**	
	Lower Tail Probability	**Upper Tail Probability**	**Lower Tail Expected Excess**	**Upper Tail Expected Excess**
Estimation materiality	10.0%	7.5%	n/a	2.0%
Outcome materiality	8.0%	6.0%	n/a	1.5%

All other things being equal, the resulting outcome materiality standards are higher from the corresponding estimation materiality standards, a logical relationship when considering the higher amount of risk associated with outcome materiality standards. For the majority of the healthy companies the resulting outcome materiality benchmark significance level, for the upper tail test, was 0.0%. This result highlights the fact that most of the healthy companies are so well capitalized that they need to suffer an adverse claim liability deviation in excess of the 99.9% percentile of their claim liability distribution in order to get downgraded into the next lower RBC level.

We also performed the above analysis on a group of 16 companies that were financially impaired. These companies were either in rehabilitation or liquidation. For arriving at the outcome materiality benchmark significance levels for the financially impaired companies we used the bright line test as well as one year adverse development from Schedule P data. As expected, we estimated much higher benchmark significance levels and benchmark exceedence ratios compared to the healthy company figures mentioned above. The outcome materiality benchmark significance levels for adverse deviation for the financially impaired companies were 18% compared to 6% for financially healthy companies.

Specifically, we performed the following steps to come up with the outcome materiality benchmark significance levels and benchmark exceedence ratios.

We employed 39 financially healthy companies from our A.M. Best data base. The calculation of an upper tail test outcome materiality benchmark significance level threshold, for the Percentile Threshold approach, followed the steps outlined below:

1) For each of these companies we measured the total risk, for all lines combined, associated with their claim liability distribution. That claim liability distribution was calculated from loss and ALAE Schedule P Part 3 triangle data using the Mack stochastic reserving method. We further assumed that the mean C_α of the stochastic distributions is equal to the carried reserves for the companies.

2) From the Bright Line Test we calculated the adverse claim liability deviation that would downgrade each company into the next lower RBC level. That adverse claim liability deviation m represents a maximum standard of materiality.

3) We then added the mean of the distribution to the adverse claim liability deviation. The area under the claim liability distribution in excess of $C_\alpha + m$ represents the upper tail outcome materiality benchmark significance level. It measures the probability of extreme claim liability outcomes that a company must experience before it gets downgraded into the next lower RBC level.

For the calculation of an outcome materiality exceedence ratio for the TVar approach the first two steps outlined above were identically repeated. As a last step we measured the average of claim liability outcomes that exceed the carried reserves by the standard of materiality. These excess losses were calculated as a ratio to the expected claim liabilities, producing the exceedence ratio threshold.

3.5 Step 4 – Estimate Materiality Standards for each Individual Line

Based on the selected outcome materiality benchmark significance levels and exceedence ratios we then calculated the outcome materiality standards for each company in our sample database. The calculation proceeds as follows:

1) For each company triangle we generate a claim liability distribution using both the Bootstrapping and the Mack method.
2) We normalize each loss reserve distribution so that the mean of the distribution is equal to the carried reserve of the company.
3) The outcome materiality standard is equal to the difference between the percentile implied by the outcome materiality benchmark significance level, as described above, and the percentile implied by the carried reserve.
4) The outcome materiality standard implied by the TVar approach is calculated as the difference between the percentile implied by the benchmark exceedence ratio and the percentile implied by the carried reserve.
5) The estimation materiality standards are calculated in a similar fashion using the estimation materiality benchmark significance levels and exceedence ratios.

3.6 Step 5 – Recognize Risk Diversification Benefits Among Multiple Lines

Few companies are monoline writers. For multi-line writers, the standards of materiality should incorporate the risk diversification associated with underwriting more than one lines of business. Aggregate claim liability distributions can be calculated from the individual line distributions. In our analysis we incorporate a Copula type of approach that performs the aggregation procedure. More information regarding the Copula approach is included in Appendix C.

The Mack and Bootstrapping stochastic reserving methods mentioned above measure the claim volatility for an individual line of business. In case where more than one lines of business are considered we need a model that aggregates the individual lines distributions. The mean of the aggregate distribution is the sum of the individual lines means. However we cannot arrive at the percentiles of the aggregate distribution by simply adding the individual lines percentiles. Straight summation makes sense only in the case of 100% correlation across all lines, a highly unlikely situation. The volatility of the aggregate distribution is influenced by two factors:

- The claim volatility for each individual line of business: The larger the claim volatility for each individual line, the larger the volatility of the aggregate distribution, all other things being equal; and
- The correlations across lines: The larger the correlation among individual lines, the larger the volatility of the aggregate distribution, all other things been equal.

Statisticians have shown that the aggregate distribution of any combination of n random variables can be written as a function of the n individual variables distributions (*Sklar theorem 1996*). This function is called Copula. We are employing one Copula model in our analysis that provides a convenient way of calculating the aggregate distribution of several lines of business. Two components are needed for the Copula model:

- The distributions of the individual lines of business; and
- The correlation coefficients among these lines.

The Copula model employed in our analysis is the Normal Copula. For the Normal Copula a correlation matrix based on the assumed correlations among the various lines must be selected. The correlation matrix for the Normal Copula should be positive-definite (i.e., invertible) for the Copula to work.

The selected correlation among the various lines is based on modeling of economic variables such as general/price inflation, wage inflation, auto inflation, and medical inflation. This is done by first building forecasting models for auto inflation and medical inflation as a function of general/price inflation. The models have an autoregressive component in that the inflationary component being modeled reverts back to it long term mean. Next we modeled the impact of each of these inflationary components on each line of business. The model used was a geometric model employed by Robert P. Butsic [3] to model the impact of different inflationary components on losses of different lines of business including social inflation. Once the impact of these inflationary components on each line of business is ascertained then we can construct a distribution of losses for each line of business by forecasting these economic variables. The correlation matrix is then estimated by empirically measuring the correlation between the simulated losses for each line of business. Our assumed correlation matrix is included in Appendix C. The advantage of such models is that correlation between the claims experience is an emergent property.

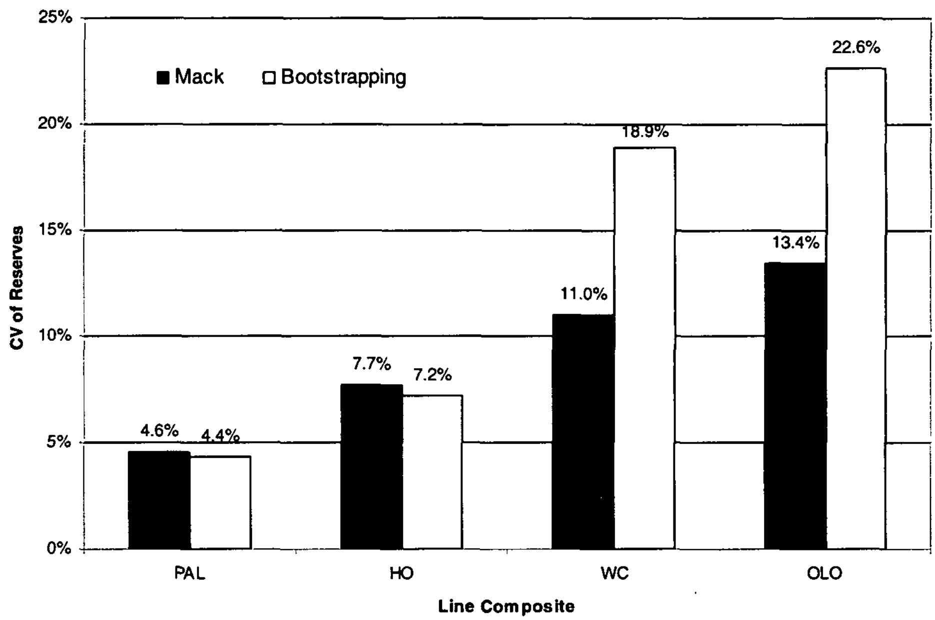
4. RESULTS and CONCLUSIONS

4.1 Reserve Volatility

As can be seen in Exhibit 4.1.1, the volatility is relatively small for Personal Auto Liability (PAL), somewhat larger for Homeowners (HO), larger still for Workers Compensation (WC), and even larger for Other Liability Occurrence (OLO.) The relative magnitude is as expected. The HO line is impacted by catastrophes and the HO claim liabilities are more volatile when compared to the OLO liabilities. OLO is impacted by some high severity claims so intuitively is more volatile. WC also has high severity claims but there is enough frequency/consistency that overall it is less volatile than OLO.

