ESTIMATING THE REQUIRED SURPLUS, BENCHMARK PROFIT, AND OPTIMAL REINSURANCE RETENTION FOR AN INSURANCE ENTERPRISE USING THE COMPOUND POISSON DISTRIBUTION

Joseph A. Boor, FCAS*
Florida Office of Insurance Regulation
Tallahassee, Florida U.S.A.
Florida State University
Tallahassee, Florida U.S.A.

August 2012

*joeboormath@gmail.com
Abstract

This paper presents an analysis of the capital needs, needed return on capital, and optimal reinsurance retention for insurance companies, all within the context of the compound Poisson distribution. As an alternative to much of the present practice, it focuses on closed form expressions and closed form approximations, rather than focusing on how to estimate such values using Monte Carlo simulation. The analysis is also done using a distribution-free approach with respect to the loss severity distribution.

It shows how the risk of extreme aggregate losses that is inherent in insurance operations may be understood (and, implicitly, managed) by using key values from the loss severity distribution. The capital and surplus needs of a company are then estimated using a VaR approach. A tractable formula for the benchmark profit need of a company is developed. An analysis of the economically optimal reinsurance retention/policy limit is performed as well. It shows that the marginal (across loss caps) profit need should equal to the marginal reinsurer loading on losses. Analytical expressions are then developed for the optimal reinsurance retention. Approximations to the optimal retention based on the normal distribution are developed and their error is analyzed in great detail. For sample data that is known to be difficult to approximate with a normal distribution, the results indicate the normal approximation to the optimal retention is acceptable.

The impacts of those results on other aspects of insurance company operations are discussed. It discusses that their is a logical limit on insurance benefits beyond which the cost of the insurance outweighs the benefits. Also, the benchmark loading for profit and the amount needed to recompense investors for diversifiable risk is discussed. An analysis of whether or not the loading for diversifiable risk is needed is performed, suggesting that some small load for the randomness of insurance claims is required to support the capital employed by an insurance company. The profit load needed in the rates is shown to be independent of how an insurance company invests its assets, and as such is mostly independent of the CAPM “beta” of the company as a whole. It is shown to be related strictly to the risk-free rate and the asset structure of an insurance company in the most common cases.
1 Introduction

Many aspects of actuarial literature give fairly precise direction with respect to the classical problem of actuarial science — estimating the losses expected under insurance contracts. However, there is no clear direction from the literature on some of the most basic questions underlying financial operations of insurance companies, such as: “How much ‘surplus’ (buffer fund) is needed to reasonably ensure that an insurance company can fulfill its promises to customers?”; “How much profit does an insurance company need to generate in order to compensate investors adequately?”; and, “What types of losses should be ‘retained’ by an insurance company and what types should be passed along to a ‘reinsurer’?” These are key questions that impact the ratemaking process, the “reinsurance” purchasing process, and the enterprise risk management and financial management processes of insurance companies.

Software programs are available that allow users to create highly parameterized mathematical models of an insurance company’s operations (one specific example may be found in Feldblum [38], another in Lowe and Stanard [50]). Those models typically require the user to select a highly specific distribution or a highly specific combination of distributions for the baseline losses, the parameter uncertainty of each coefficient used to fit each instance of a family of distributions, the baseline values of each coefficient, etc. Then, the modeling process generally requires generating a large number of Monte Carlo simulations of the possible losses the insurance company might suffer. Those simulations are intended to provide perspective on the riskiness of the insurance company. One could use such a model to estimate how likely the insurance company is to suffer more claims than it has funds to pay. Similarly, one could use such a model to estimate how much capital and surplus the company needs. Possibly one might even, by trial and error, estimate the optimal loss size at which the company should stop retaining responsibility and begin transferring the losses to a reinsurer. Those programs share a few key limitations. First, since the backbone of the process is a large set of Monte Carlo simulations, it is generally impossible for a user of the resulting analysis to flow through the calculations and personally validate the accuracy of the results. Since many users of the resulting reports are chief financial officers and other high level executives of insurance companies, they are accustomed to receiving reports that show all the underlying

\footnote{Reinsurance is so named because a risk that is already the subject of insurance passed to another reinsurer.}
calculations. The fact that so much of the simulation process lies in the background, and hence is not auditable, can lead to a lack of trust in the results of such simulations.

Further, many working actuaries still rely on the “RAND()” function from Excel, which is known (see Coddington [18]) to have autocorrelation problems. It is possible that models contain errors of this sort, and other errors in the choice of probability distributions that are esoteric in the minds of company managers. However, while those errors may be esoteric in the minds of the key users of the model output, they can affect the output greatly. So, management concerns that the modeled results are not easily audited may be well-founded. Further, insurance claims costs are not driven by basic equations such as Maxwell’s equations for electrodynamics or Newton’s laws of motion. In fact, there is no reason for insurance claim costs to arise from any specific formula. Consequently, the actual claims “severity distributions” which companies face are not specific instances of any family of probability distributions. So, there are good reasons to seek approaches to the capital need and reinsurance retention problems that involve closed form formulas. In order to avoid the bias that the use of specific distributions can bring, it is desirable to develop equations for the capital need, reinsurance retention, and related items that are as distribution-free as possible.

There is still considerable debate in the actuarial community regarding the financial issues addressed in this paper. For example, due to the random nature of insurance claims (both regarding how many claims are received and the cost of each), insurance company losses and expenses in a particular time period sometimes exceed the premiums collected to pay them. Consequently, an insurance company must have a buffer fund. Otherwise it will not be able to reliably honor its obligation to claimants. There are different views as to how large of a buffer fund (which is loosely\textsuperscript{2} equal to the so-called “surplus” funds of an insurer) solvency regulators and agencies that rate insurer’s creditworthiness should require of an insurance company. United States solvency regulation includes a number of so-called I.R.I.S.\textsuperscript{3} tests, including one requiring surplus to be at least one third of the pre-

\textsuperscript{2}Technically, under US statutory accounting loss reserves (for future payouts on claims that have happened but are unresolved) are undiscounted, credits for expenses paid early in the course of a policy that are not expect to reoccur as the policy runs its course (so-called prepaid acquisition costs) are not reflected, and assets described as long-term holdings are not valued at their market value. So, although surplus is loosely intended to represent free funds, it is technically an inexact measure of free funds (and often slightly lower than available free funds.)

\textsuperscript{3}Insurance Regulatory Information System
mium “written” after deducting for reinsurance premiums purchased. US
regulatory accounting also requires that various adverse actions be taken
against an insurer if the ratio of its surplus to its “risk-based capital” (as
discussed in Feldblum [36]) falls below certain levels. The risk-based capital
formula crudely represents one standard deviation of its annual fluctuations
in operating results. The Standard and Poor’s procedure presented by Fed-
erman [31] also notably references considerable use of individual judgment
in addition to firm parameters. Some views of the European approach are
contained in Feldblum [35], de Castries [23] and Drzik [25]. So, while most
of those approaches intuitively seem to be somehow related to the standard
deviation of results, they are not explicitly presented in that fashion. How-
ever, since they are similar to a standard deviation approach, they loosely
approximate certain minimum thresholds of the probability of ruin within
a given year. There have even been analyses that focus not on the capital
itself, but rather on possible contingent capital to be received in the event
of dire scenarios (see Lin, Chang, and Powers [48]).

There is also debate within the actuarial profession as to how to measure
the riskiness of insurance enterprises. An important paper by Butsic [8]
discusses the use of expected policyholder deficit, or “EPD” (the percentage
of the total losses across all scenarios for the total loss costs that are in
excess of the total funds the insurance company has available for claims
payments). As an alternate risk measure, the value associated with funding
losses up to a specific confidence level is thought of as the “value at risk”,
or VaR. One must note that the VaR in broad usage in other industries
relates to how much money may be lost. In the insurance context, it relates
to how much money is needed to fund a given percentile of the distribution
of possible “aggregate” (cost of all claims combined) losses experienced by
an insurance company in a given year. However, the two meanings are
mathematically identical. Another measure in common use today is the
tail-value-at-risk or “TVaR”. That refers to the average costs for all loss
scenarios that exceed the given VaR funding level. Of note, Powers [61]
espouses that the expected discounted cost of future insolvencies, or EDCI,
should be used. Given that brief description of the alternatives, one of the
objectives of this paper is to present a coherent theory for evaluating the
amount of the buffer fund needed to secure the insurer’s promise to pay
claims.

Of course, insurance markets have varying degrees of competition. In
many cases competition affects profits considerably. However, financial per-

\footnote{Effectively, “sold”}. 

4
sonnel that employed by insurance companies often compare the profits of their organizations to a benchmark profit. Stockholders certainly compare companies’ profit performance to benchmarks. Further, insurance regulators must often opine on the “underwriting profit” (profit markup in pricing) built into a company’s rates. So, it is important to understand what profit target should be used and what underwriting profit should be built into a company’s rates. However, there is no broad consensus among actuaries as to what underwriting profit benchmark to use for property/casualty insurance companies. Further, there is no consensus on what underwriting profit provides, with related profit sources, the return needed to sustain the capital provided by shareholders. As an example of the differences of opinion, a paper by Robbin [63] that is used in the course of study for casualty actuaries lists seven different approaches that actuaries may use to determine profit loads. The approaches vary significantly. One involves using a preselected (but not derived from financial theory) flat percentage profit load on all premiums sold, less a portion of the investment income relating to the funds held to pay losses\(^5\). Another method prescribes a given (but also not fully theoretically supported) profit from all income sources. One approach equalizes the discounted profit to the discounted cost of a multi-year equity commitment. In another approach, the net present value of premiums less the net present value of losses and expenses is required to be equal to the net present value cost of equity. A variant of that replaces broad market interest rates with risk-adjusted interest rates. Lastly, another approach involves requiring that the profit generated by an insurance company exceed a “hurdle rate” applied to invested capital. Of note, while each of those methodologies may have some use in some context, they do not provide a single comprehensive view of the needed profit. Such a view should logically begin with first principles such as “what profit percentage, of what type, is required, and why?” businesses

The poor understanding of profit needs is, in a sense, a logical consequence of the basic financial structure of most financial intermediaries. In financial intermediaries most of the underlying capital needed to begin the business, in contrast to being invested in plants and equipment, is invested in direct income-producing assets such as stocks and bonds. Further, due to the gap between when the premiums are received and when claims are paid, additional investment income is received from “policyholder-supplied” (as defined by Bailey [1]) funds as well. Lastly, the underwriting profit is

\(^5\)Two different methods, of the seven, represent different ways to compute investment income.
earned. Because of the multiple sources of profit, operating profit includes more than underwriting profit. Underwriting profit is the difference between premiums received and losses paid out. Operating profit also includes the investment income on policyholder supplied funds. In many contexts, it also includes the passive returns on the capital invested in the company. The presence of multiple types of profit has served to complicate and obscure the understanding of the proper profit benchmark, as illustrated by the variety of methods listed by Robbin [63]. Also of note, there is a long-running debate as to whether the profit load should be fundamentally based on the riskiness of the insurance sold by the company (as discussed by Meyers [56]) or the cost of capital of the company as exemplified in an article by Cummins and Phillips [22]. In this paper, a comprehensive view of the indicated profit needed to attract capital, and the marginal capital need associated with increasing loss sizes will be presented. Resolution of those issues within this paper will also be key to the analysis of some of the other issues (reinsurance retention, for example) discussed earlier.

Also of note, an insurance company’s role in the economy is clearly to insure the risks of its policyholders. Insuring more policyholders and accepting more risk is very desirable to a company from a sales standpoint. However, when insurance companies add more customers they increase the total standard deviation, variance, etc. of their aggregate losses. The combination of increased risk and a fixed base of assets to cover any shortfall in premiums increases the risk of failure\(^6\). Aside from the increased probability of failure, the earnings of the company will become more volatile. Should the insurance company perform an internal analysis of the risk adjusted return on capital, the perceived economic benefit of adding additional customers may be minimal. Insurers must manage the trade-off between writing additional business and managing the risk of the enterprise. Otherwise, they may not be able to fulfill their promise to pay their customers’ claims. A higher certainty of paying claims should be worth something to consumers. However, there typically is very little useful information on the ability of an insurance company to pay its claims. The only publicly available solvency information for consumers is the set of credit scores established by rating agencies. For that reason, most insurance regulatory agencies have divisions that evaluate the viability of insurance companies. However, beyond the various approaches mentioned above, there is no explicit requirement that

\(^6\)Of note, none of this discussion of the general risk and growth issues facing insurers should be thought of new or innovative. These are well-established industry principles and a fairly trivial result of probability.
insurers must fulfill regarding their ability to pay claims. As a part of this paper, some general advice will be provided on that subject. That in itself would not preclude an insurer from choosing to provide higher reliability to its customers by meeting a higher standard of solvency.

Another item discussed earlier involves the optimal caps on individual losses (so-called “specific excess retentions”). These caps are effected by insurers passing along more severe risks to specialized reinsurers through the mechanism of reinsurance. This subject has received less formal mathematical attention within the broad community of North American casualty actuaries than its importance to insurance company operations would suggest. In effect, the insurer has a choice to make between “retaining” losses through some given size (the retention), or reinsuring those upper levels of loss with a reinsurer. Retaining successively larger and larger individual losses involves successively higher and higher risk for the insurance company. However, in addition to charging for the expected losses above the retention, a reinsurer will add on an expense and profit load. Hence, insurance companies must choose between keeping some expected loss costs with a very high variance or paying a premium beyond the expected loss costs to eliminate the variance. A key paper by Centeno [13] analyzes the trade-off between purchasing “quota share” (transfer of a flat percentage share of all claims costs after deducting excess of loss reinsurance) and “excess of loss” (transfer of all claim costs above some retention “R” per individual claim) reinsurance. Excess reinsurance (transfer of the claim costs \(\sum_{\text{claims } c} (L_c - R)^+\), where \(L_c\) is the cost of the \(c^{th}\) claim) reinsurance has been partially explored. For example, Bernard and Tian [3] approached the reinsurance purchasing problem in a manner similar to that used in Centeno [13]. They asked “If an insurance company has a specified budget to spend on reinsurance, what is the best way to spend it?” Also, Cai and Tan analyzed the problem by selecting which different risks should be ceded in various available layers (see [10]). Most importantly, at present excess of loss retentions are still quite often chosen using rules of thumb such as “the size of the second largest claim we get each year”. An analysis of where the optimal retention should be set will be a key part of this paper.

The focus of this paper is to identify, investigate, and explain a com-

---

7 Of note, this is similar to what households (at least those that are understandably risk averse) do when they replace a fairly volatile set of expected losses in the form of potential house fires, automobile accidents, etc. with the purchase of insurance policies with fairly significant (15-35% of the premium) expense loads. Financial economists have long discussed the rational basis for this. A discussion of the general principle of maximizing utility by paying a surcharge beyond expected costs can be found in [47].
prehensive view of risk, reward, and risk mitigation based on the characteristics of the risks assumed at different loss sizes. Those concepts will be applied to specific scenarios: the no parameter uncertainty monoline compound Poisson distribution; claim count parameter uncertainty; lognormal severity parameter uncertainty; the multiline compound Poisson case; the compound Poisson scenario with parameter uncertainty, and especially the multiline compound Poisson parameter uncertainty scenario where portions of the parameter uncertainty are correlated across the lines of business. The analysis within this paper provides key results that can be used to

- Assess the survival probability of an insurance company given exposure to a class of possible claims;
- Assess the amount of surplus needed to maintain a given survival probability;
- Assess the benchmark profit needed to sustain operations of the company;
- Estimate the company’s optimal reinsurance attachment point;
- Model the percentiles of the aggregate loss distribution for enterprise risk management; and
- Provide closed form solutions that can be implemented in spreadsheets for presentations to management.

To avoid the bias inherent in any choice of statistical assumptions, the following protocols will be followed within the paper:

- The use of specific families of distributions will be avoided whenever possible, unless first principles suggest a given distribution is preferable.
- Further, when a distribution family is chosen, the broadest possible approach (fewest assumptions) will be used.

To fulfill the goals established earlier, the focus here will be on the marginal risk (variance) of an insurance company’s potential total aggregate losses with respect to the maximum loss amount retained by the

---

8Unless a specific example is included for illustration.
9This is also referred to as the “collective risk”.
10As noted earlier, this is often called the retention or “attachment point”
company. Secondly, it will be necessary to compute the capital\textsuperscript{11} and surplus required to assure reasonable confidence that all the company’s claims will be paid. The paper also contains analyses of the marginal capital and surplus needed as the maximum loss amount increases. Of course, once the capital and surplus requirement is established, then discussions of the appropriate return on capital and surplus, an appropriate profit benchmark, and the optimal reinsurance retention to reduce the total cost of the insurance, will follow.

Within the body of this paper, the analysis of the capped aggregate loss distribution (the cost of all the claims in a year when each individual claim is capped at some pre-selected value), its variance, and its derivative with respect to the capping point will be explored. Next, the benchmark required profit, including recognition of the indicated capital commitment, cost of capital, and the riskiness of insurance operations associated with various loss caps will be analyzed. That analysis includes a review of what the proper load should be on the non-broad market, diversifiable, insurance risk. Next, economic principles are applied to the question of how the appropriate amount of capital should be calculated, and the relevant criteria for this decision. That is followed by an analysis of where the loss cap or retention might be set (using excess of loss reinsurance). Rather than approaching the loss cap problem from the internal perspective of the insurance company, the goal will be to set the loss cap so that the total cost to the policyholder is minimized. Approximations will be developed for the key percentiles and the derivatives of the percentiles across the various loss caps.

The reliability of the approximations will be tested with a particularly extreme sample distribution. Comparisons of key approximations such as those for percentiles of the aggregate loss distribution to their true values (as estimated by Monte Carlo simulation) are included.

\textsuperscript{11}Of note, “capital” refers to capital “deployed” (invested in a company with a direct commitment of funds) as financial backing for an insurance enterprise. Insurance company statutory financial statements sent to regulators, and often other accounting statements, include an accounting entry for capital that is often far different from the meaning in the context of this paper. Specifically, the capital shown on an insurance company’s statutory balance sheet supplied to financial regulators may only be the initial capital contributed at the start of the company’s operations, and excludes all the profits retained by the company over the years and deployed to provide the financial backing the company required as it grew.
2 A Simple Model: Compound Poisson Claims with No Payment Lag

2.1 Notation, the Compound Poisson Process, and the Loss Severity Distribution

The first step in understanding the expected losses, pricing needs, variance, required capital, or almost any other non-investment aspect of the financial operations of an insurance company involves understanding the potential aggregate loss costs to be borne by the company and their distribution. To do so, actuaries often make assumptions as to how those loss costs are distributed, and then successively broaden the assumptions so that they have a realistic and appropriate approximation that yields appropriate, useful guidance to the management of the company and other interested parties. This paper follows that model.

To understand the initial model one must understand the nature of insurance claims data that is the basis for developing and titrating actuarial models. Insurance company loss data often has three deficiencies that must be dealt with in analyzing loss or claims costs. These primarily arise from the fact that usually the claims costs are ultimately estimated using the actual (sampled, effectively) loss or claims data of the company itself. Firstly, for many lines of business, it may be several years until all the claims are settled, so the most current data is incomplete (it is generally accepted among actuaries that the larger claims tend disproportionately to settle late, so the claims size mix is biased as well). Secondly, since distributions of claims by size are often highly skewed upward, there is often an inadequate sample of the larger claims, often inadequate enough to materially distort the sample as a prediction of the expected loss. Lastly, claims costs are constantly changing both as a result of inflation and as a result of changing societal attitudes, law changes, judicial interpretations, etc. All of those present significant challenges that result in widespread use of mathematical models by actuaries when they analyze loss data in rate setting, evaluating reinsurance and advising management on reinsurance terms, and evaluating the riskiness of various operations of the company (including the holistic “enterprise risk management”).

In the interest of beginning the review of marginal risk, the simplest, most relevant model will be used . . . the compound Poisson claims distribution with no lag in claims payout. In this case, consider a company that writes (sells insurance to) “n” individuals (we would consider each to represent an exposure to loss for the one year term for one policy, and call each
an “exposure”). Assume that each exposure has an equal probability of generating claims, and that each claims occurrence is completely independent of that of all the other claims. So, if $\lambda$ represents the mean claim frequency per exposure, $n\lambda$ represents the mean claim exposure for the population as a whole. Also, the claim count (frequency) for a single exposure, under those assumptions, is Poisson distributed with parameter $\lambda$ and the group claim count is Poisson distributed with parameter $n\lambda$. Similarly, if one assumes that each of the individual claim costs “$X$” is distributed independently of, but identically to all the other claims, following a probability mass function $s(x), (0 < x < \infty)$, then the total costs generated by the group of exposures (the so-called “collective risk”), by construction, will be distributed according to a compound Poisson distribution with frequency parameter $n\lambda$ and jump size distribution $s(x)$. That represents the first major assumption in this model. The second assumption is that the mean $\lambda$ and the severity distribution $s(x)$ are fully known and specified. The last assumption is that the claims are paid on or near occurrence, so multi-year commitments of capital are not needed.

For purposes of this paper, a number of values based on the formal compound Poisson mass distribution must be computed or estimated. However, the mass function and cumulative distribution of a compound Poisson distribution are typically difficult to show in closed form or not amenable to easy simplification. It will be helpful to define notation to aid in the analysis, which will be used through the paper. If $N$ (a random variable, generated by a Poisson($n\lambda$) distribution) represents the number of claims the company experiences in a year, and if each individual claim, $X_i$, $i = 1, 2, \ldots, N$, is generated from the common severity mass function $s(x)$, then one may define the random variable representing the aggregate (total combined) cost of all the claims occurring in the year as

$$T = \sum_{i=1}^{N} X_i,$$

and $T$ will, by construction, be distributed according to a compound Poisson distribution with a Poisson parameter of $n\lambda$ and a jump size mass function of $s(x)$. It is the well known that the expected value of the aggregate cost in a year (or other period) is

$$E[T] = n\lambda E[X]$$

(2)

(which is a clear conclusion, given that the number of claims, and the size of each claim are mutually independent), and that the variance of the aggregate
costs is

\[ Var[T] = n\lambda E[X^2]. \]  

(3)

The formula for the variance is a little less obvious, but is a fairly direct consequence of the collective risk model. The collective risk formula, in standard form, says that (in this context, where the number of claims and the size of each are mutually independent)

\[ \sigma^2_T = \mu_N \sigma^2_X + \mu^2_X \sigma^2_N. \]  

(4)

The \( \mu \)'s and \( \sigma^2 \)'s may be expressed in alternate notation, and simplified using the formula \( Var[X] = E[X^2] - E^2[X] \) and characteristics of the Poisson distribution, to get

\[
\begin{align*}
Var[T] &= E[n\lambda Var[X] + E^2[X]E[(n\lambda)^2] = n\lambda Var[X] + n\lambda E^2[X] \\
&= n\lambda(E[Var[X] + E^2[X]]) = n\lambda E[X^2].
\end{align*}
\]  

(5)

So, beginning with those basic properties for the compound Poisson, one may also define values in the contexts where various caps and limits are imposed on individual claims in order to reduce the volatility of the company’s aggregate losses from year to year. Specifically, one could define “\( X_r(L) \)”, or \( X \) restricted to be less than a limit \( L \), by

\[ X_r(L) = \{ X, \text{ for } X \in [0, L]; 0 \text{ otherwise} \}. \]  

(6)

Further, if we set \( s_r(x) = s(x) \) for \( x \in [0, L] \), and add a point mass at zero corresponding to the probability under \( s(x) \) of a loss exceeding \( L \) (i.e., \( (1 - F_s(L))\delta(0) \)), where \( \delta(0) \) is a Dirac delta function), and set \( s_r(x) \) to be zero elsewhere, then \( s_r(x) \) is also a possible jump size distribution. So, one may define

\[ T_r(L) = \sum_{i=1}^{N} (X_i)_r(L) \]  

(7)

and \( T_r(L) \) will then be distributed according to a compound Poisson distribution with a Poisson parameter of \( n\lambda \) and a jump size mass function of \( s_r(x) \). So, one may conclude from equations (2) and (3) that

\[ E[T_r(L)] = n\lambda E[X_r(L)] \]

(8)

and

\[ Var[T_r(L)] = n\lambda E[(X_r(L))^2]. \]

(9)
Similarly, if the individual claims are capped at \( L \), then one may define

\[
X_c(L) = \{ X, \text{ for } X \in [0, L]; \ L, \text{ for } X > L; \ 0 \text{ otherwise} \}. \tag{10}
\]

Then \( X_c(L) \) has a severity distribution defined as \( s_c(x) = s(x) \) for \( x \in [0, L) \), with a point mass at \( L \) corresponding to the probability under \( s(x) \) of a loss exceeding \( L \) (i.e. \((1 - F_s(L))\delta(L))\), and \( s_c(x) \) defined as zero elsewhere. So, one may write

\[
T_c(L) = \sum_{i=1}^{N} (X_i)_c(L), \tag{11}
\]
and \( T_c(L) \) will also be distributed according to a compound Poisson distribution with a Poisson parameter of \( n\lambda \) and a jump size mass function of \( s_c(x) \). Further, as before, one may conclude that

\[
E[T_c(L)] = n\lambda E[X_c(L)], \tag{12}
\]
and

\[
Var[T_c(L)] = n\lambda E[(X_c(L))^2]. \tag{13}
\]

The various distributions in restricted and capped form will be used extensively throughout the remainder of this paper. Often, the derivatives of their mean and variance with respect to the cap \( L \) will be needed. So, it is helpful to compute

\[
\frac{\partial}{\partial L} E[X_r(L)] = \frac{\partial}{\partial L} \left\{ \int_{0}^{L} xs(x)dx + 0(1 - F(L)) \right\} = Ls(L), \tag{14}
\]

\[
\frac{\partial}{\partial L} E[T_r(L)] = n\lambda \frac{\partial}{\partial L} E[X_r(L)] = n\lambda Ls(L), \tag{15}
\]

\[
\frac{\partial}{\partial L} E[(X_r(L))^2] = \frac{\partial}{\partial L} \left\{ \int_{0}^{L} x^2s(x)dx + 0(1 - F(L)) \right\} = L^2s(L), \tag{16}
\]
and

\[
\frac{\partial}{\partial L} Var[T_r(L)] = n\lambda \frac{\partial}{\partial L} E[X_r^2(L)] = n\lambda L^2s(L). \tag{17}
\]

(By notation, \( X_r^2(L) \) will mean \( X \), first restricted to be less than \( L \), then squared. In the course of this paper, it will always be clear in context when some alternate meaning is intended. The restriction is always to apply first.)
For the values capped at $L$, the computations are only slightly more complex as

$$\frac{\partial}{\partial L} E[X_c(L)] = \frac{\partial}{\partial L} \left\{ \int_0^L x s(x) dx + L(1 - F(L)) \right\}$$

$$= L s(L) + 1 - F(L) - L s(L) = 1 - F(L), \quad (18)$$

$$\frac{\partial}{\partial L} E[T_c(L)] = n \lambda \frac{\partial}{\partial L} E[X_c(L)] = n \lambda (1 - F(L)), \quad (19)$$

$$\frac{\partial}{\partial L} E[(X_c(L))^2] = \frac{\partial}{\partial L} \left\{ \int_0^L x^2 s(x) dx + L^2(1 - F(L)) \right\}$$

$$= L^2 s(L) + 2L(1 - F(L)) - L^2 s(L) = 2L(1 - F(L)), \quad (20)$$

and

$$\frac{\partial}{\partial L} Var[T_c(L)] = n \lambda \frac{\partial}{\partial L} E[X_c^2(L)] = 2n \lambda L(1 - F(L)) \quad (21)$$

(By notation, the capping operator “c” is always to apply before the square, expectation, etc. is computed.) Those values may be thought of as the “marginal cost” ($\frac{\partial}{\partial L} E[T_c(L)]$) and “marginal risk” ($\frac{\partial}{\partial L} Var[T_c(L)]$) of the aggregate distribution with respect to the “limit” or “cap” “L”.

Another bit of notation should be defined. One may think of the mass function

$$f(x) = n \lambda s(x), \quad (22)$$

which defines the mass function of the expected number of claims of each size. Obviously, $\int_0^\infty f(x) dx = n \lambda$, and an alternate form of equation (21) is

$$\frac{d}{dL} Var[T_c(L)] = 2L \int \limits_0^\infty f(x) dx. \quad (23)$$

This set of notation and formulas provides the starting point for the analysis of percentiles of the aggregate distribution, their responsiveness to the limit $L$, and the weakening of the compound Poisson assumptions in the sections and chapters that follow.
2.2 The Capital/Surplus Commitment and the Marginal Capital/Surplus

The previous section presented some notation, and a basic mathematical starting point for an analysis of risk by the limit or cap employed by an insurance company and the consequent profit and surplus needs that will be shown to flow from those formulas. For this to be useful in terms of analyzing profit (and, as it will be shown, reinsurance attachment points), at some point the analysis must be explicitly shifted in scope from strictly the variance, which describes the properties of the aggregate cost distribution, to the consequent required capital or surplus derived from the variance\textsuperscript{12}. Of course, since in general the higher the variance the higher the range of possible adverse events, capital and surplus are logically highly related to the variance. So the variance, as discussed in the previous section, is a good starting point. But, as will be shown, the variance itself will only provide an approximation to the needed capital and surplus of an insurance company.

2.2.1 VaR, EPD, or TVaR? Choosing the Proper Risk Constraint to Define Capital and Surplus Needs

The next question to be answered, however, is how exactly the adequacy of the surplus (and implicitly, the needed capital and surplus) should be determined. The three most common approaches in use today generally involve “value at risk” (VaR), which deals with the value of capital needed to guarantee survival of the company within a certain probability threshold; “expected policyholder deficit” (EPD), which deals with the average amount of possible losses that would go unpaid in the event of a failure, multiplied by the probability of a failure; and “tail value at risk” (TVaR) which describes the average aggregate loss beyond the VaR standard from all aggregate loss scenarios that exceed the VaR standard (or equals the EPD, in context).

As discussed earlier, one school of thought is that the purpose of surplus is to protect against failure to pay claims, but (in most cases) the range of possible adverse outcomes is infinite. So, since it is neither possible nor reasonable to require a company to hold infinite surplus, an appropriate amount of capital must be determined. Under the “VaR” approach, capital/surplus

\textsuperscript{12}Although insurance accounting refers primarily to the funds securing an insurance company’s claims-paying ability in the event of higher-than-expected aggregate claims costs as “surplus”, from here out these funds will be referred to as the “surplus” or “free funds” portion of “capital” to denote that they represent the capital investment or continuation of a capital investment in the company.
to withstand all claims up to some pre-determined percentile of the distribution of possible aggregate claims costs is required. Implicitly, a specified “one year probability of ruin” is employed. This could also be described as having a “100(1 − p)% confidence level in the claims-paying ability” of the company over the next year, where 100p% represents the probability of the company having inadequate assets to pay all its claims within the next year. In that sense, since the capital/surplus is based on a benefit to insureds and claimants, this “confidence level for survival” approach is consumer-focused, which is desirable from a general business perspective.

Another commonly-used criterion for evaluating capital/surplus is the “EPD”, or “expected policyholder deficit” as defined by Butsic [8]. This statistic represents the expected value of the losses that will not be covered with a given amount of capital/surplus. So, it represents the one year probability of ruin discussed above, multiplied by the expected amount of the unfunded losses in the event of a failure (such failures are usually called “insolvencies” in the insurance sphere). This method differs from the “VaR” statistic in that it does not just measure the probability of an insolvency, but it also measures how severe the insolvency might be. Further, under the US guaranty fund system (as discussed in Han, Lai and Witt [43]), in many circumstances the failures of one insurer are recompensed by the other insurers, so this logically relates to the costs that are transferred from one insurer to another. Therefore, the EPD approach to solvency is more focused on the costs transferred to other insurers than on the reliability of claim payments to claimants.

The last common method is the “TVaR”, or “tail value at risk”. This statistic does not have a clear operational analogue in terms of insurance operational criteria as the VaR and EPD do, but rather it represents the average total size of an excess loss. It is most often cited because it has a “sub-additivity” characteristic (as noted in [24]) that is convenient for mathematical computations, and which results in what is described as a “coherent” risk measure.

Although there is no universally held view within actuarial circles, the mathematical advantages of the TVaR approach have won it many converts, and it is very widely used in some actuarial circles and by many students of risk theory. Nevertheless, there are very compelling reasons to use VaR in the common context of property/casualty (especially long-tail casualty insurance, where the costs of the larger claims tend to not be known until long after the policy expires) insurance company solvency. Three concerns clearly favor the VaR approach. First, as mentioned above, since it assesses the reliability of the insurance service (whether or not all the claims will
be paid), the VaR approach is more customer-centered. Next, one must note that the largest losses and the upper tail of any property/casualty loss distribution usually features rare, but outsized events. Because of their rarity, such events usually are very poorly represented in the data used in loss-based cost studies. So, the values related to extreme events are often heavily influenced by assumptions and as such prone to estimation error. Of note, the VaR approach only requires understanding the events that occur within a specified survival probability. To compute EPD or TVaR, values of all possible events, no matter how rare, must be used in the computation, and they are focused on extreme events. So, any errors in estimating the upper tail of the loss distribution will have an outsized impact on the estimated values of the EPD and TVaR statistics. On the other hand, the VaR values, in general, are less influenced by assumptions/estimation error and hence are more reliable. Lastly, since meeting a TVaR threshold does not appear to provide any explicit promise of solvency to customers or competitors, it does not provide a meaningful metric or performance criteria to the players in the system. So, those concerns support the VaR approach, specifically funding to secure a predetermined probability of survival. The VaR approach will then be used throughout this study.

So, in combination with the view that any insolvency is a bad thing to consumers\textsuperscript{13}, in this author’s opinion the confidence level is more customer-driven than alternatives such as TVaR\textsuperscript{14}. Also, as will be shown later, the confidence level approach leads to a much preferable analysis of the proper level of funding to require of insurers. Therefore, in the remainder of this dissertation the focus will be on the VaR “confidence level of survival” approach.

\textsuperscript{13}For example, when a claim is paid by a guaranty fund rather than the originating insurer, the claim may often to begin the claim process all over again, depending on what records were transferred from the failing insurance company to the guaranty fund. Further, there are often caps on claim amounts that after often lower than the amounts the insurance company would have paid. Further, as the guaranty fund may have marshall assets to pay claims before paying them, there may be long delays in claim payments from guaranty funds.

\textsuperscript{14}TVaR is equivalent to Butsic’s\textsuperscript{8} expected policyholder deficit. The expected policyholder deficit is the expected value of all unpaid claims cost after the company’s assets are exhausted (the expectation is taken across all loss outcomes, including the vast majority of outcomes where all the claims can be paid and there is no deficit).
2.2.2 Issues in Choosing the Funding Percentile and the Importance of Maintaining the Right Capital

As mentioned in the introduction, while insurance regulators may require that a company’s aggregate risk fall below some threshold, insurers are free to establish higher thresholds, and in doing so provide additional consumer confidence in their ability to pay whatever claims their policyholders generate. But that generates an operating trade-off. As stated in the introduction, an insurance company’s role in the economy is clearly to insure the risks of its policyholders. So, insuring more policyholders and covering more risk is very desirable from a sales standpoint. To the extent that the size of loss distribution and expected claim count are known rather than estimated, the compound Poisson model applies. Therefore, the law of large numbers and general properties of loss distributions dictate that while the standard deviation and variance increase with additional claim counts, the ratio of the standard deviation to the mean will decrease. But, as noted, the absolute value of the standard deviation increases with increasing expected claim counts. Further, if the loss distribution is not known, and is subject to so-called “parameter risk”, the relative parameter risk is generally not reducible through increased pooling. So a strategy of insuring more policyholders brings additional risks. And, barring extensive use of reinsurance or raising significantly more capital, the only way to ensure that the risk of insolvency/claims non-payment is low requires limiting the number of policyholders the company serves. Insurers must manage, in the end, a fourfold trade-off between writing additional business, the amount of reinsurance purchased, the capital held by the insurer, and the risk to the enterprise.

That leads to a concern with the amount of capital needed to meet the desired solvency threshold. The ratio of the capital held by an insurance company to the capital needed to support its policies, should it be high, can lead an insurer to under-price policies, as noted by this author in [6] and [7] and also in Cummins and Danzon [21]. Essentially, excessive or inadequate capital can drive a company, or a large portion of an industry, to either great success (by those who have capital) when capital is scarce, or unprofitability when capital is excessive. So it is relevant to understand the proper capital for an insurance company to reliably serve its policyholders. In the context of this dissertation, the focus will primarily be on those factors rather than on the allocation of capital\(^{15}\) to the various components of an insurer’s operations.

\(^{15}\)Of note, per Myers and Read [57] the EPD has some considerable value in allocating capital.
It should be understood, though, that approval of insurance regulators is often needed to dividend out excess capital. Further, raising capital can take time and be expensive. So, it may be prudent for an insurance company to target say “up to 100 + \( x \)% of the capital needed to fund 100(1 - p)% of all possible loss scenarios”.

2.3 Normal Approximation to the Percentiles and Their Capital and Profit Implications

The basic concept underlying the analysis of the various percentiles and related quantities is that they will vary with the loss limit used by the company. For illustration, one may begin by assuming the most simplistic case for the distribution of the aggregate losses \( T \) (as defined in Section 2.1). But first, a model to simplify the calculations will be introduced. Noting that a typical insurance company will often have a considerable volume of medium-to-large claims, one would expect the Central Limit Theorem to significantly force the aggregate loss distribution close to a normal distribution. So, this will begin with a model that assumes the distribution of \( T \) is close enough to a normal distribution to be treated as a normal distribution\(^{16} \). This would characterize the situation where the claims severity distribution is not too heavily skewed and there is a large body of data. It would likely also be a better approximation where the probability of ruin is relatively large. In this case, the confidence interval for the probability of ruin and the variance directly specify the required capital/surplus, since the shape of a normal distribution is exactly that of the standard normal (corrected for mean and variance). So, if \( 1 - p \) is the desired confidence level, one must have funds sufficient to cover losses in the amount of

\[
F_{Tc(L)}^{-1}(1 - p) \approx E[Tc(L)] + \Phi^{-1}(1 - p) \times Var^{1/2}[Tc(L)],
\]

using the normal approximation.