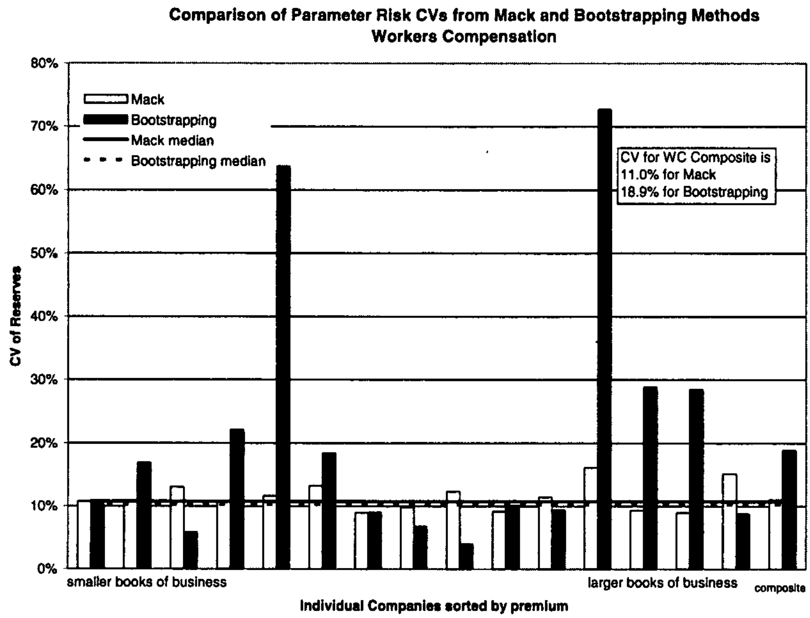
Exhibit 4.1.1

Comparison of Parameter Risk CVs from Mack and Bootstrapping Methods

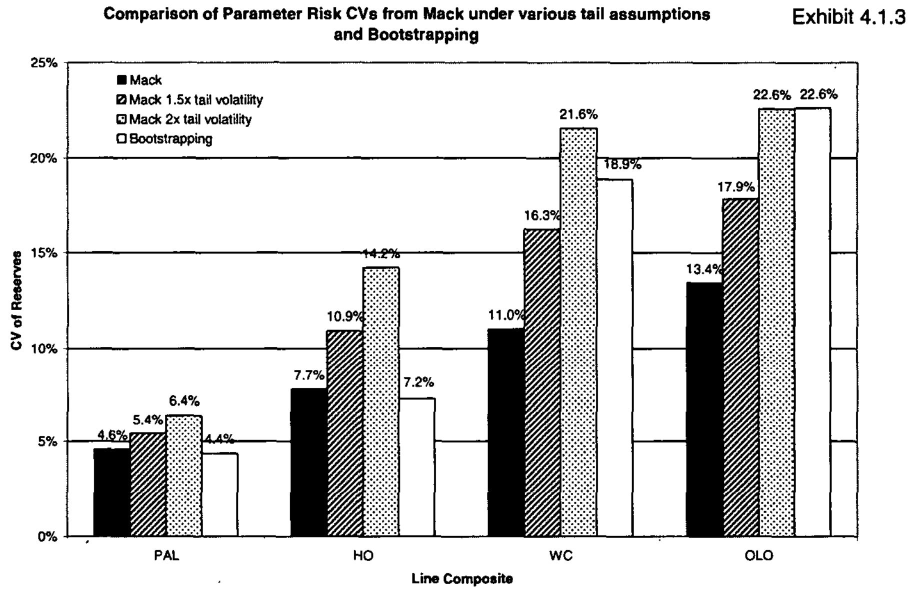


The Bootstrapping method is more sensitive to outlier development factors and so it generates significantly larger CVs for some companies, as displayed in Exhibit 4.1.2.

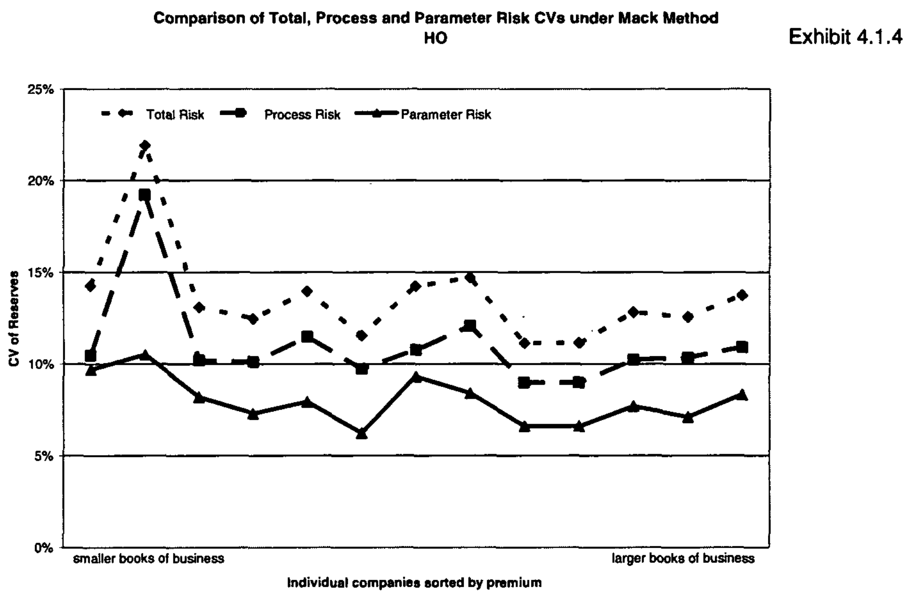
Exhibit 4.1.2



As discussed in Section 3.3, the Mack method is dependent on the assumed volatility in the tail. We tested our tail volatility assumption to determine how sensitive the analysis is to our supposition. We increased the volatility in the tail by 50% and 100%. For Workers Compensation and Other Liability Occurrence, the increased tail volatility drove the Mack CVs closer to the Bootstrapping CVs. On the other hand, the adjustment created a difference for Personal Auto and Homeowners, whereas the CVs were quite similar before increasing the tail volatility. These results are displayed in Exhibit 4.1.3 (see next page)



As shown in the Exhibit 4.1.4, total, parameter and process risk all generally follow the same

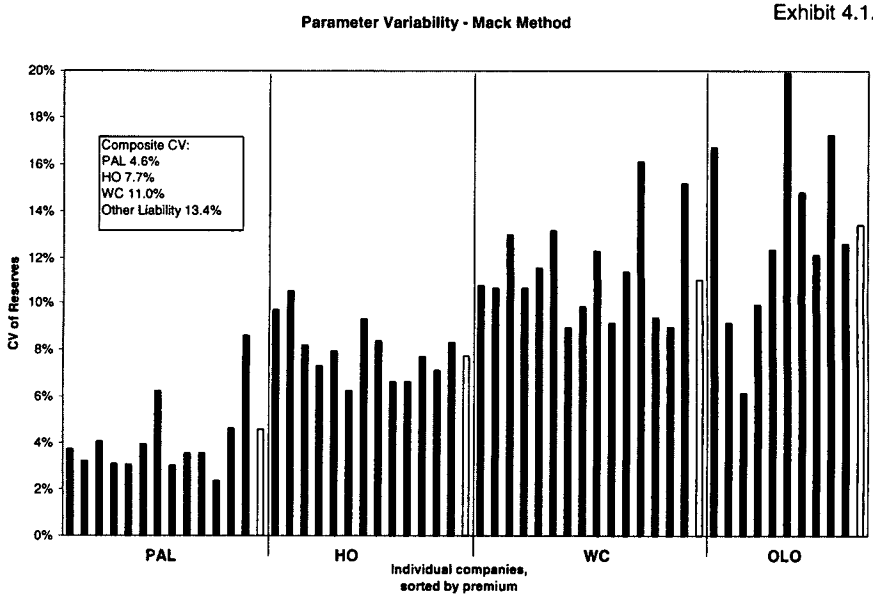


relationship: process is usually larger than parameter risk, and naturally, total risk is the largest.

We expected that parameter risk is invariant of size, while process risk should decrease by the size of the company. Our analysis calculates process risk that is independent of size. That result suggests that the stochastic methods employed in our analysis could possibly overstate process risk.

We summarize the results by size, expecting larger books of business to be less volatile, however this was not the case. For Personal Auto Liability, the most volatile companies were generally the larger ones. There was not much variation in the size of selected Workers Compensation companies with two-thirds of them categorized as small. The results were mixed with high CVs coming from both small and large companies. Each bar in Exhibit 4.1.5 represents a company and is sorted from by premium volume, with smaller companies on the left.

Exhibit 4.1.5



An analysis by size of company that employs a larger sample of companies is probably needed to draw more credible conclusions in the comparison of claim volatility by size of company.

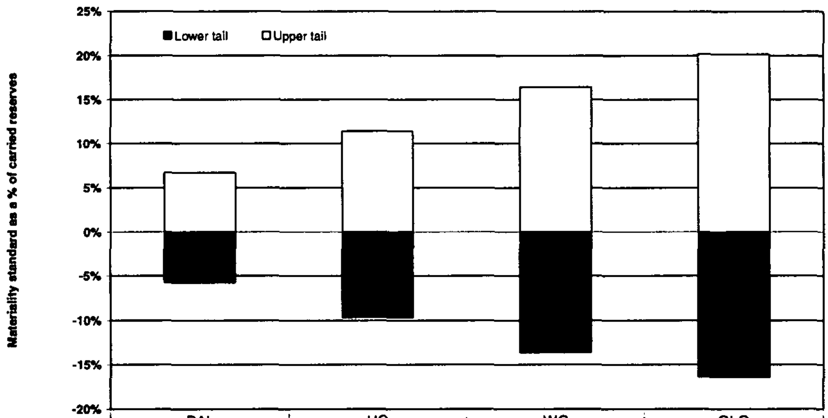
4.2 Materiality Standards

The following graphic summarizes the estimation materiality standards, based on the Mack method, for the four lines of business under consideration. All the standards shown in the remainder of the paper were calculated, unless otherwise noted, as a percentage of carried reserves and using the Percentile Threshold approach.

The resulting estimation materiality standards are higher than what actuaries are accustomed to, partly because these techniques overstate volatility unless adjustments are made for exogenous and endogenous factors.

Exhibit 4.2.1

Estimation Materiality Standards based on the Mack Method



	PAL	HO	WC	OLO
Upper tail	6.7%	11.4%	16.4%	20.2%
Lower tail	-5.8%	-9.7%	-13.6%	-16.4%

For Personal Auto the standards are close to the +/-5% range of expected that is often employed, as a rule of thumb. For the remainder of the lines the resulting standards are much higher. The calculated standards of materiality could be overstated due to the suspected overstatement of the process and parameter risk produced by the Mack method. Exhibit 4.2.1 (see previous page) graphs the upper and lower tail estimation materiality standards by company and by line. For the more volatile lines the wider range of the materiality standards is evident.