In a simplified fashion, insurance premiums cover expected losses, fully predictable\(^{17} \) expenses, and contain a profit load designed to provide an appropriate return on the capital deployed. So, the two sources of funds available to fund the adverse outcome with probability “\( p \)”, specifically to

\(^{16}\) More complex, but possibly more accurate approaches may be found in Panjer and Sharif [59], Pentikäinen [60], and Heckman and Meyers [44].

\(^{17}\) Although it is undoubtedly not absolutely true in fact that insurance overhead and sales expenses are fully predictable, they are nonetheless usually regarded as being predictable enough relative to losses to be viewed as non-random in insurance ratemaking (price setting).
fund unexpected additional losses of $F_{T_c(L)}^{-1}(1 - p)$, are the expected profit and the surplus portion of the capital committed to the line. So, using “P” to represent the expected aggregate profit (which is, at this point, independent of the capital/surplus) and $S(L, p)$ to denote the needed “free capital” or “surplus” funds for loss cap $L$ and probability of ruin $p$, we get the approximation

$$E[T_c(L)] + \Phi^{-1}(1 - p) \times Var^{1/2}[T_c(L)] \approx E[T_c(L)] + P + S(L, p), \quad (25)$$

or

$$P + S(L, p) \approx \Phi^{-1}(1 - p) \times Var^{1/2}[T_c(L)]. \quad (26)$$

That means that (tentatively holding the profit $P$ constant — the general case will be discussed later)

$$\frac{d}{dL}S(L, p) \approx \Phi^{-1}(1 - p) \times \frac{1}{2}Var^{-1/2}[T_c(L)] \times \frac{d}{dL}Var[T_c(L)]. \quad (27)$$

Using the result $\frac{d}{dT}Var[T_c(L)] = 2L \int_L^\infty f(x)dx$ from (21), one may determine that

$$\frac{d}{dL}S(L, p) \approx \Phi^{-1}(1 - p) \left[ \frac{1}{2L} \int_0^L x^2f(x)dx + \int_L^\infty f(x)dx \right] - \Phi^{-1}(1 - p) \left[ \frac{1}{2L} \int_0^L x^2f(x)dx + \int_L^\infty f(x)dx \right] \frac{n\lambda \int_L^\infty s(x)dx}{\int_L^\infty f(x)dx} \approx \Phi^{-1}(1 - p) \left[ \frac{n\lambda \int_L^\infty s(x)dx}{\int_L^\infty f(x)dx} \right]. \quad (28)$$

### 2.4 Quality of the Normal Approximation to the Percentiles

As long as the normal distribution mirrors the distribution of the possible aggregate losses, a target ruin probability $p$, and a target profit load, the marginal cost of capital/surplus with respect to the loss cap $L$ is determined.
But of course, the typical loss distribution is not normal. Should the book of business be small in relative terms and the class of business prone to out-sized large losses, the compound variables $T$ and $T_c(L)$ likely are highly skewed. To review the quality of the normal approximation to the aggregate loss distribution, it is helpful to define a function $G(L, p)$ representing the ratio of the true $100(1 - p)^{th}$ percentile of the aggregate loss distribution (with individual claims capped at $L$) $T_c(L)$ to the $100(1 - p)^{th}$ percentile of the normal distribution used to approximate $T_c(L)$. Specifically, $G$ is determined from the variance and mean of $T_c(L)$, the cumulative distribution functions of $T_c(L)$, and the standard normal cumulative distribution function $\Phi$ as

$$G(L, p) = \frac{F_{T_c(L)}^{-1}(1 - p)}{E[T_c(L)] + \Phi^{-1}(1 - p)Var^{1/2}[T_c(L)]}.$$  \hspace{1cm} (29)

It is helpful to evaluate just how well the normal approximation estimates the percentiles that underlie the surplus (and, as will be shown later, profit) needs of the company. The first step in doing so is to evaluate how substantive the “$G$” factor is in a practical context.

In light of that concern, some random sample data was reviewed. In constructing the sample, some standard, but inexact, actuarial assumptions were made, and then implemented in an exactly specified context. First, note that since insurance size of loss distributions (in this context $s(x)$) often clump at points such as $10,000$, $25,000$, $50,000$, it is entirely likely that actual loss distributions, whether $G(L, p)$ is high or not, have very large values of $\partial G(L, p)/\partial L$. But, recognizing that actual data is often an inadequate estimate of the true distribution of losses by size, and considering the need to replace the highly complex structure of the actual data with something simpler and more tractable, actuaries often work with fitted curves rather than actual data. Since most conclusions actuaries reach through analysis of loss distributions involve broad approximations about what is an optimal reinsurance structure, pricing structure, capital structure, etc., rather than precise estimates, that is practical. So, in lieu of analyzing every possible size of loss distribution, which clearly could produce almost any value of $\partial G(L, p)/\partial L$, it is appropriate to analyze one that has well-known mathematical properties, is in common use by actuaries, but still represents a case that has considerable (if not extreme) skewness and kurtosis (with consequent likelihood that $G(L, p)$ is far different than unity). Since the Pareto is recognized by actuaries as representing a distribution with a fairly large percentage of large claims, a Pareto distribution with parameter $\alpha = 2.2$
\( F_s(x) = 1 - \left( \frac{x_m}{x} \right)^{\alpha} \) was chosen. Further, the parameter \( \alpha = 2.2 \) creates a situation where the distribution \( s(x) \) has a finite mean and variance (though barely, \( \alpha = 2.0 \) generates a distribution with infinite variance), but infinite skewness and kurtosis. So logically, if \( G \) has minimal impact on the capital/surplus and marginal capital/surplus for that Pareto distribution, it should have minimal practical impact in most situations encountered by actuaries.

To perform that evaluation, the total claims costs capped at various possible claim limits and expected claim counts \( (n\lambda) \) were simulated\(^{18}\) using first a Poisson\((n\lambda)\) simulation of the number of claims occurring, then a simulation of each of the claims using the Pareto distribution\(^{19}\) with parameters \( \alpha = 2.2 \) and \( x_m = 5000 \). The results are summarized in Table 1.

It is worth making a few brief qualitative comments on the relationships present in the chart. One would expect that, given the Central Limit Theorem\(^{20}\), as the number of expected claim counts increases, the quality of the approximation increases so that the \( G \)'s are closer to unity. Further, since raising the limit from $25,000 to $50,000 and so forth increases the skewness

\(^{18}\) Simulation was done using the NtRand Excel plug-in (version 2), and 5 sample sets of 5000 subsamples. Each set was generated by two randomly chosen seeds. One seed was used to generate the random numbers for the number of claims in each of the 5000 subsamples. The other set was used by NtRand to generate 480 consequent random numbers. The \( i^{th} \) consequent random number was itself used as a seed to NtRand to generate the random numbers used to generate the 5000 individual \( i^{th} \) claims in each of the subsamples. The same two initial seeds were used for all the various expected claim count, limit, and confidence level computations within a given sample set. Derivatives within a sample set were computed using differences between \( G \) values at limits 10\% above and 10\% below (of course computed within the same sample set) each target limit \( L \). The limited mean, variance, and inverse cumulative probability distribution values for \( 1 - p = .95, .98, \) etc., were ultimately estimated as the mean of the values from the five sample sets. The derivatives of \( G \), giving consideration to the goal of evaluating an adverse scenario with respect to the size of the derivative, were set at the highest (\( \geq \) values from all other sample sets) of the five values.

\(^{19}\) \( x_m \) is referred to as a lower truncation and scale parameter. Since the Pareto distribution features zero probability of claims below \( x_m \), the value of 5000 was chosen as a reasonable splitting point between the high volume small claims that contribute only a small portion of loss, and the “meaningful” claims. So, considering that the smaller claims do not contribute greatly to volatility, this particular Pareto distribution provides a good example of what is most relevant (in terms of total costs and skewness) in insurance claims distributions.

\(^{20}\) Technically, the Central Limit Theorem only refers to situations where there are a fixed number of samples from a distribution, and in this case there are a random number of samples from the loss severity distribution. However, each aggregate loss distribution may be equivalently generated by \( n\lambda \) independent compound Poisson distributions with a Poisson parameter of unity.

22
of the severity distribution, they would be expected to increase the skew of the aggregate loss distribution for $T$ as well. Hence, the quality deteriorates slightly at the higher caps. Lastly, it can be seen that the worst performance is at the higher percentiles. Considering that it is relatively likely that $s(x)$ has a different asymptotic character than the $e^{-x^2}$ character of the normal distribution, this is also to be expected.

As one can see, for most scenarios $G$ is actually a very minor adjustment to the normal distribution estimate of the needed capital/surplus. The exceptions generally occur when the failure probability is very low (survival probability is relatively high) and at the larger claim size values (for exam-

---

Table 1: Values of $G$ per Normal Distribution Approximation for Various Mean Claim Amounts and Limits Using Compound Poisson Model

<table>
<thead>
<tr>
<th>$n\lambda$</th>
<th>$L$</th>
<th>$E[#\text{excess claims}]$</th>
<th>Values of $G(L, p)$ with $1 - p =$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>95%</td>
</tr>
<tr>
<td>60</td>
<td>$25,000$</td>
<td>1.739</td>
<td>1.007</td>
</tr>
<tr>
<td>60</td>
<td>$50,000$</td>
<td>0.379</td>
<td>1.010</td>
</tr>
<tr>
<td>60</td>
<td>$100,000$</td>
<td>0.082</td>
<td>1.015</td>
</tr>
<tr>
<td>60</td>
<td>$250,000$</td>
<td>0.011</td>
<td>1.021</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>1.003</td>
</tr>
<tr>
<td>150</td>
<td>$25,000$</td>
<td>4.349</td>
<td>1.004</td>
</tr>
<tr>
<td>150</td>
<td>$50,000$</td>
<td>0.946</td>
<td>1.005</td>
</tr>
<tr>
<td>150</td>
<td>$100,000$</td>
<td>0.206</td>
<td>1.007</td>
</tr>
<tr>
<td>150</td>
<td>$250,000$</td>
<td>0.027</td>
<td>1.013</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>1.007</td>
</tr>
<tr>
<td>400</td>
<td>$25,000$</td>
<td>11.596</td>
<td>1.002</td>
</tr>
<tr>
<td>400</td>
<td>$50,000$</td>
<td>2.524</td>
<td>1.002</td>
</tr>
<tr>
<td>400</td>
<td>$100,000$</td>
<td>0.549</td>
<td>1.003</td>
</tr>
<tr>
<td>400</td>
<td>$250,000$</td>
<td>0.073</td>
<td>1.006</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000$</td>
<td>0.003</td>
<td>1.006</td>
</tr>
</tbody>
</table>

---

21In reviewing these conclusions, it is important to understand that specific cells in the table, especially those combining higher limits and a low claim amount, could still be subject to considerable sampling error. For example, especially for some specific combinations of high limits and low expected claim counts, less than 3 claims above $1,000,000 are expected in the whole body of claims generated in a sample set of 5000 total loss simulations. So, some inaccuracy is expected. However, the general principle that $G$ is irrelevant in most practical situations still holds.
ple, the 1.221 G factor for the 60 expected claims $1 million cap combination at the 99.5% confidence level includes a cap of $1 million when the total expected aggregate loss is only $549,567.) So at the levels where one might wish to employ the cap G is much less material.

Actuaries would typically use these computations to estimate the needed capital/surplus, and due to uncertainties of parameter specifications and so forth, any estimate of the needed capital/surplus is bound to be imprecise anyway. Also, as will be shown later, this may also be used as a starting point for determining any optimal limit, but the typical problem of determining a limit (or reinsurance attachment) involves determining whether one of $250,000 or $500,000 (for example) should be used, rather whether one of $250,011 or of $250,012 should be used. Similarly, implications for profit will be considered later, but, based on the data above (from an extreme distribution), G will generally not have a meaningful impact there.

At the detailed level, for a low expected number of meaningful claims, such as 60, G is meaningful only at a combination of what would be a very high limit retained of $250,000 or $1,000,000, and then only for fairly high confidence levels of 99% and 99.5%. At a more common expected number of meaningful claims of 400 (and logically, by the compression towards a normal distribution induced by the Central Limit Theorem, for expected claim counts greater than 400), G is only really relevant for the $1,000,000 limit and the 99.5% confidence level (which is unlikely to be encountered in practice). Of note, this is based on one sample distribution, so the conclusions should be viewed in terms such as “for low numbers of expected claims”, rather than “for 60 expected claims”.

Of note, however, that analysis rests on the assumption that small values of G denote small differences between the perceived confidence level of the funding level estimated using the normal approximation and the actual confidence level that estimate represents. To evaluate the degree to which one may rely on the normal approximation to the 95th, 98th, . . . percentiles of the aggregate distribution, the error in the funding percentile introduced by the normal approximation at those confidence levels was introduced by taking \( (G - 1 + 2\sigma_{\text{est.}G})F_{T_e(L)}(1 - p) \) and dividing it by the derivative of \( F_{T_e(L)}(1 - p) \) with respect to \( p \). The results are shown in Table 2.

---

22 The derivative of \( F \) was determined using finite differences applied to \( F(.95), F(.98), F(.99), \) and \( F(.995) \). The values in the middle used the average of the finite differences in both directions, the outside values used on the finite difference with the one adjacent point. Also of note, the standard deviation of G, which was ultimately not extremely significant, was obtained as 1/5 of the observed standard deviation between five estimates of each value of G.
Table 2: Estimated Error in Confidence Level $1 - p$ for Various Expected Claim Count, Limit, and Confidence Level Combinations Using Compound Poisson Model

<table>
<thead>
<tr>
<th>$n\lambda$</th>
<th>$L$</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$25,000</td>
<td>0.40%</td>
<td>0.62%</td>
<td>0.43%</td>
<td>0.42%</td>
</tr>
<tr>
<td>60</td>
<td>$50,000</td>
<td>0.52%</td>
<td>0.77%</td>
<td>0.50%</td>
<td>0.43%</td>
</tr>
<tr>
<td>60</td>
<td>$100,000</td>
<td>0.68%</td>
<td>0.98%</td>
<td>0.66%</td>
<td>0.49%</td>
</tr>
<tr>
<td>60</td>
<td>$250,000</td>
<td>0.76%</td>
<td>1.30%</td>
<td>0.94%</td>
<td>0.61%</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000</td>
<td>0.09%</td>
<td>0.92%</td>
<td>0.76%</td>
<td>0.52%</td>
</tr>
<tr>
<td>150</td>
<td>$25,000</td>
<td>0.32%</td>
<td>0.56%</td>
<td>0.35%</td>
<td>0.38%</td>
</tr>
<tr>
<td>150</td>
<td>$50,000</td>
<td>0.37%</td>
<td>0.60%</td>
<td>0.42%</td>
<td>0.41%</td>
</tr>
<tr>
<td>150</td>
<td>$100,000</td>
<td>0.44%</td>
<td>0.77%</td>
<td>0.49%</td>
<td>0.43%</td>
</tr>
<tr>
<td>150</td>
<td>$250,000</td>
<td>0.75%</td>
<td>1.11%</td>
<td>0.76%</td>
<td>0.53%</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000</td>
<td>0.30%</td>
<td>1.07%</td>
<td>0.78%</td>
<td>0.49%</td>
</tr>
<tr>
<td>400</td>
<td>$25,000</td>
<td>0.22%</td>
<td>0.44%</td>
<td>0.21%</td>
<td>0.33%</td>
</tr>
<tr>
<td>400</td>
<td>$50,000</td>
<td>0.24%</td>
<td>0.39%</td>
<td>0.30%</td>
<td>0.37%</td>
</tr>
<tr>
<td>400</td>
<td>$100,000</td>
<td>0.35%</td>
<td>0.43%</td>
<td>0.31%</td>
<td>0.36%</td>
</tr>
<tr>
<td>400</td>
<td>$250,000</td>
<td>0.60%</td>
<td>0.68%</td>
<td>0.51%</td>
<td>0.44%</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000</td>
<td>0.43%</td>
<td>1.19%</td>
<td>0.84%</td>
<td>0.50%</td>
</tr>
</tbody>
</table>

Noting that in the observed scenarios $G$ is always greater than unity, so the confidence level is slightly less than the formula suggests, one may first review the values for $1 - p = 95\%$. The normal approximations to $F^{-1}(.95)$ in the example still have a true confidence level of over 94\%, so they may be relied on. The quality of the approximations to $F^{-1}(.98)$ are not quite as good, but generally have true confidence levels of around 97\%, and higher for larger counts and lower limits, so those may generally be relied on. For the 99\% confidence levels, it appears that the approximation is only reliable for situations with lower limits or higher claim counts. At the 99.5\% level, it appears that either a combination of lower limits and higher claim counts, or much higher claims counts than were contained in the sample are needed. But, in most situations where the limit bears a reasonable relationship to the number of counts and the tolerance is within the normal 99% or less confidence level, the normal approximation appears to be very good.
2.5 The Lognormal Approximation to the Percentiles

While the normal approximation to the upper percentiles of the aggregate loss distribution tested is of reasonable quality, one is sure to notice that it generally undershoots the mark (since the $G's$ are generally above unity). The reason is clear — the normal distribution is symmetric and is lacking skewness and kurtosis, whereas the aggregate claim distribution is affected by outsized claims from a highly skewed\(^{23}\) claim size distribution, and is somewhat affected by the limited skewness in the Poisson distribution generating the claim count. So, understanding that the following approach may be counterproductive with less severe claim size distributions, it makes sense to develop an alternate approximation using a skewed distribution that is almost as common as the normal distribution — the lognormal.

The approach suggested here requires simply using the already determined mean and variance of the distribution together with the method of moments curve-fitting algorithm. Using the stand ins \(\hat{\mu}\) and \(\hat{\sigma}^2\) for the parameters of the lognormal distribution being fitted, and \(\mu_{Tc(L)}\) and \(\sigma_{Tc(L)}^2\) for the mean and variance of the compound distribution, we have

\[
\hat{\sigma}^2 = \ln \left( \frac{\sigma_{Tc(L)}^2}{\mu_{Tc(L)}^2} + 1 \right), \tag{30}
\]

and

\[
\hat{\mu} = \ln(\mu_{Tc(L)}) - \frac{1}{2} \hat{\sigma}^2 = \ln(\mu_{Tc(L)}) - \frac{1}{2} \ln \left( \frac{\sigma_{Tc(L)}^2}{\mu_{Tc(L)}^2} + 1 \right) = \frac{1}{2} \ln \left( \frac{\mu_{Tc(L)}^4}{\sigma_{Tc(L)}^2 + \mu_{Tc(L)}^2} \right). \tag{31}
\]

Then, when seeking a percentile, say the 100\((1 - p)^{th}\) percentile, of the approximating distribution, one would do so as

\[
F_{Tc(L)}^{-1}(1 - p) \approx e^{\hat{\mu} + \Phi^{-1}(1-p)\hat{\sigma}}. \tag{32}
\]

Rather than substituting in the expressions for \(\hat{\mu}\) and \(\hat{\sigma}\) directly, it is first helpful to use the expression for the mean\(^{24}\) of the composite distribution being analyzed, \(\mu_{Tc(L)} = e^{\hat{\mu} + \hat{\sigma}^2/2}\). That means

\(^{23}\)On an uncapped basis, recall that the skewness and kurtosis of the claim size distribution is infinite.