The following table compares the estimation materiality standards produced by the Mack and Bootstrapping models. The actual upper and lower tail estimation materiality standards are calculated as follows:

Estimation Standards of Materiality – Bootstrapping vs. Mack

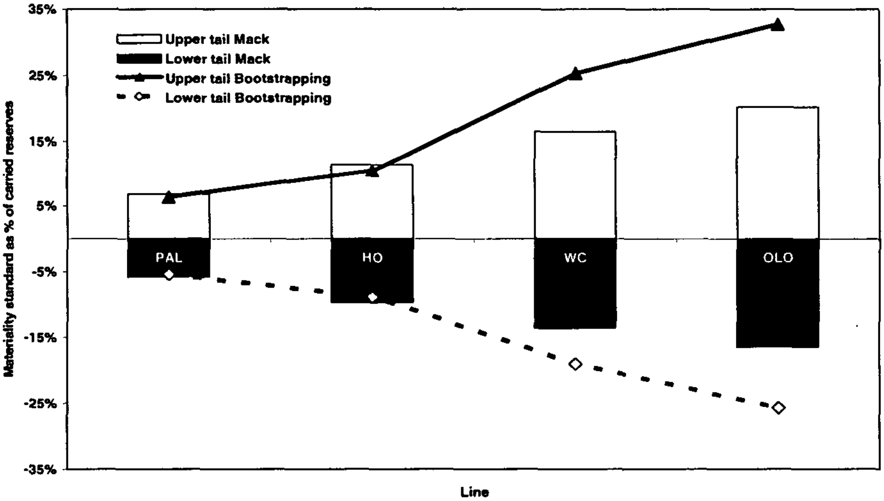
	Mack		Bootstrapping	
Line of Business	Lower Tail Test	Upper Tail Test	Lower Tail Test	Upper Tail Test
Personal Auto Liability	-5.8%	6.7%	-5.4%	6.3%
Homeowners	-9.7%	11.4%	-8.8%	10.5%
Workers Compensation	-13.6%	16.4%	-19.0%	25.3%
Other Liability	-16.4%	20.2%	-25.7%	32.7%

The resulting estimation materiality standards between the two methods are relatively close for the Personal Auto and Homeowners lines of business. For the two long-tail lines, workers compensation and other liability, the Bootstrapping statistical standards are 40% to 60% higher when compared to the Mack standards.

The additional tail volatility implied by the Bootstrapping method produces the higher estimation materiality standards. Exhibit 4.2.2 compares the estimation materiality standards for the two stochastic methods employed in the analysis.

Exhibit 4.2.2

Comparison of Estimation Materiality Standard - Mack vs. Bootstrapping



The comparison of estimation and outcome standards of materiality is summarized in the following table, for the Mack method, for both lower and upper tail tests:

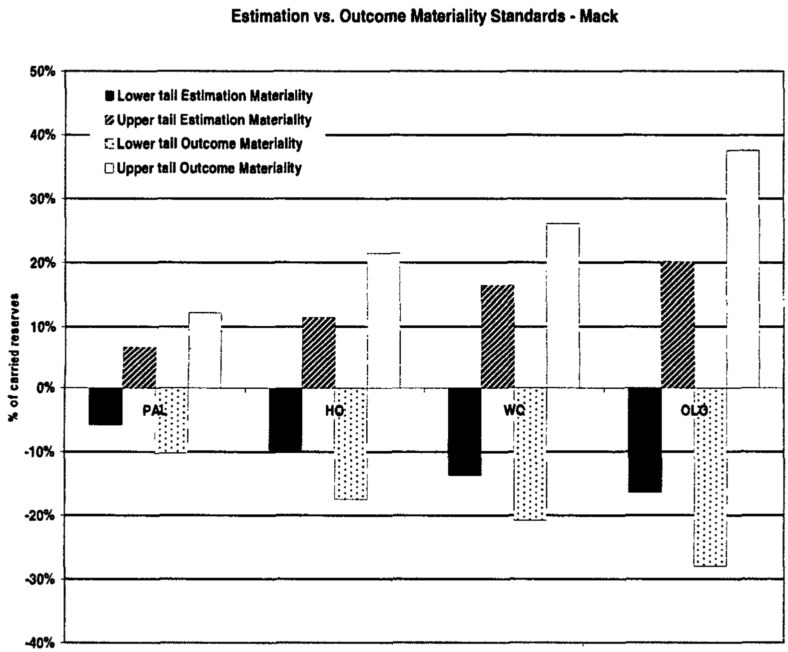
Estimation vs. Outcome Materiality Standards– Mack

	Estimation standards		Outcome standards	
Line of Business	Lower Tail Test	Upper Tail Test	Lower Tail Test	Upper Tail Test
Personal Auto Liability	-5.8%	6.7%	-10.2%	12.2%
Homeowners	-9.7%	11.4%	-17.5%	21.5%
Workers Compensation	-13.6%	16.4%	-20.8%	26.2%
Other Liability	-16.4%	20.2%	-28.0%	37.7%

The outcome materiality standards are, on average, 75% higher when compared to the estimation materiality standards. There are two reasons that explain this relationship: (1)

outcome materiality standards employ the calculation of both process and parameter risk while estimation materiality standards employ parameter risk only. The inclusion of process risk increases the outcome materiality standards; and (2) the benchmark significance level is higher for the estimation materiality standards when compared to the benchmark significance level for the outcome materiality standards. All other things been equal, the resulting outcome materiality standards should be higher since the corresponding probability of Type I error is lower. Exhibit 4.2.3 provides a comparison of the outcome and estimation materiality standards.

Exhibit 4.2.3



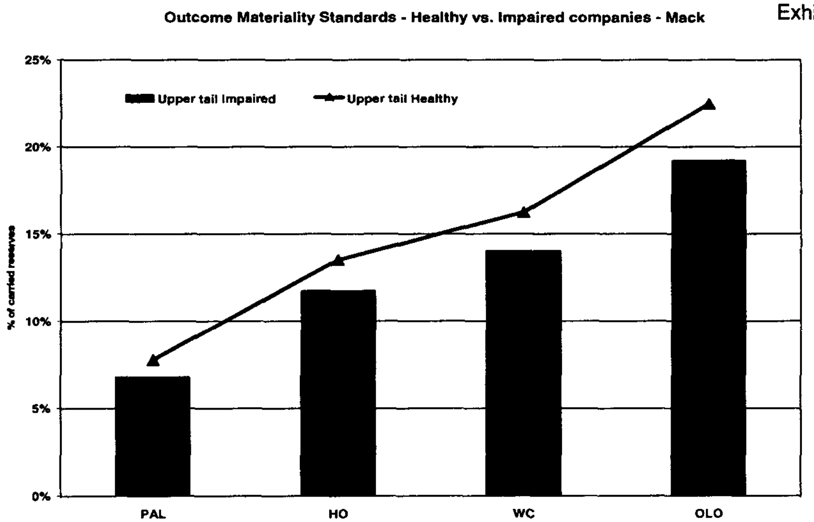
The following table compares the upper tail outcome materiality standards for financially healthy and financially impaired companies.

Outcome Materiality Standards– Upper Tail
Healthy vs. Liquidated Companies – Mack

Line of Business	Financially Healthy	Financially Impaired
Personal Auto Liability	12.2%	6.8%
Homeowners	21.5%	11.7%
Workers Compensation	26.2%	14.0%
Other Liability	37.7%	19.2%

The outcome materiality standards are much higher for the financially healthy companies when compared to the corresponding standards for the financially impaired companies. For a financially impaired company, a lower outcome materiality standard is reasonable since it provides an earlier warning flag if an adverse claim liability deviation is experienced by that company. The lower standards compensate for the greater reserve uncertainty associated with the reserves of a financially impaired company coupled by lower reserve to surplus ratios. Moreover, our selected significance level benchmarks of 18% for financially impaired companies vs. 6% for financially healthy ones, allows for a greater probability of Type I error

Exhibit 4.2.4



for the financially impaired companies, decreasing in effect their respective outcome materiality standards. Exhibit 4.2.4 (see previous page) compares the upper tail outcome materiality standards for financially healthy and financially impaired companies.

The following table compares the upper tail estimation materiality standards for the two risk measures employed in our analysis, the Percentile Threshold approach and the TVar approach.

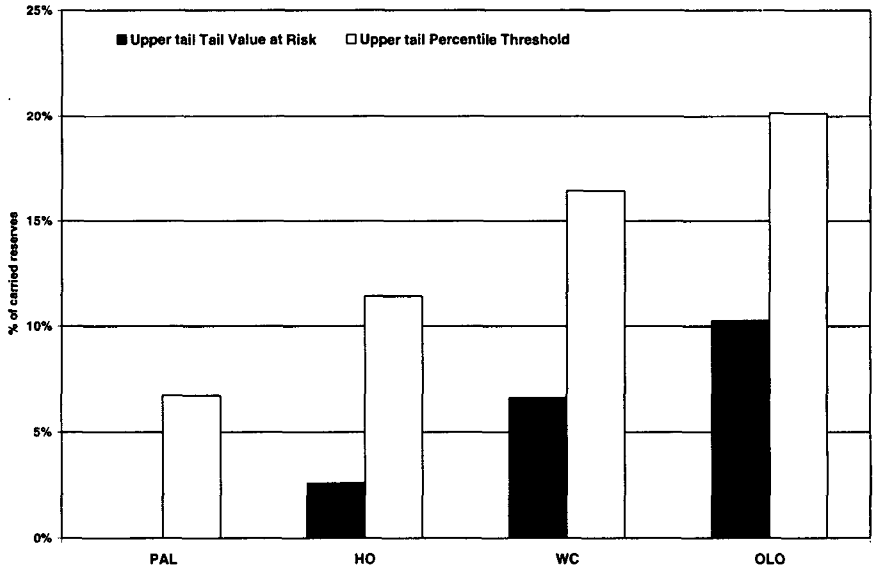
Estimation Materiality Standards – Mack

Line of Business	Percentile Threshold	Tail Value at Risk
Personal Auto Liability	6.7%	0.0%
Homeowners	11.4%	2.6%
Workers Compensation	16.4%	6.6%
Other Liability	20.2%	10.3%

The standards implied by the TVar approach are considerably lower when compared to the standards produced by the Percentile Threshold approach. The reason of that observed difference lies on the varying fundamental assumptions of the two risk measures. The Percentile Threshold approach measures the probability that the actual claim liability amount would exceed a selected dollar threshold. It does not consider the magnitude of the resulting deficiency. A $1 reserve deficiency gets the same weight as a $1 million reserve deficiency under the Percentile Threshold approach. On the other hand, the TVar approach measures the expected risk of material adverse deviation. The higher the risk of material adverse deviation, the higher measure of risk is calculated by the TVar approach. In other words, the TVar approach penalizes a company for the probability of extreme claim liability outcomes. Since for most of the property and casualty (general non-life) insurance companies there is a small chance of very large claim liability outcomes, the TVar approach, on average, assigns more reserve risk to the companies when compared to the Percentile Threshold approach. The higher risk associated with the TVar approach results in lower standards of materiality since an earlier warning flag is more appropriate in the presence of more reserving risk. Exhibit 4.2.5 compares the upper tail estimation materiality standards for the two measures of risk employed in our analysis.

Exhibit 4.2.5

Upper Estimation Materiality Standards



We also created a fictitious company that writes the four lines of business under consideration with a reserve distribution approximating the distribution of the whole industry. Employing a Normal Copula approach we calculated the CVs of the claim liabilities for the company. The resulting total risk CV is 11.0% while the parameter risk CV is 10.2%. The risk diversification associated with the underwriting of four, instead of one, lines of business results into combined CVs that are lower when compared to the CVs from the two long tail lines of business (workers compensation and other liability). Exhibit 4.2.6 (see next page) compares the resulting aggregate CVs from the four monoline writers to the CVs of the fictitious multi-line company.

Exhibit 4.2.6

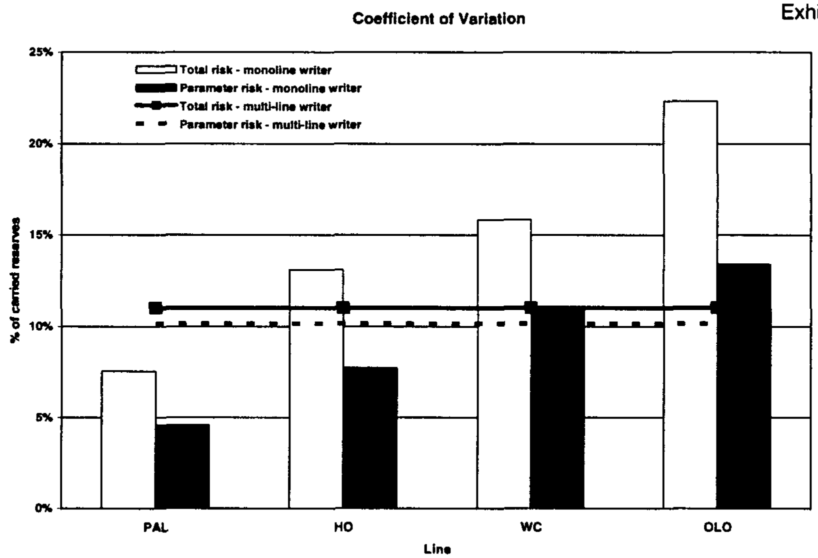
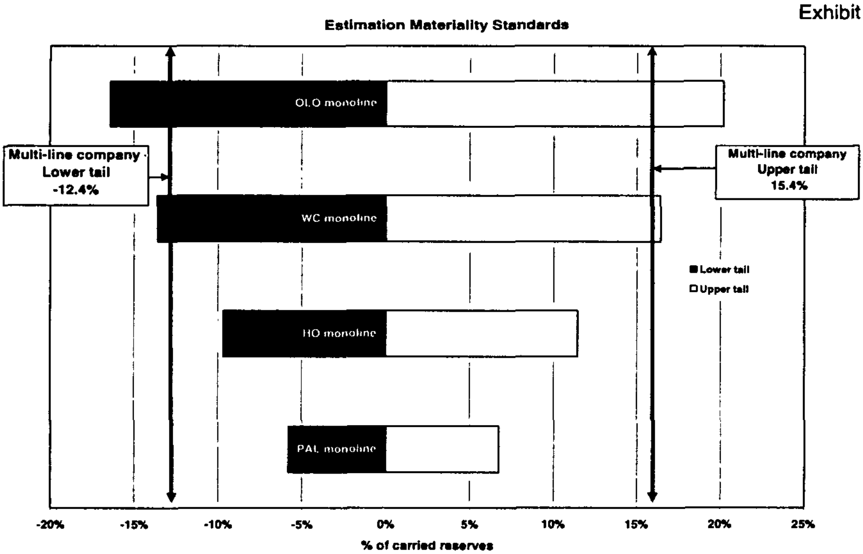


Exhibit 4.2.7 compares the estimation standards of materiality of the fictitious four lines writer to the four monoline writers. The resulting upper tail statistical standard is 15.4% while the lower tail statistical standard is 12.4%. These standards are affected by the higher weight given to the long tail lines (30% for Personal Auto Liability, 6% for Homeowners, 35% for Workers Compensation and 29% for Other Liability Occurrence.)

Exhibit 4.2.7



The following table compares the outcome and estimation materiality standards for a writer of the four lines of business.

Standards of Materiality – Mack

Type	Lower Tail	Upper Tail
Estimation materiality standards	-12.4%	15.4%
Outcome materiality standards	-14.4%	18.1%

The outcome materiality standards are, on average, 15% higher when compared to the estimation materiality standards. This relationship is reasonable in light of process risk which is considered in the outcome materiality standards but not in the estimation materiality standards.

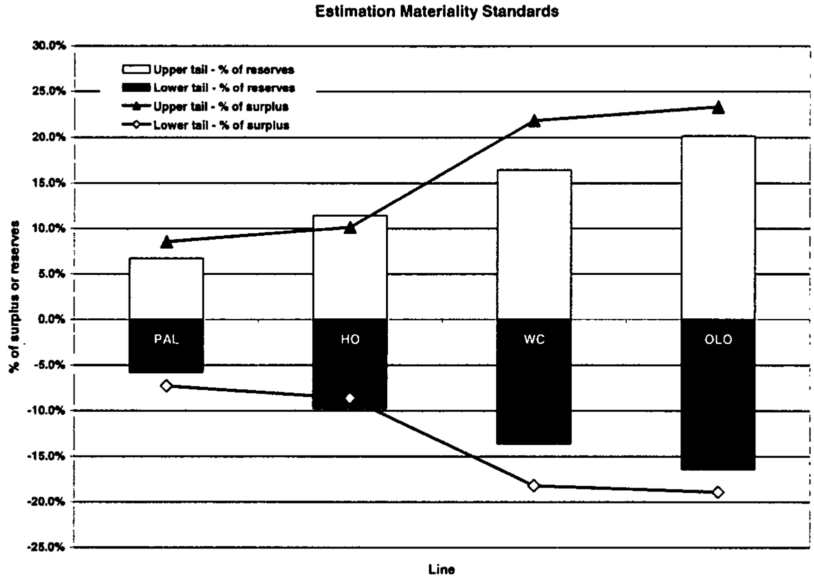
The following table summarizes the estimation materiality standards, for the four lines of business under consideration, as a percentage of individual company surplus.

Estimation Materiality Standards – Mack
(as a % of surplus)

Line of Business	Lower Tail	Upper Tail
Personal Auto Liability	-7.3%	8.5%
Homeowners	-8.6%	10.1%
Workers Compensation	-18.2%	21.9%
Other Liability	-18.9%	23.4%

The resulting percentages for the upper tail test are in the area of 10% for short tail lines and in the area of 20% for long tail lines. Exhibit 4.2.8 (see next page) compares the estimation materiality standards, as a percentage of both surplus and carried reserves, for each line of business analyzed.

Exhibit 4.2.8



The following table compares the implied volatility for each line of business analyzed, measured by the coefficient of variation, to the resulting estimation materiality standards for the upper tail test in the Percentile Threshold approach.

Comparison of Parameter Risk CVs and Estimation Materiality Standards – Mack

Line of business	CV	Upper tail estimation materiality standards
Personal Auto Liability	4.6%	6.7%
Homeowners	7.7%	11.4%
Workers Compensation	11.0%	16.4%
Other Liability	13.4%	20.2%

The standards of materiality increase for the more volatile lines. The uncertainty associated with the calculation of the claim liabilities for a volatile line is quite high, and the large associated standards of materiality reflect that uncertainty. All other things being equal, two independent actuarial estimates that measure volatile claim liabilities should be given the benefit of the underlying uncertainty before considered materially different from one another. Another way to intuitively think about this result is that lines of business, or books of business, which show a high level of volatility usually have a higher percentage of total surplus allocated to them and thus have a higher cushion to absorb adverse deviation. That results in a higher standard of materiality as a percentage of reserves.