\(^{24}\)This arises from the formula for the mean of a lognormal distribution in terms of its internal \(\mu\) and \(\sigma\) parameters.
\[ F_{T_c(L)}^{-1}(1 - p) \approx \mu_{T_c(L)} e^{\Phi^{-1}(1 - p) \mu_{T_c(L)}}. \tag{33} \]

Then the expression for \( \tilde{\sigma}^2 \) from equation (30) may be used to yield

\[ F_{T_c(L)}^{-1}(1 - p) \approx \frac{\mu_{T_c(L)}}{\sigma_{T_c(L)}^2 + 1} \exp \left( \Phi^{-1}(1 - p) \sqrt{\ln \left( \frac{\sigma_{T_c(L)}^2}{\mu_{T_c(L)}^2} + 1 \right)} \right). \tag{34} \]

However, that is still an extremely complicated expression. The first step is to define

\[ A(L) = \frac{\sigma_{T_c(L)}^2}{\mu_{T_c(L)}} + 1. \tag{35} \]

Then, using equations (12) and (13), one may get

\[ A(L) = \frac{n\lambda E[X_c^2(L)]}{n^2 \lambda^2 E^2[X_c(L)]} + 1 = \frac{E[X_c^2(L)]}{n\lambda E^2[X_c(L)]} + 1. \tag{36} \]

So, one may write

\[ F_{T_c(L)}^{-1}(1 - p) \approx \frac{\mu_{T_c(L)}}{\sqrt{A(L)}} e^{\Phi^{-1}(1 - p) \sqrt{\ln \left( A(L) \right)}}. \tag{37} \]

or, in this case, since \( \mu_{T_c(L)} = n\lambda E[X_c(L)] \),

\[ F_{T_c(L)}^{-1}(1 - p) \approx \frac{n\lambda E[X_c(L)]}{\sqrt{A(L)}} e^{\Phi^{-1}(1 - p) \sqrt{\ln \left( A(L) \right)}}. \tag{38} \]

which is then a moderately long, but tractable, estimate for the percentiles of the distributions capped at various values of \( L \).

Of course, the next step is to compute the values of \( G \) for the target claim size distribution and the various expected claim counts and capping limits used earlier. The formula for \( G \) is

\[ G(L, p) = \frac{F_{T_c(L)}^{-1}(1 - p)}{n\lambda E[X_c(L)] e^{\Phi^{-1}(1 - p) \sqrt{\ln \left( A(L) \right)}}}. \tag{39} \]

The \( G \) values computed using the formulas for the mean and variance, and the sampling-estimated values of \( F_{T_c(L)}^{-1}(1 - p) \) are shown in Table 3.
Table 3: Values of $G$ per Lognormal Distribution Approximation for Various Mean Claim Amounts and Limits Using Compound Poisson Model

<table>
<thead>
<tr>
<th>$n\lambda$</th>
<th>$L$</th>
<th>$E[#\text{excess claims}]$</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$25,000$</td>
<td>1.739</td>
<td>0.994</td>
<td>0.989</td>
<td>0.984</td>
<td>0.981</td>
</tr>
<tr>
<td>60</td>
<td>$50,000$</td>
<td>0.379</td>
<td>0.996</td>
<td>0.991</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td>60</td>
<td>$100,000$</td>
<td>0.082</td>
<td>0.999</td>
<td>0.995</td>
<td>0.992</td>
<td>0.993</td>
</tr>
<tr>
<td>60</td>
<td>$250,000$</td>
<td>0.011</td>
<td>1.002</td>
<td>1.010</td>
<td>1.019</td>
<td>1.031</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>0.982</td>
<td>1.000</td>
<td>1.032</td>
<td>1.100</td>
</tr>
<tr>
<td>150</td>
<td>$25,000$</td>
<td>4.349</td>
<td>0.998</td>
<td>0.997</td>
<td>0.994</td>
<td>0.992</td>
</tr>
<tr>
<td>150</td>
<td>$50,000$</td>
<td>0.946</td>
<td>0.998</td>
<td>0.997</td>
<td>0.996</td>
<td>0.993</td>
</tr>
<tr>
<td>150</td>
<td>$100,000$</td>
<td>0.206</td>
<td>0.999</td>
<td>1.000</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td>150</td>
<td>$250,000$</td>
<td>0.027</td>
<td>1.004</td>
<td>1.007</td>
<td>1.010</td>
<td>1.013</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>0.997</td>
<td>1.014</td>
<td>1.035</td>
<td>1.086</td>
</tr>
<tr>
<td>400</td>
<td>$25,000$</td>
<td>11.596</td>
<td>0.999</td>
<td>1.000</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>400</td>
<td>$50,000$</td>
<td>2.524</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>400</td>
<td>$100,000$</td>
<td>0.549</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>400</td>
<td>$250,000$</td>
<td>0.073</td>
<td>1.002</td>
<td>1.002</td>
<td>1.005</td>
<td>1.009</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000$</td>
<td>0.003</td>
<td>1.002</td>
<td>1.016</td>
<td>1.029</td>
<td>1.057</td>
</tr>
</tbody>
</table>

Two things are immediately obvious from the table of $G$’s. First, in this instance, the approximation error is much less than that of the normal approximation. In fact, the lognormal-based approximation is good enough that, except for the highest limits and the lower expected claim counts, it is immediately obvious from the table that the lognormal prediction results in no appreciable error in the survival probability.

One should be careful simply using the lognormal as a blanket replacement for the normal approximation. The data used in the sampling of the $F_{T(L)}^{-1}(1 - p)$ values was from a very skewed distribution. Less extreme loss size distributions may result in sufficiently less skewness to make the lognormal approximation significantly overshoot the actual percentiles $F_{T(L)}^{-1}(1 - p)$. In some circumstances, though, it may be helpful to view both and view the characteristics of the loss size distribution in order to determine an estimate for the upper percentiles. Of note, though, it must be stated that the normal and lognormal approximations take something very complex (computing the upper percentiles of an aggregate distribution —

28
and doing so at a wide variety of loss caps) and convert it to a fairly simple process involving expected claim counts and moments of the capped claim size distribution. Further, since it uses a value-at-risk approach, this model only requires data up to the amount of the cap. Hence it is not heavily dependent on assumptions about the upper end of the claim size distribution. This analysis creates a powerful tool for enterprise risk management.
3 Financial Considerations: Structure Considerations, Profit, and Reinsurance Program Design

3.1 Financial Structure of an Insurance Company

Before continuing the analysis of the compound Poisson model for potential aggregate claims costs, some background analyses of the financial structure within which insurance companies operate are in order. Specifically, some analyses of the surplus required to support operations, the profit needed to sustain surplus in the long run, and the optimal choice of a reinsurance retention are in order.

The first step in that process is to understand the general financial structure of an insurance company. Part of the lack of consensus on the proper profit loading for insurance companies stems from their structure as financial intermediaries. As such, they utilize money from all their customers to pay some of their customers (claimants); require capital from investors to secure against payments coming in higher than expected; earn investment income on both the funds supplied by their policyholder customers and on the capital supplied by the investors; and earn profits on both the passive investment of capital and their “insurance underwriting operations” of taking in premiums and paying out claims. An illustrative model of the underlying cash flows and operations may be found in Sturgis [67]. So, to truly understand how the financial operations of insurance companies affect investors and insurance customers, one must understand the capital and asset structure of insurance companies.

It is helpful to discuss the two aspects of an insurance company separately. In the spirit of being customer-centered, the insurance underwriting operations will be discussed first. They are best understood in terms of the flow of a dollar of premium through the insurance process. Not surprisingly, the flow begins with a customer purchasing an insurance policy. With that purchase, a commission often must be paid, and certain other expenses (such as the costs of reviewing each of the potential insurance customers to see whether they meet the company’s eligibility standards, entering the policies of the selected customers onto the company’s statistical records, premium tax to the state in which the policy is effective etc.) often must be paid. Therefore, if the company receives $100 from the customer, they typically only have $65-$95 left over after the up-front expenses. The costs associated with that $65-$95 are the ongoing administrative expenses associated

\footnote{Of note, insurance companies must also pay for operating expenses and (usually) commissions to salespeople, plus a few classes of miscellaneous expenses.}
with the policy (including possibly some general overhead of the company, and certainly some administrative expenses associated with administering claims), and of course the costs of the “losses”, or claims that must be paid. Many of those expenses occur throughout the life of the policy, and a large portion of the losses, especially on so-called “long-tail lines” (types of insurance contracts where there is a long lag between receipt of the premium and payment of the claims) will be paid long after the contract has run its course. So, there is an opportunity to grow the $65-$95 left after up-front expenses into somewhere between a modestly and a significantly larger sum through investments. However, the insurance company has a responsibility to make certain that it has the funds to fulfill its promise to pay the claims of its insured customers. Hence, most insurance companies (at least those that are reasonably sophisticated in the investment and asset-liability matching disciplines) will back the liabilities (the future costs of claims and expenses required after the up-front costs) with investment grade bonds that mature at roughly the same time as the liabilities come due. So the insurance underwriting operations take in premiums, invest funds in bonds until payments come due, and generate profits off the difference between the premium and the discounted costs of all the expenses and losses associated with the policies the company sells.

As a result of this process the customers receive value. The insurance company replaces an unknown risk of loss that has a relatively high standard deviation (that potentially could bankrupt the insured customer) with a certain predefined cost. Since the law of large numbers dictates that the standard deviation of a group of pooled risks will be a successively lower and lower percentage of the mean as more risks are added to the pool, the relative volatility of the whole pool of policies sold by the insurance company is less than the typical relative volatility of each risk on a standalone basis.

---

26 Even if the insurance company chose to invest in investments (such as, in this, even investment grade stocks) that offered higher returns, but at a higher risk that funds could decline to less than the cost of claims, in the United States there are solvency regulators in all or nearly all jurisdictions. In most circumstances, those regulators would not allow the insurance company to take such risks.

27 Technically, since most insurance companies offer installment premium payment plans, it may be mathematically superior to consider that premiums are not always received at the inception of the policy. However, the lag between policy inception and premium collection is usually nowhere near as long as that between policy inception and the associated liability-type expenses, so this is functionally disregarded.

28 Of note, that cost also has a higher expected value than the risk of loss, due to the insurance company's expenses.

29 This random risk of claims that occur refers solely to so-called “process risk. As will be discussed later, the typical insurance setting features is “parameter risk”, or the risk
An insurance company is a classic financial intermediary. It receives monies from policyholders, processes them, and returns them to the same class of policyholders. Its profit is derived from the spread between total funds received (including, implicitly, investment income) and the total funds paid out. Further, the lag between receipt of premiums and payment of claims allows the company to earn an investment “float”, which is at least partially returned to the policyholders through a reduced profit loading. So, an insurance company invests money for the indirect benefit of its customers, like typical financial intermediaries. Further, the value an insurance company provides is an intangible increase in economic security, like the intangible “enhanced return on invested funds” or “simplified processing of payments” provided by investment houses and banks. Aspects of the cash flows of insurance companies are discussed in Feldblum [33] and to some extent in Balcarek [2].

But, of course, that process assumes that all the elements of the underwriting operations occur at exactly their expected values. Laws of probability dictate that they almost always will do otherwise. Further, probability dictates that when losses in the aggregate are highly variable, the company’s actual final profit results will usually differ materially from their expected values. Further, even companies with relatively predictable losses will occasionally experience extreme profit fluctuations if the tail of the aggregate loss distribution is large enough. For that reason, in order to assure their customers and the public\footnote{The public at large has a stake in the ability of the insurance company to pay its claims as well as the insured customer. Businesses, automobile owners, and many professionals often must have insurance in order to obtain a license. Individuals relying on insurance customers, such as medical patients, other drivers, and employees of businesses, need to be certain that the insurance company can pay its claims should they become claimants.} that the claims will (up to a certain probability) be paid, the insurance company must set up a buffer fund, called their “surplus”\footnote{To be absolutely precise, a property/casualty insurance company operating in the United States generally must not only set up a “surplus” account, they must also set up a “loss reserve” liability for their unpaid claims that does not reflect the time value of money. Thus, the amount of the discount inherent in the loss reserves provides a buffer—as long as the loss reserves are adequate to pay the undiscounted costs of all the unpaid claims.}. Should results turn out exactly as expected, the profit will accrue to the surplus, from which it is sent as dividends to the investors or reinvested in the business. Should results turn out worse than expected, the...
surplus provides funds to backstop the premiums so that (unless results are very adverse) the promise that the company will pay all valid claims is still fulfilled. Views of the optimal amount of surplus may be found in Cummins and Nini [20] and Cummins and Sommer [19].

In that regard, the surplus represents the key capital stock of an insurance company. An auto manufacturer, for example, requires capital to invest in manufacturing facilities, tooling, and engineering. On the other hand, an insurance company requires only a pool of available funds (plus a comparatively smaller amount for office furnishings, equipment, and a computer system), which are themselves re-invested. As such, most of the capital of an insurance company is understandably more liquid (easier to sell) in the event demand falls off. An example of the view of required return for industrial companies may be found in Fama and French [30].

However, state insurance regulators have considerable influence on how much surplus is held and what classes of investments are held. For most solvent mid-range to large insurers, as long as the insurer’s bond portfolio (supplemented with the premiums that are in the process of being collected) is adequate to fund its expected liabilities, rating agencies and most state regulators often allow the remaining funds of the company to be invested in investment grade stock. Consequently, the vast majority of the capital of an insurance company actually earns a passive investment return, even if the insurance company sells no policies. This must be contrasted with that of, say, an auto manufacturer or some other industrial company, whose capital assets depreciate in value rather than grow, and additionally generate passive costs (real estate taxes, maintenance, etc.) rather than passive returns. One view of the required return for an insurance company may be found in Masterson [52]. Other useful sources on the accounting and financial structure of insurance companies include Troxel and Bouchie [69], Tarbell [68], Cabral and Feldblum [9], Nealon and Yit [58], and Bloomer [5].

So, although the accounting records of an insurance company show no distinction in how the assets are allocated, it is nonetheless very practical to think of an insurance company as holding two classes of operations operating out of two separate accounts. Matsuyama’s article [53] on the asset mix of a typical life insurer illustrates how insurers set up separate accounts in this fashion. The insurance underwriting asset account group involves the direct fulfillment of the insurance promise. Money is taken in, invested in investment-grade bonds, operating expenses and insurance claims are paid (barring adverse results) out of this account. When profits are generated, though, they are forwarded to the surplus or “capital” account. The capital account has two objectives. Its first objective is to backstop the underwrit-
ing account should losses be adverse. Its second objective is to generate passive investment returns. Given that the maturies of bonds match, it would be possible to even assign bond assets to lines of business. For a view of assigning capital funds to line of business, one may review Gründl and Schmeiser [42].

It is also important to reflect the reality that insurance companies sometimes elect to take measured risks in the underwriting/bond account. For example, an insurer might choose to take advantage of an upward sloping yield curve and invest so that its underwriting/bond portfolio comes due two years or so after its insurance liabilities. Such a strategy might produce additional profits or it might produce additional losses. Logically, since such a strategy is divorced from the basic insurance underwriting operations, the profit or loss from such activities should logically fall directly to the surplus account rather than be thought of as part of insurance underwriting operations. As will be shown later, given that the return and risk of the alternate strategy are consistent with market norms, a return appropriate for the risk will accrue to the investors, directly from the investment strategy employed. As such, the profit load indicated for underwriting operations should be neutral or unaffected relative to the investment strategies employed with the capital/surplus funds. This will become more apparent in a forthcoming section.

Of note, at first blush it might appear that, since most of the capital of a typical insurance company is invested in stocks, it is already returning an adequate yield to its capital providers. Therefore, one might posit that no further profit is required. Consider, though, that some of the assets (at least representing computer systems, furniture and equipment, and possibly real estate) do not generate investment income, and do depreciate. Further, the earnings from the company’s invested capital assets (dividends and capital gains only, should the capital be invested in stock) are taxed as part of the taxation of the insurance company’s profits before they could be remitted to the owner/investors. This is referred to as an “additional layer of taxation” relative to the profits simply being taxed once on receipt by the ultimate investor. Lastly, there is a debate within the actuarial profession as to whether the appropriate long term investment return to the investors of the insurance company should be adjusted to reflect the volatility of the insurance underwriting operations. That is because such volatility, at least by first principles, is totally unrelated to the stock market volatility which the Capital Asset Pricing Model or “CAPM” views as the basis for excess returns above the risk-free interest rate (see Sharpe [66], Markowitz [51], Fama and French [29], Fama and MacBeth [28], Ross [65], Roll [64], Chen, Roll and
3.2 The Required Return — A Model Incorporating Pure Volatility

Now that the general financial structure of an insurance company has been explained, it is important to understand conceptually what profit load is required to sustain operations. That ultimately is driven by the mathematically indicated required overall return. However, there is an ongoing discussion within the actuarial profession regarding the relationship between the profit that would reasonably be expected by investors and the diversifiable volatility of the returns from an insurance company’s insurance underwriting operations. Some appropriate resolution of that matter, or appropriately determined position is required in order to understand the indicated overall return.

Specifically, there are a number of actuaries and financial professionals that believe that an insurer’s required returns should simply follow the CAPM approach (or some related multi-factor model, as exemplified in Cummins and Phillips’ article [22]). There is another group that believes that a reward for pure risk volatility is required. Following the logic of the pure CAPM, since an insurer’s claims are (theoretically) completely fortuitous and unrelated to the stock market (or other common factors, for the most part), no loading for the risk of insurance underwriting operations is needed. Aspects of this approach were also addressed by Feldblum [34].

Following the risk loading concept requires a different approach. This faction (see an article by Meyers [56] mentioned earlier) looks at the indicated profit load from a customer value perception. They argue that the service provided by insurance companies is the transfer of risk, so that insurance companies deserve compensation for the residual risk (after the benefits of pooling they provide) associated with service they provide. Therefore, they argue, insurers deserve compensation for the risk they assume. With regard to this second view, one must in the end consider that the purpose of profit is to attract investment capital. So, logically, while the opportunity to earn profits is more related to the risk the insurance company’s insured customers perceive they are transferring to the company, the profit needed to sustain their business is primarily related more to the demands of investors. One must consider, though, that there are economic forces beyond CAPM, specifically that insurance company managers should seek to create
more returns than CAPM indicates. That view suggests that efficient treatment of risk should create an opportunity to add extra value, or “alpha”, for shareholders beyond the returns needed to attract capital. That forms the alternate view, and some decision between the two must be made in order to create a final formula for the benchmark profit needed by an insurance company.\footnote{It is important to note that specifying a benchmark return that is adequate is not a sufficient condition for a company to survive. It must, by dint of customer service and other business activities, earn the specified return. In the end, the primary factor in a company’s long run success will of course be how well it operates in its chosen market.}

That does not, however, completely divorce the needed profit from risk. Prominent mention should be made of the fact that, as discussed earlier, the needed capital to run an insurance company with a $100(1 − p)\%$ probability of survival is strongly related to the standard deviation of its residual risk. Therefore, the more residual risk, the more capital needed. Since the profits are partially earned from the insurance underwriting operations, the more residual risk, the more capital needed in relation to sold “premium” (sales, but said in insurance terminology), which means that the premium must generate higher profits to support the corresponding capital).