As we pointed earlier, the results quoted above might be overstated as the stochastic methods employed in this paper presumably overstate process and parameter risk. Thus both the CVs and the materiality standards derived above are overstated. We performed the adjustments to reduce the overstatement, described earlier in this paper, on three companies for three different lines of business to get an approximate impact of the overstatement of volatility by the stochastic methods. The impact of the overstatement of the CV was calculated by subtracting the CV obtained from the true volatility of the adjusted industry triangle from the CV of the unadjusted individual company triangle. Using the benchmark significance levels we calculated the adjusted standards of materiality. This procedure was performed separately for each line of business. The results are presented in the following tables:

Outcome Materiality Standards– Mack, Upper Tail or Adverse Deviation

Line of Business	Before Adjustment	After Adjustment
Personal Auto Liability	12.2%	5.7%
Workers Compensation	26.2%	18.0%
Other Liability	37.7%	16.7%

Estimation Materiality Standards– Mack, Upper Tail or Adverse Deviation

Line of Business	Before Adjustment	After Adjustment
Personal Auto Liability	6.7%	3.6%
Workers Compensation	16.4%	12.5%
Other Liability	20.2%	11.5%

As these tables show, the impact of this overstatement can be significant. To make a thorough assessment of the impact of the adjustment is beyond the scope of this paper.

APPENDICES

A. Technical Appendix – Stochastic Methods Employed

The deterministic methods provide a best estimate of the claim liabilities. In comparison, stochastic methods provide a claim liability distribution around the best estimate, in addition to the best estimate. We employed two stochastic methods in our analysis. Each of these methods represents the two families of stochastic methods described below:

"Chain Ladder" family of methods. These methods employ cumulative loss and expense triangle data and generally are based on the premise that the underlying assumptions of the chain ladder method (CLM) are correct. The Thomas Mack method is probably the best-known representative of this family. It provides the first two moments of the claim liability distribution (i.e., the mean and the variance of the distribution)

"Simulation" family of methods. These techniques provide an empirical distribution of the claim liabilities. Our representative of this family is Bootstrapping, a powerful, yet simple, technique that employs simulations and avoids the fitting of complicated analytical models.

A more detailed description of these two methods follows:

Mack method

The Mack method [12] specifies the first two moments of the claim liability distribution only. It essentially calculates the standard error of the claim liability distribution based on the inherent uncertainty of the underlying data. Our research employed the following notation: Let C_{ik} denote the cumulative loss payments for accident year i, $1 \le i \le I$ and development year k, $1 \le k \le I$, where I is the total number of accident years. The values of C_{ik} are known for $i+k \le I+1$. We want to estimate the values of C_{ik} for $i+k > I+1$. The value of the reserves for accident year i is:

$$R_i = C_{iI} - C_{i,I+1-i}, \qquad \text{(A.1)}$$

where C_{iI} represents the true ultimate loss for accident year i. The expected ultimate loss amount for accident year i is calculated by the formula:

$$\mathbf{C}_{iI} = C_{i,I+1-i} \text{ x } \mathbf{f}_{I+1-i} \text{ x} \ldots \text{x } \mathbf{f}_{I-1}, \qquad \text{(A.2)}$$

where $2 \le i \le I$ and $\mathbf{f}_k$ are the observed volume weighted ATA factors from maturity k to k+1 for $1 \le k \le I\text{-}1$. Notice the bolded figure $\mathbf{C}_{iI}$ that represents an estimate of the ultimate loss for accident year i employing historical ATA factors $\mathbf{f}_k$ for $1 \le k \le I\text{-}1$. The true value of the ultimate loss for accident year I is denoted by C_{iI} and depends on the actual ATA factors f_k whose values are currently unknown.

There are three major assumptions that form the base of this paper:

(1) $E(\frac{C_{i,k+1}}{C_{i,k}} / C_{i1},\ldots,C_{ik}) = f_k$ for $1 \le i \le I$ and $1 \le k \le I\text{-}1$, i.e. the expected value of the loss development factor $\frac{C_{i,k+1}}{C_{i,k}}$ equals f_k, where f_k is the unknown "true" development factor which is the same for all accident years. Moreover the loss development factor $\frac{C_{i,k+1}}{C_{i,k}}$ equals f_k irrespective of the prior development $C_{i1},\ldots,C_{ik}$.

(2) The variables $\{C_{i1},\ldots,C_{iI}\}$ and $\{C_{j1},\ldots,C_{jI}\}$ for different accident years $i \neq j$ are independent (i.e. the loss payments in an accident year are independent from the loss payments in another accident year). Under this assumption, the ATA estimators $\mathbf{f}_k$ are unbiased i.e. $E(\mathbf{f}_k) = f_k$.

(3) The 3[rd] major assumption of the paper satisfies the principle of the theory of point estimation that among all the unbiased estimators of the ATA factors preference should be given to the one with the smallest variance. This principle can be restated as:

$$\mathrm{Var}(C_{j,k+1} \,/\, C_{j1},\ldots,C_{jk}) = C_{jk} \times \alpha_k^2, \tag{A.3}$$

where $1 \le j \le I$, $1 \le k \le I\text{-}1$ with unknown proportionality constants α_k^2 for $1 \le k \le I\text{-}1$.

With the help of the previous stated assumptions, we calculated the mean squared error (mse) of the ultimate losses for accident year i. This mse of the ultimate loss is defined as:

$$\mathrm{mse}(\mathbf{C}_{iI}) = E[(C_{iI} - \mathbf{C}_{iI})^2 \,/\, C_{ik} \text{ for } i+k \le I+1]. \tag{A.4}$$

That mean square error is a conditional expectation of the actual triangle data, since Mack measures future claim volatility given a run-off triangle. It can easily be shown that the mse of the ultimate losses and the reserves for a particular accident year i are equal, i.e. $\mathrm{mse}(\mathbf{C}_{iI}) = \mathrm{mse}(\mathbf{R}_i)$. The square root of the mean squared error of the reserves is called the standard error (s.e.) of the reserves. Based on the previously stated assumptions the standard error of the reserves is calculated for every accident year i, s.e.$(\mathbf{R}_i)$, and for all accident years combined, s.e.$(\mathbf{R})$. The resulting formulas are as follows:

$$(\text{s.e.}(\mathbf{C}_{iI}))^2 = \mathbf{C}_{iI}^2 \sum_{k=I+1-i}^{I-1} \frac{a_k^2}{f_k^2} \left(\frac{1}{C_{ik}} + \frac{1}{\sum_{j=1}^{I-k} Cjk}\right), \text{ and} \tag{A.5}$$

$$(\text{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I} \left\{ (\text{s.e.}(\mathbf{R}_i))^2 + \mathbf{C}_{iI} \left(\sum_{j=i+1}^{I} \mathbf{C}_{jI} \right) \sum_{k=I+1-i}^{I-1} \frac{2a_k^2 / f_k^2}{\sum_{n=1}^{I-k} C_{nk}} \right., \tag{A.6}$$

$$\text{where: } \boldsymbol{\alpha}_k^2 = \frac{1}{I-k-1} \sum_{J=1}^{I-k} C_{jk} \left(\frac{C_{j,k+1}}{C_{jk}} - \mathbf{f}_k \right)^2, \; 1 \le k \le I\text{-}2.$$

Using the identity that: $E(X - c)^2 = Var(X) + [E(X) - c]^2$ where c is a constant, we can re-write the mean square error as:

$$mse(\mathbf{R_i}) = Var(R_i \mid D) + [\, E(R_i \mid D) - \mathbf{R_i}\,]^2, \tag{A.7}$$

where D is the observed triangle data, (i.e., we can decompose the total claim liabilities risk into the sum of pure future random error $Var(R_i \mid D)$ and the deviation between the model estimated claim liabilities and the true Expected Claim liabilities (i.e., the parameter risk)). All the components of the mean square error can be calculated based on the implicit assumption of the Mack model that the chain ladder estimated link ratios are unbiased, minimum variance estimators of the true unknown loss development factors.

Bootstrapping method

Bootstrapping [6] is based on theory developed by England and Verrall. In some of their earlier work, they proved that the reserve estimates from the CLM are identical to reserves produced by an over-dispersed Poisson generalized linear model (GLM). As a result, the residuals produced from a chain ladder model fitted to a historical triangle can be treated as residuals of a regression model. The residuals of regressions should be approximately independent and identically distributed around zero. The Bootstrapping technique samples, with replacement, the residuals of the CLM. The resulting simulated residuals can be considered as residuals from a triangle that have approximately the same statistical characteristics as the triangle that produced the original residuals. Using appropriate residuals (the so-called Pearson residuals) we can produce new sets of incremental payments and subsequently new reserve indications from each simulation.

The Bootstrapping algorithm steps are as follows:

Using all years volume weighted loss development factors (LDFs) from the original triangle, a "fitted" triangle is calculated by applying these LDFs to the latest diagonal of the original triangle.

Fitted incremental values are compared to actual incremental values to calculate unscaled residuals. The formula for the Pearson residual is = (actual – fitted) / sqrt(fitted).

Residuals are normalized by an appropriate scale factor: $\sqrt{\frac{n}{n-p}}$, where n is the number of data point in the triangle and p is the number of the parameters in the over-dispersed Poisson GLM model. The scaling factor adjusts for the difference in the degrees of freedom between the parameter free Bootstrapping model and the over-dispersed Poisson GLM model.

The model re-samples these scaled residuals with replacement. The re-sampling is performed once per simulation. Using these re-sampled residuals, an incremental "bootstrap" loss triangle is created based on the Pearson residuals formula. These incremental losses are converted to cumulative, from which all years volume weighted LDFs are calculated. These are then used to "complete the square," by application of the LDFs to the latest diagonal. Reserves are then calculated for each simulation, and a distribution is assembled using the results of all the simulations. This step captures parameter risk only.