There is also a view that in spite of the absence of pure diversifiable variance from most stock pricing models, it still matters to some or many investors. Proponents would argue that investors carefully select the individual stocks within their portfolios, and in the process consider the volatility on a stock-by-stock basis. In this view, there should be a certain market premium for diversifiable volatility (the volatility that is uncorrelated with the market as a whole, and hence amenable to reduction through diversification of stock holdings across a variety of different companies) as well as intrinsic volatility (the volatility of the market as a whole, which is, in light of the market representing all investment options, does not allow for further diversification).

Therefore, it makes sense to attempt to evaluate the degree to which the market might be rewarding diversifiable volatility. In the interests of obtaining the reward for the added volatility, a review of the well-known characteristics in broadly popular models (as described in \cite{66, 29} and \cite{14}) of stock pricing was done. The only characteristic factor that relates to diversifiable volatility, even indirectly, is the “Small minus Big” factor in the Fama-French model. While the premium for investing in small-cap stocks could also be due to items such as the lack of information imposed by the limited resources of analysts covering small-cap stocks, one could also make a point that small-cap stocks should be intrinsically more volatile.
So, in that view, the small-cap premium is partially or wholly due to that increased volatility.

To evaluate which view is most correct, a data analysis comparing returns of stocks with low and high diversifiable variance was performed. First, due to the fact that many insurance companies do not fully analyze their loss reserves, and hence their full underwriting results, until the year-end accounting, it was determined that annual data should be analyzed for this study. Next, the “Small minus Big” (“SMB”) factors on Dr. Ken French’s web page, as well as the market excess return and risk free return were converted into yearly returns. Three time periods were analyzed, 1963-2010 year ends (all available data at the time of the analysis), the years\(^{33}\) 1963-2005, and 1981-2005, to reflect more current data. For each set of points the portion of the returns that was explained statistically by the market excess returns for the same time period was removed, creating a set of returns that are orthogonal to the market excess return across each time period. The mean returns for the portion of SMB that were truly uncorrelated to the market return are shown below.

Table 4: Historical Average SMB Returns—Portion Uncorrelated with the Market

<table>
<thead>
<tr>
<th>Year Ends Used</th>
<th>Mean Uncorrelated to Market SMB Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1963 – 2010</td>
<td>1.75%</td>
</tr>
<tr>
<td>1963 – 2005</td>
<td>1.45%</td>
</tr>
<tr>
<td>1980 – 2005</td>
<td>0.29%</td>
</tr>
</tbody>
</table>

Consider that the most recent data point of the two that exclude the unusual 2008 market activity shows a very small premium for the truly uncorrelated portion of SMB. So, there is arguably a question of whether or not investor sentiment has changed to eliminate SMB. But, it is also true that the values are slightly inconsistent. Overall, this suggests that the pure

\(^{33}\)This was designed to omit 2008, since there was a precipitous decline in stock prices, less so for small than large stocks, in 2008. But, it is generally acknowledged that a large factor behind this decline was a lack of transparency in financial statements and a lack of communication from corporations to their shareholders about how the companies were run. So, there is strong reason to believe that the downturn of 2008 was not really the sort of downturn that investors accepted when they purchased the stocks. Therefore, some calculations excluding 2008, rather ending at the previous multiple of five (2005) were done.
uncorrelated (to the market return) portion of SMB is likely somewhere between twenty-five and one hundred seventy-five basis points.

That presents the information for the average return. But how that excess return is expressed as a function of the variance must still be analyzed. The first step involves establishing the mathematical form of the relationship. For reference in doing so, it is helpful to describe the CAPM approach in somewhat non-standard terms. The estimate of the excess return in the CAPM approach is generally expressed as the market excess return times the “β” factor, where $\beta = \frac{\rho_{stock,market} \sigma_{stock}}{\sigma_{market}}$. But note that it is also true that $\beta$ is also equal to the square root of the shared variance component of the stocks returns divided by the market standard deviation $\beta = \sqrt{\frac{\rho^2 \sigma^2_{stock}}{\sigma_{market}}}$. So the excess return expected for a stock may be written as (Stock’s Excess Return) = ((excess return of market)/$\sigma_{market}$) × (Square root of shared variance component).

In other words, a stock’s required return may be computed by multiplying a normalized (per unit of market standard deviation) market return on standard deviation, times the square root of the portion of the stock’s variance (of returns) that is shared with the market. As summary, the portion that involves the square root of a shared variance, as the square root of a variance, should be thought of as a shared standard deviation. So, logically the expected excess return for a given stock is determined by applying a market loading of amount (excess return of market)/$\sigma_{market}$ on the standard deviation component the stock shares with the market. Therefore, when evaluating any loading for the uncorrelated (with the market) volatility of a stock, the most consistent approach is to assume that each stock’s profit loading would be computed by multiplying some constant loading factor for uncorrelated returns by the standard deviation of the uncorrelated portion of the stock’s excess returns.

To estimate that loading factor, estimates of both the SMB excess return and the standard deviation of each are needed. But, to compute the standard deviation difference underlying the SMB excess returns, it is not relevant to consider the variance of the SMB data (since that has had the benefit of pooling). So, the 1995-2005 and 1995-2010 annual returns of twenty randomly selected$^{34}$ small company stocks and twenty randomly selected

---

$^{34}$Selection protocols were as follows. The first forty stocks returned alphabetically to a query for NASDAQ stocks with under $one billion in capitalization for which data back to 1995 was available were selected for the small company stocks. The first forty stocks returned alphabetically to a query for NYSE stocks with over $twenty billion in capitalization for which data back to 1995 was available were selected for the large company stocks. Of those, the first twenty meeting the longevity criteria were chosen. Considering
large company stocks were computed. Then the portion of each stock’s excess (over the risk-free return per Fama-French) return over the ten year period 1995-2005 and the fifteen year period 1995-2010 that was orthogonal to the market excess return in each period was computed. The variance of each stock’s uncorrelated returns was computed. The excess returns were summarized by group, along with the difference in the diversifiable excess returns, for each period. The results are shown below.

Table 5: Conversion of Uncorrelated SMB Returns to Load on Standard Deviation of Diversifiable Risk

<table>
<thead>
<tr>
<th>Data Used</th>
<th>“Small”</th>
<th>“Large”</th>
<th>Difference</th>
<th>Likely Return</th>
<th>Implied S.D. Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 – 2005</td>
<td>68%</td>
<td>28%</td>
<td>40%</td>
<td>0.29%</td>
<td>0.7%</td>
</tr>
<tr>
<td>1995 – 2010</td>
<td>63%</td>
<td>29%</td>
<td>36%</td>
<td>1.75%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

In conclusion, the data analyzed suggests\(^{35}\) that a load in a range of around 1-5% would be priced on the standard deviation of diversifiable risk. This is also contrasted with the CAPM approach, where the load on the shared standard deviation (between a stock’s returns and those of the broad market) is in the range of 25-45%. So, in the rest of this paper a load on diversifiable risk will be assumed to be required to satisfy investors. In the interest of distinguishing the diversifiable risk loading ratio from probability (“\(P\)” and “\(p\)”), a cumulative distribution (“\(F\)”), etc. a neutral name of “\(\nu\)” will be used for this loading ratio on the standard deviation of diversifiable returns.

This also has implications for the true excess return underlying the heavy incidence of stocks with a shorter history, this data has impacts of unknown magnitude from survivorship bias.

\(^{35}\)Of interest, the actual differences in the mean returns of the small stocks selected and the large stocks selected was about 9% over 1995-2005 and about 7% over 1995-2010, which must be considered in light of a much higher volatility between the returns of various stocks (an 18% standard deviation, for example, for the fifteen year small stock average returns between companies—4% standard deviation of the mean across all companies in the sample). This indicates a higher loading on standard deviation. But, unless there is corresponding evidence that some portion of what the SMB represents acts to reduce the excess returns generated by SMB, it should represent a cap on the total excess return difference used in computing the loading ratio. Note also that the small stock data is more volatile than the large stock data, which would suggest that the small stock perceived additional returns would be significantly more enhanced by survivorship bias.
CAPM approach. Since the standard deviation is never negative, the return of the universe of investments available must already include some portion of the loading on diversifiable standard deviation. Therefore, technically the overall market return should be decomposed into a portion related to the sum of all individual stock’s diversifiable standard deviation loads, plus a remainder which would represent the true market return on non-diversifiable risk. However, in the context of this paper, a brief review of the forthcoming algebra of the overall return provided to stockholders of an insurance company will affirm that such a decomposition would not affect the results.

3.3 The Indicated Profit Loading in the Rates

In computing the needed benchmark profit ratio on sales (the profits to be earned on underwriting operations, as discussed in Section 3.1) for an insurance company, one must compute the total profit needed so that the combination of the profit earned on sales (after tax) from insurance underwriting operations plus the profits earned from passive investment of capital (also after tax) sum to the needed total return. Also, the total target would have to reward the company’s investors for all forms of return that are intrinsic to the stock market. The three relevant returns to investors on their capital investment in the insurance company would include the risk-free return, the loading for market risk, and the loading for diversifiable risk discussed earlier.

To begin to discuss the required return, one must first relate the capital required to the premium sold (technically, in this scenario, required capital is actually generated by the losses insured). Recall that the needed surplus for an aggregate loss in the variable amount $T_c(L)$ was defined in conjunction with the profit by equation (26) as

$$P(L, p) + S(L, p) = E_{T_c(L)}^{-1}(1 - p) - E[T_c(L)].$$

(40)

Since the required profit will relate to the surplus commitment, a small system of equations is involved to separate the profit and surplus. Also, to truly understand the profit load and marginal profit required in operating

\footnote{As will be shown in Section 3.3 equation (52), any indicated return beyond the risk-free rate is simply earned (or expected to be earned) by the portfolio and passed through after dilution for taxes and non-interest earning assets. So whatever return is expected by the market, however it is decomposed, will be passed through. Should a beta between insurance losses and the stock market be determined, the decomposition of market returns into premiums for diversifiable and intrinsic risk may be needed.}

\footnote{As noted earlier, profits are not per se given . . . a company must earn them.}
an insurance company, it is necessary to understand some of the details of insurance company operations and costs a little better.

First, earlier in this paper, in the interest of brevity, it was implicitly assumed that all of the insurance company’s capital was invested in liquid assets such as stocks and bonds. That is true to the extent that most of an insurance company’s capital is invested in liquid assets. It also reflects the key knowledge needed to understand the concepts presented earlier. But it is not entirely true. A portion of an insurance company’s capital assets will be tied up in the costs of computer systems and hardware, furniture and equipment, and real estate. That portion of the assets is referred to as “non-admitted” (not counted towards solvency) assets, so (since $N$ is already in use), their aggregate amount will be referred to as $U$ (for “un”-admitted). Hence, the total capital invested will be counted as $C = S + U$.

Next, as mentioned previously, taxation is a key issue (see Feldblum [32] for a more extensive discussion of property/casualty insurance company taxation, as well as Feldblum [37] and van der Meer and Smink [70] for a more detailed study of investment strategies of insurers). But it is also true that the surplus funds are taxed generally (excluding dividends) at the capital gains rate, whereas underwriting profits are taxed at the full rate. Hence, one must define $X_A$ as the tax rate applicable to the passive returns on assets backing the surplus, and $X_U$ as the tax rate on underwriting profits.

It is also necessary to introduce symbols for the various key interest rates. Following standard notation, $r_F$ denotes the risk-free interest rate and $r_M$ represents the excess return of the market as a whole. As indicated before, factors such as SMB and the $\nu$ factor for standard deviation would act to reduce the pure market return. However, later in this section, the analysis will show that the CAPM return is not needed to compute the required profit load in most circumstances. Of course $\nu$ continues to represent the factor loading on standard deviation uncorrelated to the market.

---

38In practice, the return on the bond investments supporting the underwriting returns are taxed, too. However, in this model, at this point in the paper, release of profit at the time a policy is sold is assumed. Further, the underwriting returns are assuming to be computed on a discounted basis. Hence, the “unwinding” or progress from discounted values to the final nominal values of the losses (which reduces profits in a year), given a similar interest rate on the bonds as is used in discounting losses, will offset the investment income on the bonds backing the liabilities (which would otherwise increase profits). It is possible in practice to marginally improve the profits of an insurance company by investing in tax-free (or tax-advantaged) municipal bonds that earn investment income equal to the company’s profit from insurance underwriting operations, but that is beyond the scope of this paper.
The CAPM posits that the required return on a portfolio of investments that is needed to satisfy investors is equal to the risk-free rate \( r_F \), plus a factor “beta”, or \( \beta \), times the return of the market as a whole less the risk-free rate (the “excess return”, \( r_M \), of the market). The \( \beta \) of the portfolio is computed by first computing the covariance of the portfolio’s excess returns with those of the market as a whole. In equation form,

\[
\beta = \frac{\text{Cov}[r_A, r_M]}{\text{Var}[r_M]},
\]

where \( r_A \) represents the excess returns of the portfolio of assets invested. The indicated required excess return on portfolio \( A \) is then given by

\[
E[r_A] = \beta E[r_M],
\]

and the total return required for portfolio \( A \) (per CAPM) would be \( E[r_F] + \beta E[r_M] \). The excess return equation (42) may also be restated in terms of the standard deviation shared between \( A \) and the market by defining a loading \( l_M \), where \( l_M = E[r_M]/\sqrt{\text{Var}[r_M]} \). Specifically, one may write

\[
E[r_A] = \beta l_M V a r^{1/2}[r_M] = l_M \frac{\text{Cov}[r_A, r_M]}{\text{Var}[r_M]} V a r^{1/2}[r_M] = l_M \frac{\text{Cov}[r_A, r_M]}{\text{Var}[r_M]} V a r^{1/2}[r_M]
\]

\[
= l_M \text{Var}^{1/2}[r_A] \rho[r_A, r_M] = l_M \sqrt{\text{Var}[r_A]} \rho^2[r_A, r_M]
\]

\[
= l_M \times \text{(stand. dev. of variance component shared by } r_A \text{ and } r_M),
\]

where \( \rho \) is the standard notation for the correlation. Therefore, the excess return on a stock may also be characterized as a market loading factor times the standard deviation of the variance component that is shared with the market.

For the market loading, \( \beta_S \) will denote the beta of the invested surplus, \( \beta_C \) will denote the beta on invested capital, and \( \beta_U \) could represent the beta of the insurance underwriting results. In this case, however, as fully random events are unrelated to the stock market we implicitly infer that \( \beta_U = 0 \). A somewhat alternate approach may be found in Bingham [4].

Lastly, it will be necessary to introduce a few symbols used in property/casualty ratemaking. Generally, loss or claims costs are denoted by “L”, but \( L \) is already in use (as is \( C \)), so \( T \) will be used in the tradition of its use earlier in this paper within the context of the probability distribution of the total losses. Expenses must be considered as well. Insurance cost structures have a relatively novel feature in that most insurance salespeople are paid entirely by a fixed percentage of premium called commission, which
is sometimes a very significant portion of the premium. Of less prominence, most premium taxes, unlike most sales taxes, are actually included within the premium rather than tacked on at the point of sale. These constitute what are called “variable expenses”, which float up and down directly and proportionately to any change in insurance pricing. Although many actuaries treat all expenses as fluctuating with premium, there are others that treat flat costs that do not vary with premium, such as maintaining a computer system, as so-called fixed expenses\textsuperscript{39}. Following actuarial convention, one would set $VE$ as the variable expense in the aggregate, $VER$ as the ratio of variable expense to premium, $FE$ as the aggregate fixed expense costs expected\textsuperscript{40}, and $FER$ is generally undefined since fixed expenses do not float with premium. A more detailed explanation of the ratemaking equations may be found in McClenahan [54]. Then the following equation applies, either for pricing in the aggregate with $FE$ representing aggregate costs, or for individual policies where $FE$ is fixed expenses per policy, in computing the needed premium:

$$\text{Premium Charge} = \frac{E[T] + FE}{1 - VER}. \tag{44}$$

The typical actuarial practice at the time of this writing is different from this. In most actuarial ratemaking analyses today, profit is subsumed into variable expenses. But, one may see that, with respect to this analysis, profit relates solely to the volatility of the loss, and (as will be seen), the capital commitment, which itself relates to the volatility of the loss. However, assuming a fixed portfolio of losses, the profit load should logically be treated as a fixed expense. That yields the corrected equation

$$\text{Premium Charge} = \frac{E[T] + P + FE}{1 - VER}. \tag{45}$$

To complete the notation, one should first note that the loss content above is for total losses. Next, recognizing that a focus of this paper is on management options that involve using features that limit losses, one should consider reinsurance, or the transfer of losses from the insurer to a

\textsuperscript{39}There is no intention here to present a view or belief that any particular expense should be treated as fixed or variable. Such an approach might require different short-term and long-term views and could introduce much complexity. The intention here is simply to acknowledge that at least some expenses are variable and that some may be treated as fixed in some circumstances.

\textsuperscript{40}Technically, these would typically be the fixed expenses per policy, as would the variable expenses.
reinsurer. Explicitly, if the insurer transfers all risk above, say, $L$, without limit, it transfers $E[T] - E[T_c(L)]$ to the reinsurer in loss content. The reinsurer charges the $E[T] - E[T_c(L)]$ in return, plus a loading denoted by $Q(L, \infty)$ (attempting to use a variable name not used elsewhere). That results in a pricing equation of

\[
\text{Premium Charge with Policy Limit of } L = \frac{E[T_c(L)] + P(L, p) + FE}{1 - VER}, \quad (46)
\]

or

\[
\text{Premium Charge} = \frac{E[T] + P(L, p) + Q(L, \infty) + FE}{1 - VER}, \quad (47)
\]

where the $P(L, p)$ term reflects the potentially different profit load that is required when the insurer’s potential losses are individually capped at $L$, and the required surplus reflects a specific funding percentile determined by “$p$”.

Then, the first step is to compute the total indicated dollar profit, given exactly the surplus $S(L, p)$ needed for a cap of $L$ on a group of policies with certainty $1 - p$. We can see that the pre-tax profit earned by the company will simply be the return on invested surplus plus the underwriting profit\(^{41}\) across all policies. So, if the total return at some cap $L$ and funding level $p$ is defined as $R(L, p)$, and the portfolio backing the surplus is invested in the broad market\(^{42}\), specifically, a broad market mirroring the market as a whole, then the pre-tax return would be $R(L, p) = (r_F + r_M)S + P(L, p)$.

But, taxes are involved, so

\[
R(L, p) = (r_F + \beta S)M(1 - X_A)S(L, p) + (1 - X_U)P(L, p). \quad (48)
\]

Consequently, the rate of return on all capital generated by all operations is

---

\(^{41}\)In the interests of full disclosure the term “underwriting profit” is technically used in the property/casualty actuarial and insurance accounting context to refer to the underwriting result of a given year on an undiscounted basis. The term is used in the discounted context here because it refers to the profit from “insurance underwriting operations” which presumably involve handling policies, paying claims, and some fairly straightforward investment policies. Note that in this context, some basic investing is subsumed into operations, but more advanced investing, such as the effect of duration mismatch, is presumed to be included in the investment returns.

\(^{42}\)It will later become evident that this assumption is superfluous.
Now, by that equation, and since the non-admitted assets $U$ are not invested in securities (hence $\beta_U = 0$), we can infer that (since underwriting results are assumed to be purely random and hence uncorrelated to the market as a whole in the examples in this paper)

$$\beta_C = (1 - X_A)\beta_S \frac{S(L,p)}{C} + (1 - X_U) \frac{P(L,p)}{C}. \quad (50)$$

That means that the CAPM approach supplemented with the adjustment for the diversifiable standard deviation would require

$$\frac{R(L,p)}{C} = r_F + (\beta_S r_M)(1 - X_A) \frac{S(L,p)}{C} + (1 - X_U) \nu Var^{1/2}[T_c(L)]. \quad (51)$$

Alternately, the equation for the total return is

$$R(L,p) = r_F C + (\beta_S r_M)(1 - X_A) S(L,p) + (1 - X_U) \nu Var^{1/2}[T_c(L)]. \quad (52)$$

So, combining equations (48) and (52), one may conclude (simplifying $S(L,p)$ as $S$ and $P(L,p)$ as $P$ for clarity)

$$(r_F + \beta_S r_M)(1 - X_A) S + (1 - X_U) P$$

$$= r_F C + (\beta_S r_M)(1 - X_A) S + (1 - X_U) \nu Var^{1/2}[T_c(L)], \quad (53)$$

which leads to

$$P = r_F \frac{C - (1 - X_A) S}{1 - X_U} + \nu Var^{1/2}[T_c(L)]. \quad (54)$$

Therefore, the absolute required dollar profit is equal to the investment drag associated with taxation of the risk-free rate and the investment of a portion of capital into non-income bearing investments, plus a separate portion related solely to the standard deviation of aggregate losses. As promised earlier, the exact $\beta_S$ and investment classes for the surplus are not
relevant and the income tax rate for underwriting is only relevant insofar as it relates to recovering the investment drag on capital.

Last, to finish the calculation of the capital and profit, we require that

\[ S(L, p) + P(L, p) = F_{T_c(L)}^{-1}(1 - p) - E[T_c(L)]. \]

So, one may write that as

\[ F_{T_c(L)}^{-1}(1 - p) - E[T_c(L)] = S(L, p) + P(L, p) \]

\[ = r_F \frac{C - (1 - X)S}{1 - X_U} + S + \nu Var^{1/2}[T_c(L)] \]

\[ = r_F \frac{S + U - (1 - X)S}{1 - X_U} + S + \nu Var^{1/2}[T_c(L)] \]

A little algebra will show that solving for \( S \) and \( P \) yields

\[ S(L, p) = \]

\[ \frac{\left( F_{T_c(L)}^{-1}(1 - p) - E[T_c(L)] - \nu \times Var^{1/2}[T_c(L)] \right) (1 - X_U) - r_F U}{1 + r_F X - X_U}, \]  
(56)

and

\[ P(L, p) = \]

\[ \frac{\left( F_{T_c(L)}^{-1}(1 - p) - E[T_c(L)] \right) r_F X_A + \nu (1 - X_U) \times Var^{1/2}[T_c(L)] + r_F U}{1 + r_F X - X_U}. \]  
(57)

That is a closed-form expression for the indicated profit.

### 3.4 Approximating the Indicated Profit and Surplus

The expression in the previous section has an important bearing for the marginal profit. But it is not very tractable to evaluate. One could begin by using the normal approximation to the percentile to get

\[ P(L, p) \approx \frac{[\Phi^{-1}(1 - p)r_F X_A + \nu (1 - X_U)] \times Var^{1/2}[T_c(L)] + r_F U}{1 + r_F X - X_U}, \]  
(58)
and the lognormal approximation to get

\[
P(L, p) = \frac{\left( \frac{\mu_{T_c(L)}}{\sqrt{A(L)}} e^{-\frac{1}{2}(1-p)\ln\left(\frac{A(L)}{\mu_{T_c(L)}}\right)} \right) r_F X_A + \nu (1-X_U) \times Var^{1/2}[T_c(L)] + r_F U}{1 + r_F X_A - X_U},
\]

where

\[
A(L) = \frac{\sigma^2_{T_c(L)}}{\mu^2_{T_c(L)}} + 1.
\]

As discussed earlier in this paper, in the specific case of the compound Poisson scenario, one would know that

\[
\mu_{T_c(L)} = n\lambda E[X_c(L)],
\]

\[
Var[T_c(L)] = \sigma^2_{T_c(L)} = n\lambda E[X_c^2(L)],
\]

and consequently

\[
A(L) = \frac{n\lambda E[X_c^2(L)]}{n^2 \lambda^2 E^2[X_c(L)]} + 1 = \frac{E[X_c^2(L)]}{n\lambda E[X_c(L)]} + 1.
\]

When one considers the general low errors induced in practical circumstances by the normal and lognormal approximations to the percentiles of \(F^{-1}_{T_c(L)}\), one may see that equations (58) and (59), together with equations (61), (62), and (63), create quality estimates of the long-range profit needed to support the surplus and volatility of an insurance company. In a similar vein, one may note that the needed surplus of the company (per equation (56)) is

\[
\left( \Phi^{-1}(1-p) - \nu \right) \times Var^{1/2}[T_c(L)] (1-X_U) - r_F U
\]

\[
1 + r_F X_A - X_U,
\]

per the normal approximation; and

\[
\left[ \left( \frac{\mu_{T_c(L)}}{\sqrt{A(L)}} e^{-\frac{1}{2}(1-p)\ln\left(\frac{A(L)}{\mu_{T_c(L)}}\right)} \right) - \nu \times Var^{1/2}[T_c(L)] \right] (1-X_U) - r_F U
\]

\[
1 + r_F X_A - X_U,
\]

per the lognormal approximation.
3.5 The L-Derivative of Profit and an Application to the Specific Case of the Compound Poisson

As will be seen in the later analysis, it is just as important to understand the marginal profit associated with increasing the loss cap as it is to understand the indicated profit load in isolation. So, the goal will be to differentiate \( P(L,p) \) by \( L \). But, differentiating the \( F_{Tc(L)}^{-1}(1-p) \) component of \( P(L,p) \) by \( L \) is complex. Thus, it is desirable to evaluate whether one may approximate the derivative of \( F_{Tc(L)}^{-1}(1-p) \) with some proxy. A view at the normal approximation equation (58), and a comparison to the lognormal approximation equation in (59) would suggest that the normal approximation has the advantage of relative simplicity (although of course this may come at a price paid in accuracy). Obviously, standard exercises in real analysis show that even when series of functions converge to a target function pointwise, their derivatives may not converge to the derivative of the target function. In fact, they may not even be differentiable at all. But, importantly, the characteristic function proof of the Central Limit Theorem shows that (under suitable assumptions) the Fourier transforms of larger and larger sample sizes converge to the Fourier transform of the normal distribution. Further, the Fourier transform “\( \hat{s}'_T \)” of the derivative of a function \( s'_T \) is simply the complex number “\( i \)” multiplied by the placeholder variable in transform space (often “\( \omega \)” times the Fourier transform of the original function \( s_T \). So, the Fourier transform proof of the Central Limit Theorem could also be extended to show that the derivatives of the Z-scores associated with a sum of multiple selections from a common distribution will approach the derivatives of the percentiles of the standard normal distribution. Therefore, the following analysis uses the derivative of \( \Phi^{-1}(1-p)Var^{1/2}[Tc(L)] \) (the normal approximation to the percentile minus the mean).

Then, one may begin with the approximation (58). Then, the approximation to \( P(L,p) \) is a linear function of \( Var^{1/2}[Tc(L)] \). So, one may readily see that for any normal approximation,

\[
\frac{\partial P(L,p)}{\partial L} = \frac{1}{2} \frac{\Phi^{-1}(1-p)r_FX_A + \nu(1-X_U)}{(1 + r_FX_A - X_U) \times Var^{1/2}[Tc(L)]} \times \frac{\partial}{\partial L} Var[Tc(L)]. \tag{66}
\]

That may also be restated to reflect the compound Poisson assumptions using (21) as

\[
\frac{\partial P(L,p)}{\partial L} \approx \frac{\Phi^{-1}(1-p)r_FX_A + \nu(1-X_U)}{(1 + r_FX_A - X_U) \times Var^{1/2}[Tc(L)]} n \lambda L (1 - F(L)). \tag{67}
\]
That in turn simplifies to

\[
\frac{\partial P(L,p)}{\partial L} \approx \Phi^{-1}(1-p)r_F X_A + \nu(1-X_U) \left( \frac{n\lambda(1-F_X(L))}{1 + \int_{x=0}^{L} s^2(x)dx} \right) \times \sqrt{\frac{L^2}{1 - F_X(L)}}.
\]

(68)

One may also use the alternate formulation

\[
\frac{\partial P(L,p)}{\partial L} \approx \Phi^{-1}(1-p)r_F X_A + \nu(1-X_U) \left( \frac{n\lambda(1-F_X(L))}{1 + \int_{x=0}^{L} s^2(x)dx} \right) \times \sqrt{\frac{L^2}{1 - F_X(L)}}.
\]

(69)

However, in terms of the impact on the company’s policyholders, it must be noted that the marginal effect on policyholders’ total premiums paid, with respect to \(L\), noting the linear multiplier in equation (46) of \(1/(1-V_{ER})\), is

Marginal profit load \(\approx \frac{\Phi^{-1}(1-p)r_F X_A + \nu(1-X_U)}{(1 + r_F X_A - X_U) \times \sqrt{E[X^2(L)]} \times (1 - V_{ER})} \times \sqrt{n\lambda L(1 - F(L))} \times (1 + r_F X_A - X_U) \times \sqrt{E[1 - F_X(L)]} \times (1 - V_{ER})\),

(70)

or in the general case

Marginal profit load \(\approx \frac{1}{2} \frac{\Phi^{-1}(1-p)r_F X_A + \nu(1-X_U)}{(1 + r_F X_A - X_U) \times \sqrt{Var[T(L)]} \times (1 - V_{ER})} \times \sqrt{n\lambda L(1 - F_X(L))} \times (1 + r_F X_A - X_U) \times \sqrt{Var[T(L)]} \times (1 - V_{ER})\),

(71)

Summing up the results of this section, it provides the means to identify the required profit associated with various levels of loss limit and certainty, and the marginal cost of increasing the loss limit. However, the quality of the estimates must still be evaluated.

3.6 Estimating the Optimal Retention and the Error in the Estimate

The last item remaining, the optimal cap on individual losses (the so-called “specific excess retention”) effected when insurers pass along more severe
risks to specialized “reinsurers” through the mechanism of “reinsurance”\textsuperscript{43} has received less formal mathematical attention within the broad community of North American casualty actuaries than its importance to insurance company operations would suggest. In effect, the insurer has a choice to make between “retaining” losses beyond a given size (the “retention”), which involves a relatively high share of risk, or “reinsuring” those upper levels of loss with a reinsurer, which involves paying for not just the expected costs of claims above the retention, but also the reinsurer’s expenses and profit. Further, within broad limits an insurer is free to select its individual retention. So, there is a choice between keeping some expected loss costs with a very high variance or paying a premium beyond the expected loss costs to eliminate the variance\textsuperscript{44}. The optimal retention “R”\textsuperscript{45} for specific excess reinsurance (mathematically, transfer of the claim costs \(\sum_{\{\text{claims} \leq c\}}(L_c - R)^+\), where \(L_c\) is the cost of the \(c^{th}\) claim) forms the subject matter of concern.

As discussed in the introduction, there has been some discussion in the literature of the optimal retention or attachment point at which it is best to begin ceding loss costs to a reinsurer. Some authors (Centeno [13], Cai and Tan [10], and Bernard and Tian [3], for example) have compared the benefits of purchasing different types of reinsurance in different layers. In practice, however, for most casualty lines, insurers simply select a single retention and purchase reinsurance that limits the most they pay on a single claim, without picking some layers to insure and others to retain\textsuperscript{46}. Therefore, guidance on selecting an optimal retention would be helpful to many insurers. Further, in some contrast to previous work, this paper will approach the problem from the standpoint of making the final insurance premium paid by the customer, reflecting both the profit needed to provide a return on capital and the reinsurer’s expenses, as low as possible. Of course that will

\textsuperscript{43}Reinsurance is so named because a risk that is already the subject of insurance passed to another “reinsurer”.

\textsuperscript{44}Of note, this is similar to what households (at least those that are understandably risk averse) do when they replace a fairly volatile set of expected losses in the form of potential house fires, automobile accidents, etc. with the purchase of insurance policies with fairly significant (15-35\% of the premium) expense loads. Financial economists have long discussed the rational basis for this. A discussion of the general principle of maximizing utility by paying a surcharge beyond expected costs can be found in [47].

\textsuperscript{45}“R” is used here while “L” is used elsewhere to avoid the confusing notation \((L_c - L)^+\).

\textsuperscript{46}An exception to this is that unlimited reinsurance is rarely available, so there is usually a claim size cap on the reinsurance protection, after which costs begin to revert to the insurance company. In well-designed insurance programs, these caps are usually far beyond the retention and generally are beyond the largest claim an insurer experiences in a given year.
still be within the context of meeting the solvency criteria of an acceptably low probability of failure. Also, since the most common approach used currently is to set the reinsurance retention at, say, a certain percentage of the insurance company’s surplus (see Gilmore [41]), it would be ideal to form a simplified basis for determining the retention to be used.

Using equation (71) and equation (47), the first step is to compute the optimal reinsurance attachment point, and the optimal layers to reinsure, when an insurance company purchases “specific excess of loss” reinsurance. This common form of reinsurance covers only the losses excess of some specified retention “\(L\)”, possibly up to some limit of reinsurance, and is the most common cost-effective form of reinsurance. In addition to the articles mentioned in Chapter 1, it is worth noting articles dealing with reinsurance purchasing such as Kaluszka [46], Gajek [39], and Gajek and Zagrodny [40].

As noted in the introduction, the goal is to set the reinsurance retention so that the total cost to the policyholders of the insurance company, including the expected retained losses, reinsurance premium, expenses, and required profit, is minimized. The process begins with equation (47), which states

\[
\text{Premium Charge} = \frac{E[T] + P(L,p) + Q(L,\infty) + FE}{1 - VER}. \tag{72}
\]

One may differentiate that by the retained limit \(L\) to get

\[
\frac{\partial}{\partial L} \left( \frac{\partial \text{Premium Charge}}{\partial L} \right) = \frac{\frac{\partial P(L,p)}{\partial L} + \frac{\partial Q(L,\infty)}{\partial L}}{1 - VER}. \tag{73}
\]

Clearly, as long as \(\frac{\partial}{\partial L} \left( \frac{\partial \text{Premium Charge}}{\partial L} \right) < 0\), the marginal (with respect to \(L\)) costs are lower, and hence the total cost will be incrementally lower, as the retention is increased. Hence, it does not make sense to reinsure the layers (potential loss values covered by the reinsurance) for which \(\frac{\partial}{\partial L} \left( \frac{\partial \text{Premium Charge}}{\partial L} \right) < 0\). However, should the derivative pass through zero and turn negative, so that \(\frac{\partial}{\partial L} \left( \frac{\partial \text{Premium Charge}}{\partial L} \right) > 0\), reinsuring the loss is preferable to retaining it.

Consider that while \(\frac{\partial P(L,p)}{\partial L}\) is possibly complex, \(\frac{\partial Q(L,\infty)}{\partial L}\) could simply be the expected losses excess of \(L\), or \(E[T] - E[T_c(L)],\) multiplied by a flat expense and profit loading\(^{47}\) times the loss content in the layer \(E[T] - E[T_c(L)]\). In conclusion, one may determine that:

\(^{47}\)This observation is based on the pricing structures most often seen for reinsurers and the industry practice of applying flat loadings to broad layers such as “The dollars of loss excess of $500,000 per claim up to a limit of $1,000,000 per claim excess of $500,000 ($1,500,000 in what is termed “ground-up” basis).
1. For a “layer” \((L_1, L_2]\) a reinsurer offers to reinsure containing \(L\) for which \(\frac{\partial P(L, p)}{\partial L} < -\frac{\partial Q(L, \infty)}{\partial L}\), the layer\(^{48}\) should be retained. Therefore, as long as the marginal profit on the layer is less than the reinsurer’s per unit loading for expenses and profit, the layer should be retained.

2. For \(L\) for which \(\frac{\partial P(L, p)}{\partial L} > -\frac{\partial Q(L, \infty)}{\partial L}\), the layer should be reinsured. Thus, as long as the marginal profit on the layer is more than the reinsurer’s per unit loading for expenses and profit, the layer should be reinsured.

3. Lastly, given reasonable continuity, at the switching point between retaining loss and reinsuring loss (otherwise known as the “retention” or “attachment point”) \(\frac{\partial P(L, p)}{\partial L} = -\frac{\partial Q(L, \infty)}{\partial L}\).

That set of results creates a unique vantage point for setting the retention. Contrary to the approach in the prior article listed earlier, this approach does not involve analyzing each possible layer or band (losses excess of some limit \(L\), up to to limit \(M\), as reinsurance is generally sold) of reinsurance. This offers two practical benefits. First, by focusing on a single point rather than on layers, the single point has potential to be within the values where the loss distribution is understood. So, it has potential to not involve replacing claims data with extensive assumptions about the general level of claims near the point, or above it. Hence, the single point approach tends to rely less on assumptions. Second, since the values and layers above the optimal retention need not be analyzed at all, it avoids the opportunistic approach to purchasing reinsurance. Specifically, it avoids the trap of picking and choosing whether to purchase each individual layer based on heavy use of assumptions (which often underestimate the costs in higher layers); not purchasing reinsurance in the higher layers; and then lacking reinsurance (and suffering heavy costs or heavy risk) when the costs in the upper layers far exceed the assumptions.

As a result, one would conclude that the optimal retention occurs where

\[
\frac{\partial P(L, p)}{\partial L} = -\frac{\partial Q(L, \infty)}{\partial L},
\]

with the optimal choice of retention logically at the closest retention offered by the reinsurer. Since reinsurers usually offer only broadly even retentions such as $100,000, $250,000, $500,000, $1,000,000, etc., the optimal choice

\(^{48}\)Of course the partial derivative exists on a point by point basis, not across a layer for \(L\). But one may think of whether an inequality holds throughout a layer, or throughout most of a layer.
of retention would be one of the offered retentions that are closest to the optimal retention. Absent a particularly second derivative, generally the offered retention that is closest to the optimal retention would be the best choice of retention in the context of this paper.