Process risk is introduced by treating each incremental from the bootstrap triangle as the mean of a gamma random variable with variance proportional to the mean. The subsequent steps are identical to those shown above.

Tail variability is modeled by using an inverse power curve fit (the so-called Sherman inverse power curve). The parameters of a linear regression are fitted to available age to age factors (ATA) from all accident years as follows:

$$\text{ATA} = 1 + a \times t^{-b}, \tag{A.8}$$

where a and b are the fitted parameters while t represents the development year. The fitting procedure employs the natural logarithms of the ATA factors and the resulting formula is:

$$\ln(\text{ATA-1}) = \ln(a) - b \ln(t). \tag{A.9}$$

With the use of a linear regression the a and b parameters are calculated based on a least square error approach. The development factors in the tail of the triangle vary at each simulation since the ATA factors from the historical years vary at each simulation too.

Separate fits are used for parameter and total risk. The length of the tail is different by line, as described in the body of the paper.

B. Technical Appendix – Financially Impaired Companies

The following companies were analyzed to establish upper bounds on the significance level. These companies were impaired in 2002 according to Best's Insolvency Study, Property/ Casualty U.S. Insurers 1969-2002.

A.M. Best #	Company name as listed in A.M. Best database	State
03627	Aberdeen Insurance Co	TX
02681	Acceptance Insurance Co	NE
03754	American Growers	NE
00685	American Professionals Insurance Co	IN
12181	Aries Insurance Co, Inc	FL
02141	Casualty Reciprocal Exchange	MO
02412	Equity Mutual Insurance Co	MO
10561	Grange Mutual Insurance	OR
02592	Highlands Casualty Co	TX
02239	Highland Insurance Co	TX
02812	Highlands Lloyds	TX
11860	Legion Indemnity Co	IL
02352	Legion Insurance Co	PA
02348	National Automobile & Casualty Insurance Co	CA
00213	NN Insurance Co	WI
10626	Oak Casualty	IL
02880	Pacific Automobile Insurance Co	CA
02376	Pacific National Insurance Co	CA
03658	PAULA Insurance Co	CA
10420	Security Indemnity Insurance Co	NJ
00858	State Capital Insurance Co	NC
02489	Statesman Insurance Co	IN
12110	Villanova Insurance Co	PA
00942	Wasatch Crest Mutual Insurance Co	UT
10630	Western Specialty Insurance Co	IL

C. Technical Appendix – Copula

Copula theory provides a convenient way to calculate the aggregate distribution of several random variables, given a predetermined correlation matrix among these variables. We started with n=4 lines of business where the mean of the claim liabilities μ, (an (nx1) vector), and the nxn correlation matrix C of the claim liabilities between lines are already given. An assumption that needs to be satisfied is that the correlation matrix C is positive definite (an nxn matrix C is positive definite if it is symmetric and if $x'Cx > 0$ for every n-dimensional column vector $x \neq 0$). In the following steps we will describe the normal copula methodology.

1. The claim liability distribution for each line of business is calculated based on the Mack or Bootstrapping methods.
2. Employing the so-called Cholesky decomposition method, we can calculate a randomly generated n-variate normal vector X with each of its vectors satisfying the predetermined correlation matrix C. The required steps for this Cholesky decomposition are as follows:
   1. Since C is a positive definite matrix we can prove, with the help of intermediate algebra, that C can be factored as follows:
      $C = L \times L'$, (where L is a lower triangular matrix from the Cholesky decomposition and L' is the transpose of L);
   2. We introduce a linear transformation X, i.e. $X = \mu + L \times z$, , where z is an nx1 vector from a standard normal distribution, i.e. $z \sim N(0,1)$;
   3. Then: $E(X) = \mu + E(L) \times E(z) = \mu$, since $E(z)=0$ &
      $Var(X) = E((X-\mu)x(X-\mu)') = E((L \times z)x(L \times z)') =$
      $E((L \times (z \times z)') \times L') = E(L \times L') = C$;
      since $Var(z)= E(z \times z') = I$ (i.e. the identity matrix) and $L \times L' = C$;

   and

      - The end result is an n-variate normal vector X, where $X \sim N(\mu,C)$, i.e. the n-variate normal vector X has the required mean μ and the required correlation matrix C.

Copula theory has been gaining acceptance among actuaries. For example [13], "Correlation and Aggregate Loss Distributions With An Emphasis On the Iman-Conover Method", written by Stephen J. Mindenhall and published in the Winter 2006 CAS Forum, explains in more detail the multivariate Normal Copula approach described above.

We assumed the following correlations between lines:

	Personal Auto Liability	Homeowners	Workers Compensation	Other Liability
Personal Auto Liability	1.00	0.40	0.38	0.60
Homeowners	0.40	1.00	0.40	0.40
Workers Compensation	0.38	0.40	1.00	0.19
Other Liability	0.60	0.40	0.19	1.00

D. Technical Appendix – Detailed Calculation of Materiality Standard

Step 1: Historical paid loss triangle, company A and application of stochastic methods

We start with a historical paid loss triangle for the legal entity A. **We assume that A is a mono-line writer.** We denote the random variable of the unpaid claim liabilities by X. Employing the Mack method we can calculate the first two moments of the claim liability distribution, i.e. the mean, E(X), and the corresponding coefficient of variation, CV(X). Using the Bootstrapping method we calculate an empirical distribution of the claim liabilities. As a byproduct of this empirical distribution we can calculate the mean and the coefficient of variation of the claim liabilities.

For the calculation of estimation materiality standards only parameter risk was considered while for the calculation of outcome materiality standards two types of risk (i.e., process and parameter risk) were considered.

Step 2: Calculation of benchmark significance levels/exceedence ratios

In all the steps of our analysis, except the second step, we employ data from individual companies in order to calculate standards of materiality. For the calculation of the benchmark significance levels and benchmark exceedence ratios, we employ a subset of the industry-wide data, not company specific data. The benchmarks were calculated separately for a group of 39 financially healthy companies and a group of 16 financially impaired companies. These benchmarks were employed in the calculation of outcome materiality standards.

For each company in our database we employed the Mack method to calculate the mean, E(X), and coefficient of variation, CV(X), of their respective total risk claim liability distribution. For the calculation of these distributions we employed loss and ALAE triangles from Schedule P, Part 3 Summary. For simplicity, we added the losses for all lines of business written by a company before calculating its claim liability distribution. We implicitly assumed that each company had not experienced any change in its exposures, among their various lines of business, over the past 10 years. A more detailed, but also more time-consuming approach for each company would be to calculate the claim liability distribution for each of their individual lines and then calculate the aggregate distribution based on the combination of these individual lines distributions.

Another underlying assumption is that each company's claim liability distribution has a log-normal form.

We then recorded the risk based capital amount (RBC) for each company, as provided in their respective annual statements, on the "Five-Year Historical Data" page. Based on the RBC amount we calculated the different NAIC-mandated regulatory, or company action levels. So for example if a company had an RBC amount of $10,000 then we have the following levels:

RBC Action Levels	"Required Policyholder Surplus"
No action required (>100%)	$10,000 or more
Company action required (75%-100%)	$7,500 to $10,000
Regulatory action required (50%-75%)	$5,000 to $7,500
Regulatory control authorized (35%-50%)	$3,500 to $5,000
Regulatory control mandated (<35%)	$3,500 or less

As a next step, we measured the difference between the surplus as regards to policyholders and the RBC capital amount that would downgrade each company to the next lower RBC action level. So for example, if the surplus of the company is $12,500 then the calculated difference is $2,500(=12,500-10,000), while if the surplus of the company is $6,000 then the calculated difference is $1,000(=6,000-5,000.) If the company sustained an adverse claim liability deviation greater or equal to the calculated difference it would be downgraded to the next lower RBC level. The difference indicated above can serve as a maximum standard of materiality, within the solvency perspective of materiality.

For each company under consideration we calculated a maximum standard of materiality m. We also assumed that the claim liability distribution of each company is log-normal, while the mean and variance of these distributions have already been calculated by the Mack method. Given the first two moments of a log-normal claim liability distribution we can easily calculate percentiles.

The benchmark significance level is the area in the tail of the company's claim liability distribution, in excess of the mean plus the maximum materiality standard (i.e. E(X)+m.) This area represents the probability of extreme claim liability outcomes that, if materialize, would downgrade the company to the next lower RBC level.

The benchmark exceedence ratio is equal to the expected losses in excess of E(X)+m, as a ratio to the expected claim liabilities E(X). This ratio represents the expected risk of material adverse deviation as a percentage of carried reserves that, if materialize, would downgrade the company to the next lower RBC level.

Steps 3a and 3b describe in more detail how to calculate the percentiles and expected losses, in excess of a given threshold, for a log-normal distribution.

Finally, we calculated the weighted average, across all companies, benchmark significance levels and benchmark exceedence ratios using the carried reserves of each company as a weight. For healthy companies the weighted average benchmark significance level is 6.0% while the weighted average benchmark exceedence ratio is 1.5%. These benchmarks were employed for the calculation of upper tail test outcome materiality standards. For the lower tail test outcome materiality standards we selected judgmentally a benchmark significance level equal to 8.0%. The selection of higher benchmark significance level for the lower tail test makes sure that the resulting outcome standards of materiality are higher for the upper tail test when compared to those of the lower tail test.

The benchmark significance levels and benchmark exceedence ratio for the estimation materiality standards were calculated based on judgment, as explained in the text.