Alternately, if the marginal expense and profit load is\
\[ l(L) \frac{E[T_c(L)]}{\partial L} \] (a reasonable assumption, reflecting the common approach of loading a percentage of the expected loss in the layer for expenses and profit). Given that \( l(L) \) is constant in a layer (as discussed earlier, this is a common pricing approach for reinsurers), it is then also true that \( Q(L, M) = l \times (E[T_c(M)] - E[T_c(L)]) \) in the layer from \( L \) to \( M \). It then follows from (74) that
\[
\frac{\partial P(L, p)}{\partial L} = -l(L) \frac{\partial (E[T] - E[T_c(L)])}{\partial L} = l(L) \frac{\partial E[T_c(L)]}{\partial L} \tag{75}
\]
Continuing the assumption that the value of \( l(L) \) is identical, or essentially identical, across an entire layer, one might say that the optimal retention point is where
\[
\frac{\partial P(L, p)}{\partial L} = l \frac{\partial E[T_c(L)]}{\partial L} \tag{76}
\]
with that formula being a guide to the nearest possible retention. It then follows that, using formula (57), the optimal retention would occur at the value of \( L \) where
\[
\frac{\partial}{\partial L} \left( \Phi^{-1}((1-p)r_F X_A + \nu(1-X_U)) \times Var^{1/2}[T_c(L)] + r_F U \right) \quad \frac{1 + r_F X_A - X_U}{l} \cdot \frac{\partial E[T_c(L)]}{\partial L} \tag{77}
\]
That equation merits review. It consists of an extensive set of scalars and three functions, \( \Phi^{-1}(1-p) \), \( E[T_c(L)] \), and \( Var^{1/2}[T_c(L)] \). Estimating or even precisely calculating the values and derivatives of the last two involves moderate difficulty. Computing the derivative of \( \Phi^{-1}(1-p) \) is, however, mathematically forbidding. So, recognizing that the quality of this approximation must as yet be determined, the next step is to convert the equation to an approximating equation
\[
\frac{\partial}{\partial L} \left( \Phi^{-1}(1-p)r_F X_A + \nu(1-X_U) \right) \times Var^{1/2}[T_c(L)] + r_F U \quad \frac{1 + r_F X_A - X_U}{l} \cdot \frac{\partial E[T_c(L)]}{\partial L} \tag{78}
\]
53
which represents a general approximating formula for use in a wide variety of situations.

Further, one may also develop a general estimating formula for the error by noting that equation (77) may be expressed as

$$\frac{\partial}{\partial L} A(F_{T_c(L)}^{-1}(1-p) - E[T_c(L)]) + B\text{Var}^{1/2}[T_c(L)] + C = D\frac{\partial E[T_c(L)]}{\partial L}, \quad (79)$$

where the convenience constants are

$$A = \frac{r_F X_A}{1 + r_F X_A - X_U}, \quad B = \frac{\nu(1 - X_U)}{1 + r_F X_A - X_U}, \quad C = \frac{r_F U}{1 + r_F X_A - X_U}, \quad \text{and} \quad D = l. \quad (80)$$

Further, the alternate normal approximating equation has a slightly different root at \(L + \Delta L\), and

$$\frac{\partial}{\partial L} (A\Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L + \Delta L)] + B\text{Var}^{1/2}[T_c(L + \Delta L)] + C) = D\frac{\partial E[T_c(L + \Delta L)]}{\partial L}. \quad (81)$$

Then one need only subtract equation (79) from equation (81), and divide by \(\Delta L\). Subtracting left hand sides gives

$$\frac{\partial}{\partial L} \left( A\Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L + \Delta L)] - F_{T_c(L)}^{-1}(1-p) + E[T_c(L)] \right)$$

$$\approx A\Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L + \Delta L)] - F_{T_c(L)}^{-1}(1-p) - E[T_c(L)]$$

$$\approx A\Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L + \Delta L)] - F_{T_c(L)}^{-1}(1-p) - E[T_c(L)] \quad (82)$$

By partially expanding \(\text{Var}^{1/2}[T_c(L + \Delta L)]\) in a Taylor series with first order terms, that simplifies to (pending proof of the \(O\left(\frac{(\Delta L)^2}{L}\right)\) nature of the error term)
\[ \frac{A}{\Delta L} \left( \Phi^{-1}(1-p) \frac{\partial}{\partial L} \text{Var}^{1/2}[T_c(L)] + \Delta L \Phi^{-1}(1-p) \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)] \right) \\
+ \mathcal{O} \left( \frac{(\Delta L)^2}{L} \right) \\
- \frac{\partial}{\partial L} \left( F_{T_c(L)}^{-1}(1-p) - E[T_c(L)] \right) + B \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)] \\
= \frac{A}{\Delta L} \left( \Phi^{-1}(1-p) \frac{\partial}{\partial L} \text{Var}^{1/2}[T_c(L)] - \frac{\partial}{\partial L} \left( F_{T_c(L)}^{-1}(1-p) - E[T_c(L)] \right) \right) \\
+ A \Phi^{-1}(1-p) \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)] + \mathcal{O} \left( \frac{(\Delta L)^2}{L} \right) \\
+ B \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)] \\
\approx \frac{A}{\Delta L} \left( \Phi^{-1}(1-p) \frac{\partial}{\partial L} \text{Var}^{1/2}[T_c(L)] - \frac{\partial}{\partial L} \left( F_{T_c(L)}^{-1}(1-p) - E[T_c(L)] \right) \right) \\
+ (A \Phi^{-1}(1-p) + B) \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)]. \] (83)

Before going further, it is necessary to show the \( \mathcal{O} \left( \frac{\Delta L^2}{L} \right) \) nature of the first order error term. One may show that the term is (for some \( L^* \) near \( L \) and constant \( K )

\[ K \frac{\Delta L^2}{2} \frac{\partial^3}{\partial L^3} \text{Var}[T_c(L^*)] \]
\[ \approx K \frac{\Delta L^2}{2L} \frac{\partial^3}{\partial L^3} \text{Var}[T_c(L)] \]
\[ = K \frac{\Delta L^2}{2L} L \sqrt{n\lambda} \left\{ - \frac{2s(L) + Ls'(L)}{E^{1/2}[(X_c(L))^2]} \right. \\
- \frac{L(1 - F_X(L))(1 - F_X(L) - Ls(L))}{E^{3/2}[(X_c(L))^2]} \left. + 3 \frac{L^3(1 - F_X(L))^3}{E^{5/2}[(X_c(L))^2]} \right\}. \] (84)
or,

\[ K \frac{\Delta L^2}{2} \frac{\partial^3}{\partial L^3} \text{Var}[T_c(L^*)] = K \frac{\Delta L^2}{2L} \sqrt{\frac{nX}{2}} \left\{ \frac{-2Ls(L) + L^2s'(L)}{E^{1/2}[(X_c(L))^2]} \right. \\
- \frac{2L^2(1 - F_X(L))(1 - F_X(L) - Ls(L))}{E^{3/2}[(X_c(L))^2]} \\
+ \frac{3L^4(1 - F_X(L))^3}{E^{5/2}[(X_c(L))^2]} \right\}. \]

(85)

Due to the fact that the variance exists all the terms in the numerators converge to zero as \( L \to \infty \). So the term is \( O\left( \frac{\Delta L^2}{L} \right) \). Of course, it is implicit that the second derivative is smooth enough that the value of the third derivative of the variance at \( L \) is close its value at \( L^* \). Further some review will show that a suitable requirement to ensure that the third derivative converge to zero faster or at the same rate as the second derivative is a requirement that \( \lim_{L \to \infty} \frac{|Ls''''(L)|}{|s''(L)|} < \infty \). Such a quality will also ensure limited error resulting from the use of \( \frac{\partial^2 E[T_c(L)]}{\partial L^2} \) below\(^{49}\). Note that all Pareto distributions have that property, although some less skewed distributions such as the normal and lognormal do not.

Continuing the main thrust of the analysis, one may subtract right hand sides of (79) and 81) to get

\[ \frac{D \partial E[T_c(L + \Delta L)]}{\partial L} - \frac{D \partial E[T_c(L)]}{\partial L} \approx D \frac{\partial^2 E[T_c(L)]}{\partial L^2}. \]

(86)

Then, combining the left and right hand sides gives

\[ \frac{A\Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L + \Delta L)]}{\partial L} - \frac{A\Phi^{-1}(1-p) - E[T_c(L)]}{\partial L} \]

\[ + \left( A\Phi^{-1}(1-p) + B \right) \frac{\partial^2 E[T_c(L)]}{\partial L^2} \text{Var}^{1/2}[T_c(L)] \]

\[ \approx D \frac{\partial^2 E[T_c(L)]}{\partial L^2}, \]

(87)

\[ (88) \]

\(^{49}\)Note that in the case of the severity parameter variance, if the value of \( a \) is of sufficient size, the quality \( \lim_{L \to \infty} \frac{|s''''(L)|}{|s''(L)|} < \infty \) is also appropriate. As with \( \lim_{L \to \infty} \frac{|s''(L)|}{|s''(L)|} < \infty \), all Pareto distributions have this property.
or

\[ \Delta L \approx A \frac{\partial \Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L+\Delta L)]}{\partial L} - \frac{\partial (F_{T_c(L)}^{-1}(1-p)-E[T_c(L)])}{\partial L} D \frac{\partial^2 E[T_c(L)]}{\partial L^2} - (A\Phi^{-1}(1-p) + B) \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)]. \] (89)

Once the actual values for \( A, B, \) and \( D \) are reflected, the result is

\[ \Delta L \approx C \left( \frac{\partial \Phi^{-1}(1-p)\text{Var}^{1/2}[T_c(L)]}{\partial L} - \frac{\partial (F_{T_c(L)}^{-1}(1-p)-E[T_c(L)])}{\partial L} \right), \] (90)

where

\[ C = \frac{r_F X_A}{l(1+r_F X_A - X_U) \frac{\partial^2 E[T_c(L)]}{\partial L^2} - [r_F X_A \Phi^{-1}(1-p)+\nu(1-X_U)] \frac{\partial^2}{\partial L^2} \text{Var}^{1/2}[T_c(L)]}. \] (91)

Note that this expression, which is independent of the confidence level \((1-p)\), forms an approximation of the relative error in the retention estimate to the error in estimating the derivative. So, by calculating the value of \( C \) for a given model, limit and expected claims count, one may estimate the accuracy of the estimate of the optimal retention.

### 3.7 The Limits to Reinsurance

Some additional insight into the structure of an optimal reinsurance portfolio may be helpful. Recall that equations (67) and (68) imply

\[ \frac{\partial P(L, p)}{\partial L} \approx \frac{[\Phi^{-1}(1-p) r_F X_A + \nu(1-X_U)]}{(1+r_F X_A - X_U) \times \text{Var}^{1/2}[T_c(L)]} \times n\lambda L \int_{L}^{\infty} s(x) dx. \] (92)

That particular expression may be used to evaluate the “retention” or “attachment point” above which (at least locally) it is more cost effective to reinsure losses than to retain them for the company’s account. Further, if the reinsurance is locally a loading on expected losses, then

\[ \frac{\partial Q(L, \infty)}{\partial L} \approx l \frac{\partial}{\partial L} \left( E[T] - \int_{0}^{L} x f(x) dx - \int_{L}^{\infty} f(x) dx \right), \] (93)
where “\(T\)” is the reinsurer’s expense and profit loading on expected losses. Equivalently, one could say

\[
\frac{\partial Q(L, \infty)}{\partial L} \approx -l [Lf(L) + \int_{L}^{\infty} f(x) dx - Lf(L)] = -l \int_{L}^{\infty} f(x) dx.
\] (94)

So, at least in local ranges, the marginal value is the loading factor times the expected number of losses excess of \(L\). Further, this basic linearity of expected values means that this characterization of the derivative of \(Q\) may also be stated in a more arbitrary fashion as

\[
\frac{\partial Q(L, \infty)}{\partial L} \approx -l \frac{\partial}{\partial L} E_{T^*} \left[ \sum_{i=1}^{N} (X_i - L)^+ \right],
\]

where \(E_{T^*}\) denotes the expectation across all scenarios of individual losses making up the aggregate loss outcomes.

Using this information it is possible to make some global comments with respect to which losses should be reinsured and which should not. It seems intuitively clear that at low values of \(L\), purchasing reinsurance is typically not the preferable option. For example, at very low values of \(L\), especially those below unity, the number of expected claims excess of \(L\) is far greater than the expected costs of the entire loss costs \(E[T_c(L)]\). It is generally recognized that, at lower levels the relative variance across ranges of \(L\) has a lower ratio of the standard deviation to the mean, than at higher ranges. For example, at \(L = 1\), the mean is the expected number of claims \(n\lambda\). But, since almost all the claims in any group of insured claims will be of at least \$1 in size, the standard deviation will approximately equal the standard deviation of the Poisson distribution, or \(\sqrt{n\lambda}\). Therefore, at small sizes the ratio of the standard deviation to the mean will be very low at \(\frac{1}{\sqrt{n\lambda}}\). This means that at low loss sizes the surplus (and hence the profit required to support the risk) will be very low, favoring retaining the risk rather than ceding it. Further, in many cases, it will not be optimal to reinsure the lower layers. Considering the fact that reinsurance is used extensively by insurers, business realities show that, at least for intermediate values of \(n\lambda\), there are values of \(L\) at which reinsurance is more cost-effective than retaining risk for many insurance companies. Hence, for many insurers, there will be some point for which it is optimal to begin purchase reinsurance and for which it is preferable to retain the risk for lower values.
However, there is also a point \( L^* \) at which it will be optimal to stop purchasing reinsurance. This point involves understanding not just the pure standard deviation approach, but also the values corrected by \( G \) as noted in (29). Note that the \( 100(1 - p)^{th} \) percentile is larger\(^{50} \) than \( n\lambda \) at the cap \( L = 1 \). For relatively small retentions, where the expected cost of individual claims is nonetheless at least unity, continuity would indicate that the \( 100(1 - p)^{th} \) percentile is still larger than the retention (for small \( p \)). For reasonably large values of \( n\lambda \), though, and what may be described as modest retentions, that would imply that the \( 100(1 - p)^{th} \) percentile, for \( p \) reasonably small, would be larger than the mean, and hence larger than the loss limit \( L \). However, the \( 100(1 - p)^{th} \) percentile of the distribution is capped at the \( 100(1 - p)^{th} \) percentile of the unlimited distribution \( T \). But, may \( L \) increase to infinity. So at some point \( L^* \) (perhaps the highest of a number of points) \( L \) will equal the \( 100(1 - p)^{th} \) percentile of the uncapped distribution. At that point, further reinsurance is clearly not needed to meet the goal of limiting the size of the \( 100(1 - p)^{th} \) percentile, since all losses beyond that point occur with a lower probability than the confidence level itself. Consequently, at some point at or possibly below the crossing point (actually, \( L^* = F_T^{-1}(1 - p) \)) the effectiveness of reinsurance will diminish to the point where it is no longer optimal.

The cap \( L^* \) produces another implication for the \( G(L, p) \) defined in (29). As defined, it represents the ratio of the \( 100(1 - p)^{th} \) percentile of the actual distribution of \( T_c L \) to that estimated using a normal distribution with mean and variance equal to that of \( T_c L \). Further, it was shown that the \( 100(1 - p)^{th} \) percentile of \( T_c L \) stops increasing at some \( L^* \). But the variance of \( T_c L \), and its mean, will continue to grow, perhaps modestly, as \( L \) grows beyond \( L^* \). That implies that, given reasonable continuity, \( G(L, p) \) will initially grow. But \( G(L, p) \) will then reverse course and shrink at some higher point. So, \( G' \) will be usually be positive for at least some of the smaller points \( L \), but will be negative for very large \( L \).

In closing, a point of practical advice is offered about this analysis of this section. The analysis above is contingent on having a deep understanding of the distribution of potential aggregate claim costs and the distribution of expected claim costs among various layers. Often, the number of large claims present in an insurance company’s claims data is not large enough to reliably determine the upper reaches of the claim size distribution. So, in seeking to determine the cost/benefits trade-off, it is generally desirable to

\[^{50}\text{In this instance, no specific assumptions about the type of claim distribution is required. } n\lambda \text{ is simply the expected number of claims.}\]
undertake an extensive dialog with the reinsurer, understanding that its staff may be primarily interested in justifying a higher price, in order to make the best decision possible. As a result of such a dialog and by independently evaluating any data or other support provided by the reinsurer, the insurance company’s actuary has an opportunity to enhance his or her understanding of the true distribution of claim costs and consequently to make an optimal decision.
4 Application of the Financial Structure Results to the Compound Poisson Model

Now that the basic financial goals to be met by an insurance company have been specified, we return to the compound Poisson model and apply them in practice.

4.1 The Normal Approximation to the Optimal Reinsurance Retention

Of note, one may derive a specific formula for the optimal retention when a compound Poisson distribution is specified and the normal approximation to the derivative is used. First, note that, per (94),

\[
\frac{\partial Q(L, \infty)}{\partial L} \approx -l \int_{L}^{\infty} f(x)dx = -n\lambda l(1 - F_X(L)). \tag{96}
\]

Also, by (92),

\[
\frac{\partial P(L, p)}{\partial L} = \Phi^{-1}(1 - p) r_F X_A + \nu(1 - X_U) \frac{n\lambda l(1 - F_X(L))}{(1 + r_F X_A - X_U) \text{Var}^{1/2}[T_c(L)]}. \tag{97}
\]

Since at the optimal retention \(L = R\), where \(\frac{\partial P(L, p)}{\partial L}(L = R) = -\frac{\partial Q(L, \infty)}{\partial L}(L = R)\), we can combine the two to get the following estimating formula for the optimal retention:

\[
n\lambda l(1 - F_X(R)) \approx \frac{\Phi^{-1}(1 - p) r_F X_A + \nu(1 - X_U)}{(1 + r_F X_A - X_U) \text{Var}^{1/2}[T_c(R)]} n\lambda R(1 - F_X(R)). \tag{98}
\]

That in turn simplifies to

\[
\frac{\text{Var}^{1/2}[T_c(R)]}{R} \approx \frac{\Phi^{-1}(1 - p) r_F X_A + \nu(1 - X_U)}{l(1 + r_F X_A - X_U)}. \tag{99}
\]

Using the formula for the variance of the capped compound Poisson distribution in (13), that in turn reduces to

\[
\frac{E[X^2_c(R)]}{R^2} \approx \frac{\Phi^{-1}(1 - p) r_F X_A + \nu(1 - X_U)^2}{n\lambda^2 l(1 + r_F X_A - X_U)^2}. \tag{100}
\]
Although that formula is an approximation, not an equality, it may be used as an equality in order to approximate the optimal retention. Of note, calculating the roots of that approximation would likely require a mathematical formula for the loss severity, and even then it might have to be computed numerically. Nevertheless, any actual density function is likely to be very discontinuous as values lump around key values such as $10,000 and so forth. Further, as noted in Section 2.4 near equation (29), it is common for actuaries to replace historical loss data with fitted curves in performing these analyses. So, it is reasonable to use formulas whose solutions require algebraic or numerical analysis.