Step 3a: Estimation materiality standards, Company A

Mack – Percentile Threshold approach

We have already calculated:

(a) The mean of the claim liability distribution, E(X);

(b) The coefficient of variation of the claim liability distribution CV(X); and

(c) The benchmark significance level r for estimation materiality. This is 7.5% for the upper tail test and 10.0% for the lower tail test.

We make the additional assumption that the claim liabilities follow a log-normal distribution with parameters μ and σ, i.e. $X \sim LN(\mu,\sigma)$. The logarithm of X then is normally distributed with parameters μ and σ, i.e. $\ln(X) \sim N(\mu,\sigma)$. From introductory statistical theory we can calculate μ and σ by:

$$\mu = \ln(E(X)) - \frac{\sigma^2}{2}, \text{where } \sigma = \sqrt{\ln(1+CV(X)^2)} \qquad \text{(D.1)}$$

The purpose of the percentile threshold approach is to calculate a range of reasonable estimates around the carried reserves that is outside the upper and lower tails of the distribution, as defined by the benchmark significance levels.

For the calculation of the upper tail estimation materiality standard, we subtracted the mean reserves from the 92.5th (=1-0.075) percentile implied by the benchmark significance level:

$$\textbf{Materiality standard - Upper tail} = E(X) * \exp^{\varphi(0.925)*\sigma-\sigma^2/2} - E(X), \qquad \text{(D.2)}$$

where $\phi(0.925)$ represents the 92.5^{th} percentile of the standard normal distribution function. The first component of the preceding formula represents the 92.5^{th} percentile of the log-normal distribution X of the claim liabilities.

For the calculation of the lower tail estimation materiality standard we subtract the 10^{th} percentile implied by the benchmark significance level from the mean reserves:

$$\textbf{Materiality standard - Lower tail} = E(X) - E(X) * \exp^{\varphi(0.10)*\sigma-\sigma^2/2}, \qquad \text{(D.3)}$$

where $\phi(0.10)$ represents the 10.0^{th} percentile of the standard normal distribution function. The second component of the preceding formula represents the 10.0^{th} percentile of the log-normal distribution X of the claim liabilities.

For the purpose of this analysis we employed a mean of the claim liabilities equal to the carried reserve for legal entity A.

Bootstrapping – Percentile Threshold approach

The Bootstrapping stochastic method calculates an empirical distribution of the claim liabilities. The Bootstrapping method produces a few thousand random realizations of the empirical claim liability distribution though a simulation approach. The first step is to linearly transform the stochastic claim liability distribution to make sure that the mean of that distribution is equal to the carried reserves for legal entity A. The transformed distribution has the same coefficient of variation as the original stochastic empirical distribution. The *percentile* function in excel can calculate the various percentiles of the resulting transformed distribution.

The upper tail estimation materiality standard is calculated as follows:

Materiality standard - Upper tail =

92.5^{th} percentile of simulated claim liability distribution - E(X).

The lower tail estimation materiality standard is calculated as follows:

Materiality standard - Lower tail =

E(X) - 10^{th} percentile of simulated claim liability distribution.

Mack – Expected exceedence/TVar approach

For the Mack approach we were provided with the mean, E(X), and the coefficient of variation, CV(X), of the claim liability distribution. Again we assume that the claim liabilities X follow a lognormal distribution with parameters μ and σ. The selected benchmark exceedence ratio is equal to 2.0%.

The purpose of the expected exceedence approach is to calculate a standard of materiality that when added to the carried reserves, the expected losses in excess of these carried reserves plus the materiality standard, is equal to 2.0% of the carried reserves, (for estimation materiality standards.) In other words, if the company experiences actual losses in excess of the expected losses plus the standard of materiality, then the expected material adverse deviation is equal to 2.0% of the carried reserves. A risk of material adverse deviation exists when the actual losses exceed expected losses (i.e., E(X)), by the selected materiality standard. By construction the TVar measure of risk focuses only on the upper tail of the distribution.

Available optimization routines in Excel™, such as *SOLVER*, can help us calculate the standard of materiality m. When we add this standard of materiality to the carried reserves E(X) then the expected losses in excess of E(X)+m are equal to 2% of the carried reserves.

The formula for the expected losses, in excess of the carried reserves plus the materiality standard (i.e. E(X)+m), is as follows:

$$\left\{1 - \frac{E[X;E(X)+m]}{E(X)}\right\} \text{x } E(X), \tag{D.4}$$

where E[X;E(X)+m] represents the expected losses from the claim liability distribution limited to E(X)+m (the so called limited expected value function.)

With an assumption of a log-normal distribution for X ~ LN(μ,σ), we calculated the expected losses limited to an upper limit c as follows:

$$E[X;c] = \exp^{\mu+\sigma^2/2} \text{ x } \Phi\left(\frac{\ln(c)-\mu-\sigma^2}{\sigma}\right) + c \text{ x } \left[1 - \Phi\left(\frac{\ln(c)-\mu}{\sigma}\right)\right], \tag{D.5}$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

Again, for the purpose of our analysis we employed a mean of the claim liabilities equal to the carried reserve for legal entity A.

Bootstrapping – Expected exceedence/TVar approach

The empirical distribution produced by the Bootstrapping stochastic reserving method is linearly transformed, as explained in the "*Bootstrapping – Percentile Threshold approach*" section. With the help of *SOLVER*, we can calculate a standard of materiality m that when added to the mean E(X) of the claim liability distribution, the expected losses in excess of E(X)+m are equal to 2.0% of the carried reserves. Again, when a company experiences actual losses that exceed expected losses (i.e., E(X)) by the selected materiality standard amount m, then the expected risk of material adverse deviation is equal to 2.0% of the carried reserves.

The analysis proceeds as follows: We start with a few thousands simulations of the transformed empirical distribution. From each simulated value we subtract the mean of the distribution plus the materiality standard (i.e. E(X)+m.) If the difference:

Simulated value – E(X) – m,

is positive, then the difference represents a material adverse deviation, since the simulated losses exceed the expected loss amount plus the materiality standard amount. If, on the other hand, the difference is negative, then we set it equal to zero since we are interested only in material adverse deviations. We average the material adverse deviations over all the simulated values and we divide this average material adverse deviation by the expected claim liability amount. *SOLVER* ensured that we selected a standard of materiality m that would produce exactly a 2.0% expected risk of material adverse deviation, as a percentage of carried reserves.

Step 3b: Outcome materiality standards, Company A

For the calculation of the outcome materiality standards, we employ exactly the same methodologies described in step 3a for the two stochastic methods, the Mack and Bootstrapping, and the two measures of risk, the percentile threshold approach and the expected exceedence/TVar approach. The only difference is in the benchmark significance level r for outcome materiality. This is 6.0% for the upper tail test and 8.0% for the lower tail test. The outcome benchmark exceedence ratio is 1.5%.

Step 4: Outcome materiality standards, Company B

We assumed that company B was a multi-line writer. The additional analysis, compared to the mono-line company A case, relates to the calculation of the aggregate claim liability distribution from the combination of all lines written by company B.

As a first step, we calculate the claim liability distributions for each of the n lines of business written by company B. Moreover, we assume an nxn correlation matrix C that describes the correlations among these various lines. Based on the Cholesky decomposition methodology described in section "*Normal Copula theory basics*", we can calculate an n-variate normal array X that satisfies the correlation matrix C provided. As a last step, we re-sort the n lines claim liability distributions produced by the Mack and Bootstrapping methods based on the ranking of the nx1 vectors in X. This way, we can achieve the predetermined correlation among the various lines claim liability distributions. We then add all these resorted line distributions together to create an aggregate distribution that represents the combined all-lines liabilities for company B.

Having produced the aggregate distribution for all lines combined we then calculate estimation and outcome materiality standards for company B employing the same techniques described in steps 3a and 3b.

The following Exhibits 1 through 5 illustrate the calculation of outcome materiality standards for company A for both the Mack and Bootstrapping stochastic methods and both the Percentile Threshold and TVar measures of risk approaches.

5. REFERENCES

[1] A.M. Best Company, Best Insurance Reports, 2003.
[2] A.M. Best, "Special Report, Best's Insolvency Study, Property/Casualty U.S. Insurers 1969-2002," May 2004.
[3] Butsic, Robert P., "Solvency Measurement for Property - Liability Risk-Based Capital Applications," *Journal of Risk and Insurance*, December 1994, Vol. 61, No. 4, 656-690.
[4] CAS, "Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves", 1988.
[5] CAS Working Party on Quantifying Variability in Reserve Estimates, "The Analysis and Estimation of Loss & ALAE Variability: A Summary Report", *CAS Fall Forum*, 2005, 1-118.
[6] England, Peter and Richard Verrall, "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving," *Insurance: Mathematics and Economics*, 1999, Vol. 25, 281-293.
[7] Faivre, Fabien, "Copula: A New Vision for Economic Capital and Application to a Four Line of Business Company," *ASTIN Bulletin*, May 2001, Vol. 31, No. 1, 213-249.
[8] International Accounting Standards Committee Foundation, "Draft Statement of Principles," ED 5.
[9] Kaye, Paul, "Risk Management in Insurance, A Guide To Risk Measurement, Capital Allocation And Related Decision Support Issues," *CAS Discussion Paper Program*, 2005, 1-34.
[10] Kreps, Rodney E., "Management's Best Estimate of Loss Reserves", *CAS Fall Forum*, 2002, 247-258.
[11] Levine, George M., "Considerations Regarding Materiality and Range of Reserves in Connection with Actuarial Standard of Practice #36", *CAS Fall Forum*, 2002, 259-286.
[12] Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates," *CAS Spring Forum*, 1994, 101-182.
[13] Mindenhall, Stephen J., "Correlation and Aggregate Loss Distributions With An Emphasis On the Iman-Conover Method," *CAS Winter Forum*, 2006, 103-204.
[14] Sherman, Richard E., "Extrapolating, Smoothing, and Interpolating Development Factors," *CAS Proceedings*, 1984, Vol. LXXI, Nos. 135 and 136, 122-155.

Appendix
Exhibit 1

Mack model, company A

AY/DY	1	2	3	4	5	6	7	8	9	10
i/k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
i=1	14,772	31,846	41,871	47,589	51,385	54,529	57,055	58,882	60,174	61,063
i=2	13,405	28,201	37,197	42,643	46,982	49,743	51,464	52,737	53,780	54,575
i=3	11,687	23,476	29,419	34,024	37,629	39,628	41,595	42,741	43,635	44,280
i=4	11,166	22,916	29,185	34,145	36,918	38,879	40,281	41,420	42,287	42,912
i=5	12,772	27,516	36,160	41,694	45,327	48,014	50,015	51,429	52,505	53,281
i=6	15,567	33,596	44,299	51,209	55,602	58,800	61,250	62,982	64,300	65,250
i=7	15,460	36,635	49,218	56,789	61,882	65,441	68,168	70,096	71,563	72,620
i=8	16,556	36,393	49,600	57,159	62,286	65,868	68,612	70,553	72,029	73,093
i=9	18,261	38,847	51,179	58,978	64,268	67,964	70,796	72,799	74,322	75,420
i=10	19,151	41,276	54,379	62,666	68,287	72,214	75,223	77,351	78,969	80,136
LDFs	2.155	1.317	1.152	1.090	1.058	1.042	1.028	1.021	1.015	1.103
CDFs	4.617	2.142	1.626	1.411	1.295	1.224	1.175	1.143	1.120	1.103
ak2	145	35	4	5	1	2	1	0	0	

AY	**Ci,10**	**Ri**	**s.e.(Ri)**			**s.e.(Ri) / Ri**		
			Process risk	**Parameter risk**	**Total risk**	**Process risk**	**Parameter risk**	**Total risk**
i=2	60,213	6,433	5,419	5,419	7,664	84.2%	84.2%	119.1%
i=3	48,855	6,114	4,398	4,397	6,219	71.9%	71.9%	101.7%
i=4	47,345	7,064	4,267	4,263	6,031	60.4%	60.3%	85.4%
i=5	58,786	10,772	5,312	5,297	7,501	49.3%	49.2%	69.6%
i=6	71,992	16,390	6,504	6,488	9,186	39.7%	39.6%	56.1%
i=7	80,123	23,334	7,266	7,227	10,249	31.1%	31.0%	43.9%
i=8	80,645	31,045	7,338	7,279	10,336	23.6%	23.4%	33.3%
i=9	83,212	44,365	7,800	7,549	10,855	17.6%	17.0%	24.5%
i=10	88,415	69,264	9,007	8,137	12,138	13.0%	11.7%	17.5%
Total:	619,587	214,782	19,648	19,265	27,517	9.1%	9.0%	12.8%

Note: The model judgmentally incorporates process and parameter tail variability of 9% respectively.

Appendix
Exhibit 2

Company:	**A**
Risk:	**Total**
Stochastic Method:	**Mack**
Measure of risk:	**Percentile threshold approach**

(1)	(2)	(3)	(4)
E(X)	**CV**	**μ**	**σ**
221,517	12.8%	12.300	12.8%

Benchmark significance levels:	
(5) Upper tail test	6.0%
(6) Lower tail test	8.0%

Outcome materiality standards:	
(7) Upper tail test	46,417
(8) Lower tail test	37,858

Standards as a % of carried reserves	
(9) Upper tail test	21.0%
(10) Lower tail test	17.1%

Notes:

(1) The carried reserves for company A. Notice that the Mack indicated reserves R_i of $214,782 are slightly different.

(2) The total risk coefficient of variation produced by the Mack method.

(3) = ln[(1)] - ln[1+(2)^2] / 2

(4) = ln[1+(2)^2] $^{1/2}$

(5) & (6), calculated from the Bright Line Test.

(7) = E(X) * $\boxed{\exp^{\varphi(0.94) * \sigma - \sigma^2/2}}$ - E(X)

(8) = E(X) - E(X) * $\boxed{\exp^{\varphi(0.08) * \sigma - \sigma^2/2}}$

where ϕ(x) is the xth percentile of the standard normal distribution function.

(9) = (7) / (1)

(10) = (8) / (1)

Company:	**A**
Risk:	**Total**
Stochastic Method:	**Mack**
Measure of risk:	**TVar approach**

(1)	(2)	(3)	(4)
E(X)	**CV**	**μ**	**σ**
221,517	12.8%	12.300	12.8%

(5) Threshold of material adverse deviation	246,644
(6) Implied Materiality Standard m	25,127
(7) Expected losses in excess of E(X) + m	3,323
(8) Benchmark expected exceedence ratio	1.50%
(9) Outcome materiality standard as a percentage of carried reserves	11.3%

Notes:

(1) The carried reserves for company A. Notice that the Mack indicated reserves of $214,782 are slightly different.

(2) The total risk coefficient of variation produced by the Mack method.

(3) = ln[(1)] - ln[1+(2)^2] / 2

(4) = ln[1+(2)^2] $^{1/2}$

(5) = (1) + (6)

(6) Calculated with excel's *solver* function in order to produce an expected exceedence ratio equal to 1.50% of carried reserves.

(7) = $\{ 1 - \boxed{\frac{E[X;E(X)+m]}{E(X)}} \} * E(X)$,

where: E[X;E(X)+m] = exp[(3)+(4)^2/2] * Φ[{ln[(5)+(6)]-(3)-(4)^2}/(4)] + [(5)+(6)] * [1-Φ{[ln[(5)+(6)]-(3)]/(4)}]

where Φ(x) is the standard normal cumulative distribution function.

(8) = (7) / (1)

(9) = (6) / (1)

Company: **A**
Risk: **Total**
Stochastic Method: **Bootstrapping**
Measure of risk: **Percentile threshold approach**

(1)
E(X)=
221,517

	(2) Bootstrapping model	(3) Transformed distribution
E(X)	216,877	221,517
CV(X)	9.6%	9.6%

Simulations	**Simulations**	**Transformed Sample**
1	249,793	255,137
2	218,344	223,015
3	240,363	245,505
4	232,179	237,146
5	235,964	241,011
6	257,774	263,289
7	211,185	215,702
8	223,796	228,584
9	222,998	227,768
10	208,965	213,436
…..	…..	…..
4,990	223,374	228,152
4,991	229,695	234,608
4,992	244,934	250,174
4,993	223,934	228,724
4,994	218,231	222,900
4,995	186,571	190,562
4,996	229,320	234,226
4,997	229,418	234,326
4,998	243,117	248,318
4,999	174,439	178,171
5,000	228,333	233,218

Benchmark significance levels:	
(4) Upper tail test	6.0%
(5) Lower tail test	8.0%

Outcome materiality standards:	
(6) Upper tail test	35,399
(7) Lower tail test	28,412

Standards as a % of carried reserves	
(8) Upper tail test	16.0%
(9) Lower tail test	12.8%

Notes:
(1) The carried reserves for company A.
(2) The mean, E(X), and the coefficient of variation, CV(X), provided by the boostrapping model.
5,000 simulations were performed.
(3) = (2) x (1) / [E(X) from (2)]
The transformed distribution has mean equal to the carried reserves and CV equal to the one calculated from the bootsrapping model.
(4) & (5), calculated from Bright Line Test.
(6) = 94.0th percentile of (3) - E(X)
(7) = E(X) - 8.0th percentile of (3)
(8) = (6) / (1)
(9) = (7) / (1)

Company:	**A**
Risk:	**Total**
Stochastic Method:	**Bootstrapping**
Measure of risk:	**TVar approach**

(1)
E(X)=
221,517

	(2) **Bootstrapping model**	(3) **Transformed distribution**	(4) **Expected material adverse deviation**
E(X)	216,877	221,517	
CV(X)	9.6%	9.6%	
Simulations	**Simulations**	**Transformed Sample**	
1	249,793	255,137	19,095
2	218,344	223,015	0
3	240,363	245,505	9,463
4	232,179	237,146	1,104
5	235,964	241,011	4,969
6	257,774	263,289	27,247
7	211,185	215,702	0
8	223,796	228,584	0
9	222,998	227,768	0
10	208,965	213,436	0
…..	…..	…..	…..
4,990	223,374	228,152	0
4,991	229,695	234,608	0
4,992	244,934	250,174	14,132
4,993	223,934	228,724	0
4,994	218,231	222,900	0
4,995	186,571	190,562	0
4,996	229,320	234,226	0
4,997	229,418	234,326	0
4,998	243,117	248,318	12,275
4,999	174,439	178,171	0
5,000	228,333	233,218	0

(5) Threshold of material adverse deviation	236,042
(6) Implied Materiality Standard m	14,525
(7) Expected losses in excess of E(X) + m	3,323
(8) Benchmark expected exceedence ratio	1.50%
(9) Outcome materiality standard as a percentage of carried reserves	6.6%

Notes:
(1) The carried reserves for company A.
(2) The mean, E(X), and the coefficient of variation, CV(X), provided by the bootstrapping simulation model.
5,000 simulations were performed.
(3) = (2) x (1) / [E(X) for (2)]
The transformed distribution has mean equal to the carried reserves and CV equal to the one calculated from the bootstrapping model.
(4) = max{(3)-(1)-(6),0}
(5) = (1) + (6)
(6) Calculated with excel's *solver* in order to produce an expected exceedence ratio equal to 1.5% of carried reserves.
(7) Average of (4).
(8) = (7) / (1)
(9) = (6) / (1)