Recall that the actual, authentic, values of the needed surplus and profit are given by

\begin{align}
S(L,p) = \frac{\{F_{T_c(L)}^{-1}(1-p) - E[T_c(L)] + \nu Var^{1/2}[T_c(L)]\} (1 - X_U) - r_F U}{1 + r_F X_A - X_U}
\end{align}

(101)

and

\begin{align}
P(L,p) = \frac{\{F_{T_c(L)}^{-1}(1-p) - E[T_c(L)]\} r_F X_A + \nu (1 - X_U) Var^{1/2}[T_c(L)] + r_F U}{1 + r_F X_A - X_U}.
\end{align}

(102)

So, using equation (17) for the \(L\)-derivative of the variance, the true optimal retention occurs when

\begin{align}
-\frac{\partial Q(L, \infty)}{\partial L} &= -\frac{\partial P(L,p)}{\partial L} \\
&= \frac{\partial}{\partial L} \{F_{T_c(L)}^{-1}(1-p) - E[T_c(L)]\} r_F X_A \\
&\quad \times \nu (1 - X_U) \\
&\quad \times \frac{1}{1 + r_F X_A - X_U} \\
&\quad \times \frac{1}{1 + r_F X_A - X_U} \times \int_{L}^{\infty} s(x) \, dx.
\end{align}

(103)
That would infer that the formula (100), derived from the normal approximation to the $100(1 - p)^{th}$ percentile, would be a good approximation (as it is based on (97) and (98)) only as long the derivative of the normal approximation to the percentile is also a good approximation to the derivative of the percentile. Specifically, it is key that

$$\frac{\partial}{\partial L} \left\{ E[T_c(L)] + \Phi^{-1}(1 - p)\text{Var}^{1/2}[T_c(L)] \right\} \approx \frac{\partial F^{-1}_{T_c(L)}(1 - p)}{\partial L},$$

(104)
or, equivalently,

$$\frac{\partial \Phi^{-1}(1 - p)\text{Var}^{1/2}[T_c(L)]}{\partial L} \approx \frac{\partial \left\{ F^{-1}_{T_c(L)}(1 - p) - E[T_c(L)] \right\}}{\partial L},$$

(105)
and that the $C$ value (which relates approximation error relative to $F^{-1}_{T_c(L)}(1 - p) - E[T_c(L)]$ to error in the retention approximation), per equation (91), is relatively modest in size.

### 4.2 The Quality of the Derivative and Retention Estimates

In the previous section, the derivative of the normal approximation to the inverse cumulative distribution was used to generate an approximation of the optimal retention. In this section, an analysis of the quality of that approximation is included.

As a first step in evaluating whether or not that approximation was valid, the relation between the derivatives of $F^{-1}_{T_c(L)}(1 - p) - E[T_c(L)]$ at different prospective retentions are far greater than the differences between the derivatives of $F^{-1}_{T_c(L)}(1 - p) - E[T_c(L)]$ and $\Phi^{-1}(1 - p)\sqrt{\text{Var}[T_c(L)]}$ at a single combination of $p$ and $L$. Therefore, the value of $\Phi^{-1}(1 - p)\sqrt{\text{Var}[T_c(L)]}$ at the $250,000$ retention would not mistakenly be thought to underlie a $100,000$ or $1,000,000$ retention (except for very high retentions in comparison to claim counts). Of note, it may also be helpful to display the percentage differences between the two derivative formulas (as a percentage of the most relevant derivative, that of $F^{-1}_{T_c(L)}(1 - p) - \mu_{T_c(L)}$). The results are shown in Table 7.
Table 6: Comparison of Derivative of Normal Approximation to Derivative of $F^{-1}_{T_c(L)}(1 - p) - \mu_{T_c(L)}$ under Compound Poisson Assumptions and Pareto(2.2, 5000) Loss Severity Distribution

<table>
<thead>
<tr>
<th>n\lambda</th>
<th>L</th>
<th>$\Phi^{-1}\sigma'$</th>
<th>$D[F^{-1} - \mu]$</th>
<th>$\Phi^{-1}\sigma'$</th>
<th>$D[F^{-1} - \mu]$</th>
<th>$\Phi^{-1}\sigma'$</th>
<th>$D[F^{-1} - \mu]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$25,000$</td>
<td>0.9534</td>
<td>1.0703</td>
<td>1.1905</td>
<td>1.4131</td>
<td>1.3485</td>
<td>1.5871</td>
</tr>
<tr>
<td>60</td>
<td>$50,000$</td>
<td>0.3712</td>
<td>0.3879</td>
<td>0.4635</td>
<td>0.6253</td>
<td>0.5250</td>
<td>0.6163</td>
</tr>
<tr>
<td>60</td>
<td>$100,000$</td>
<td>0.1491</td>
<td>0.2230</td>
<td>0.1862</td>
<td>0.3200</td>
<td>0.2109</td>
<td>0.3280</td>
</tr>
<tr>
<td>60</td>
<td>$250,000$</td>
<td>0.0460</td>
<td>0.0575</td>
<td>0.0574</td>
<td>0.1306</td>
<td>0.0650</td>
<td>0.2278</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000$</td>
<td>0.0080</td>
<td>-0.0005</td>
<td>0.0100</td>
<td>-0.0005</td>
<td>0.0114</td>
<td>-0.0005</td>
</tr>
<tr>
<td>150</td>
<td>$25,000$</td>
<td>1.5075</td>
<td>1.6253</td>
<td>1.8823</td>
<td>1.9504</td>
<td>2.1321</td>
<td>2.3507</td>
</tr>
<tr>
<td>150</td>
<td>$50,000$</td>
<td>0.5869</td>
<td>0.6542</td>
<td>0.7328</td>
<td>0.7961</td>
<td>0.8301</td>
<td>1.0356</td>
</tr>
<tr>
<td>150</td>
<td>$100,000$</td>
<td>0.2358</td>
<td>0.3510</td>
<td>0.2944</td>
<td>0.5332</td>
<td>0.3334</td>
<td>0.3588</td>
</tr>
<tr>
<td>150</td>
<td>$250,000$</td>
<td>0.0727</td>
<td>0.1395</td>
<td>0.0907</td>
<td>0.1890</td>
<td>0.1028</td>
<td>0.2510</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000$</td>
<td>0.0127</td>
<td>-0.0013</td>
<td>0.0159</td>
<td>-0.0013</td>
<td>0.0180</td>
<td>-0.0013</td>
</tr>
<tr>
<td>400</td>
<td>$25,000$</td>
<td>2.4618</td>
<td>2.7325</td>
<td>3.0738</td>
<td>3.3841</td>
<td>3.4817</td>
<td>4.2927</td>
</tr>
<tr>
<td>400</td>
<td>$50,000$</td>
<td>0.9584</td>
<td>1.0295</td>
<td>1.1967</td>
<td>1.5538</td>
<td>1.3555</td>
<td>1.8212</td>
</tr>
<tr>
<td>400</td>
<td>$100,000$</td>
<td>0.3850</td>
<td>0.4511</td>
<td>0.4807</td>
<td>0.5573</td>
<td>0.5445</td>
<td>0.6814</td>
</tr>
<tr>
<td>400</td>
<td>$250,000$</td>
<td>0.1187</td>
<td>0.1232</td>
<td>0.1482</td>
<td>0.3297</td>
<td>0.1679</td>
<td>0.3393</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000$</td>
<td>0.0208</td>
<td>-0.0025</td>
<td>0.0259</td>
<td>0.0399</td>
<td>0.0294</td>
<td>0.0563</td>
</tr>
</tbody>
</table>

One may see that the approximation errors appear at first blush to be significant. However one must consider two items. First, even if the percentage difference between the actual derivative and the approximation is small, the slope of the curve determining the optimal retention may be low, so that an acceptable estimate, recognizing that retentions offered by reinsurers are usually limited to very even amounts such as $25,000, $50,000, $100,000, $250,000, etc. may differ from the true value by as much as 25%. Second, consider, for example, the $1,000,000 retention when 60 claims are expected. Per the third column, roughly 0.1% of these claims is expected to happen in a sample of 60 claims. So, there is approximately a 0.1% chance of one happening in a given year. So, at the 99.5% confidence level, zero of these claims are expected. So, the derivative of $F^{-1}_{T_c(L)}(.995)$ at $L = 1,000,000$ is zero, leaving only a modest derivative of $\mu_{T_c(L)}$ term ($n\lambda(1-F_X($1,000,000$))$). Since the derivative of $\mu_{T_c(L)}$ at that point is still positive, but is subtracted from $F^{-1}_{T_c(L)}(.995)$ to determine the total needed surplus, the derivative of the needed surplus and profit by the limit “$L$” is
Table 7: Percentage Approximation Error in Normal Derivative Using Compound Poisson Model

<table>
<thead>
<tr>
<th>$n\lambda$</th>
<th>$L$</th>
<th>$E[#\text{excess claims}]$</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$25,000$</td>
<td>1.739</td>
<td>-10.916 %</td>
<td>-15.756 %</td>
<td>-15.037 %</td>
<td>-34.815 %</td>
</tr>
<tr>
<td>60</td>
<td>$50,000$</td>
<td>0.379</td>
<td>-4.302 %</td>
<td>-25.885 %</td>
<td>-14.821 %</td>
<td>-44.124 %</td>
</tr>
<tr>
<td>60</td>
<td>$100,000$</td>
<td>0.082</td>
<td>-33.141 %</td>
<td>-41.822 %</td>
<td>-35.698 %</td>
<td>-29.993 %</td>
</tr>
<tr>
<td>60</td>
<td>$250,000$</td>
<td>0.011</td>
<td>-20.037 %</td>
<td>-56.040 %</td>
<td>-71.456 %</td>
<td>-82.348 %</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>-1647.251 %</td>
<td>-2031.884 %</td>
<td>-2288.307 %</td>
<td>-2522.985 %</td>
</tr>
<tr>
<td>150</td>
<td>$25,000$</td>
<td>4.349</td>
<td>-7.245 %</td>
<td>-3.494 %</td>
<td>-15.749 %</td>
<td>-1.371 %</td>
</tr>
<tr>
<td>150</td>
<td>$50,000$</td>
<td>0.946</td>
<td>-10.285 %</td>
<td>-7.948 %</td>
<td>-19.847 %</td>
<td>-19.145 %</td>
</tr>
<tr>
<td>150</td>
<td>$100,000$</td>
<td>0.206</td>
<td>-32.830 %</td>
<td>-44.796 %</td>
<td>-7.075 %</td>
<td>-15.828 %</td>
</tr>
<tr>
<td>150</td>
<td>$250,000$</td>
<td>0.027</td>
<td>-47.918 %</td>
<td>-51.994 %</td>
<td>-50.045 %</td>
<td>-73.864 %</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>-1078.568 %</td>
<td>-1321.830 %</td>
<td>-1484.007 %</td>
<td>62.155 %</td>
</tr>
<tr>
<td>400</td>
<td>$25,000$</td>
<td>11.596</td>
<td>-9.906 %</td>
<td>-9.170 %</td>
<td>-18.891 %</td>
<td>-6.115 %</td>
</tr>
<tr>
<td>400</td>
<td>$50,000$</td>
<td>2.524</td>
<td>-6.909 %</td>
<td>-22.984 %</td>
<td>-25.573 %</td>
<td>4.942 %</td>
</tr>
<tr>
<td>400</td>
<td>$100,000$</td>
<td>0.549</td>
<td>-14.646 %</td>
<td>-13.742 %</td>
<td>-20.090 %</td>
<td>-24.791 %</td>
</tr>
<tr>
<td>400</td>
<td>$250,000$</td>
<td>0.073</td>
<td>-3.650 %</td>
<td>-55.050 %</td>
<td>-50.535 %</td>
<td>-56.694 %</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000$</td>
<td>0.003</td>
<td>-943.604 %</td>
<td>-35.058 %</td>
<td>-47.799 %</td>
<td>-89.481 %</td>
</tr>
</tbody>
</table>

actually negative at $L = $1,000,000. Necessarily, the derivative of the standard deviation approach, $\Phi^{-1}(1 - p)\sigma_{T_c(L)}$ by $L$ is positive, so it is impossible to equalize the two. However, one must consider that this only occurs for losses that are too unlikely to occur to be contained within $F^{-1}_{T_c(L)}(.995)$, so they are really beyond the confidence level that the profit and surplus is intended to provide.

To seek whether or not differences between the two derivatives are substantial enough to cause an insurance company reinsurance manager to make an incorrect decision about the optimal retention point, it suffices to evaluate $C$ as described in equation (91). Using the sample\(^{51}\) data and finite differences of $\pm 10\%$, the necessary second derivatives were computed. They were combined with the following very reasonable\(^{52}\) assumptions for the various

---

\(^{51}\)Due to the nature of the Pareto and compound Poisson distributions, it is actually reasonably tractable to compute the derivatives analytically. However, for later cases, the problem has potential to become more complex, so the uniform approach of computing the values numerically from the sample data was used. Also of note, in order to minimize the sampling error in computing derivatives, the values used to compute the derivatives were consistently either the same sample of the same combination of samples. While the values may have more relative than exact precision, exact precision is not needed for these calculations.

\(^{52}\)Specifically, they are reasonable considering the financial environment in 2012
constants:

\( r_F \) = risk free rate = 3%,

\( X_A \) = tax rate on asset investments = 15%,

\( X_U \) = tax rate on discounted underwriting profit = 35%,

\( l \) = reinsurer’s expense and profit markup = 15%, and

\( \nu \) = profit loading on diversifiable standard deviation = 1%.

That combination yields Tables 8 and 9.

Table 8: Estimated \( \mathcal{C} \) Ratios of Percentage Error in Retention Estimate to Error in Derivative Estimate — under Compound Poisson Scenario

<table>
<thead>
<tr>
<th>( n\lambda )</th>
<th>( L )</th>
<th>( n\lambda(1 - F_X(L)) )</th>
<th>( 1 - p = .95 )</th>
<th>( 1 - p = .98 )</th>
<th>( 1 - p = .99 )</th>
<th>( 1 - p = .995 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$ 25,000</td>
<td>1.739</td>
<td>-127</td>
<td>-127</td>
<td>-127</td>
<td>-127</td>
</tr>
<tr>
<td>60</td>
<td>$ 50,000</td>
<td>0.379</td>
<td>-1.227</td>
<td>-1.230</td>
<td>-1.233</td>
<td>-1.235</td>
</tr>
<tr>
<td>60</td>
<td>$ 100,000</td>
<td>0.082</td>
<td>-10.514</td>
<td>-10.580</td>
<td>-10.624</td>
<td>-10.665</td>
</tr>
<tr>
<td>60</td>
<td>$ 250,000</td>
<td>0.011</td>
<td>-216.416</td>
<td>-219.405</td>
<td>-221.444</td>
<td>-223.343</td>
</tr>
<tr>
<td>60</td>
<td>$ 1,000,000</td>
<td>0.001</td>
<td>-15,885,777</td>
<td>-17,773,771</td>
<td>-19,303,217</td>
<td>-20,953,363</td>
</tr>
<tr>
<td>150</td>
<td>$ 25,000</td>
<td>4.349</td>
<td>-50</td>
<td>-50</td>
<td>-51</td>
<td>-51</td>
</tr>
<tr>
<td>150</td>
<td>$ 50,000</td>
<td>0.946</td>
<td>-475</td>
<td>-476</td>
<td>-477</td>
<td>-477</td>
</tr>
<tr>
<td>150</td>
<td>$ 100,000</td>
<td>0.206</td>
<td>-4,184</td>
<td>-4,199</td>
<td>-4,209</td>
<td>-4,218</td>
</tr>
<tr>
<td>150</td>
<td>$ 250,000</td>
<td>0.027</td>
<td>-84,020</td>
<td>-84,727</td>
<td>-85,205</td>
<td>-85,647</td>
</tr>
<tr>
<td>150</td>
<td>$ 1,000,000</td>
<td>0.001</td>
<td>-5,953,577</td>
<td>-6,325,632</td>
<td>-6,600,628</td>
<td>-6,874,126</td>
</tr>
<tr>
<td>400</td>
<td>$ 25,000</td>
<td>11.596</td>
<td>-19</td>
<td>-19</td>
<td>-19</td>
<td>-19</td>
</tr>
<tr>
<td>400</td>
<td>$ 50,000</td>
<td>2.524</td>
<td>-178</td>
<td>-178</td>
<td>-178</td>
<td>-178</td>
</tr>
<tr>
<td>400</td>
<td>$ 100,000</td>
<td>0.549</td>
<td>-1,589</td>
<td>-1,592</td>
<td>-1,594</td>
<td>-1,596</td>
</tr>
<tr>
<td>400</td>
<td>$ 250,000</td>
<td>0.073</td>
<td>-29,162</td>
<td>-29,322</td>
<td>-29,429</td>
<td>-29,528</td>
</tr>
<tr>
<td>400</td>
<td>$ 1,000,000</td>
<td>0.003</td>
<td>-2,133,131</td>
<td>-2,193,710</td>
<td>-2,236,046</td>
<td>-2,276,249</td>
</tr>
</tbody>
</table>

In summary, the approximations are very good except at the larger loss caps and smaller expected claim counts/high confidence level scenarios. Reviewing the third column in the chart, these deal with situations that are beyond the point where reinsurance impacts the capital and profit needed to secure a given confidence level. Recall that the theory of upper limit to needed reinsurance, \( L^* \), as articulated in Section 3.7, predicts that behavior and suggests that such limits are beyond the optimal limit for reinsurance. So, in summary, for determining the retention itself\(^{53}\), the normal approx-

\(^{53}\)If the upper limit of reinsurance is desired, it may be more practical to use some
Table 9: Estimated Percentage Error in Retention Estimate Due to Use of Normal Approximation — under Compound Poisson Scenario

<table>
<thead>
<tr>
<th>nλ</th>
<th>$L$</th>
<th>nλ(1 − FX(L))</th>
<th>1 − p = .95</th>
<th>1 − p = .98</th>
<th>1 − p = .99</th>
<th>1 − p = .995</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$25,000$</td>
<td>1.739</td>
<td>0.06 %</td>
<td>0.11 %</td>
<td>0.12 %</td>
<td>0.41 %</td>
</tr>
<tr>
<td>60</td>
<td>$50,000$</td>
<td>0.379</td>
<td>0.004 %</td>
<td>0.04 %</td>
<td>0.23 %</td>
<td>1.13 %</td>
</tr>
<tr>
<td>60</td>
<td>$100,000$</td>
<td>0.082</td>
<td>0.078 %</td>
<td>1.42 %</td>
<td>1.24 %</td>
<td>1.07 %</td>
</tr>
<tr>
<td>60</td>
<td>$250,000$</td>
<td>0.011</td>
<td>1.00 %</td>
<td>6.42 %</td>
<td>14.42 %</td>
<td>30.00 %</td>
</tr>
<tr>
<td>60</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>-13.60 %</td>
<td>-18.77 %</td>
<td>-22.96 %</td>
<td>-27.48 %</td>
</tr>
<tr>
<td>150</td>
<td>$25,000$</td>
<td>4.349</td>
<td>0.02 %</td>
<td>0.01 %</td>
<td>0.08 %</td>
<td>0.01 %</td>
</tr>
<tr>
<td>150</td>
<td>$50,000$</td>
<td>0.946</td>
<td>0.06 %</td>
<td>0.06 %</td>
<td>0.20 %</td>
<td>0.21 %</td>
</tr>
<tr>
<td>150</td>
<td>$100,000$</td>
<td>0.206</td>
<td>0.48 %</td>
<td>1.00 %</td>
<td>0.11 %</td>
<td>0.29 %</td>
</tr>
<tr>
<td>150</td>
<td>$250,000$</td>
<td>0.027</td>
<td>2.25 %</td>
<td>3.33 %</td>
<td>5.05 %</td>
<td>11.02 %</td>
</tr>
<tr>
<td>150</td>
<td>$1,000,000$</td>
<td>0.001</td>
<td>-8.35 %</td>
<td>-10.87 %</td>
<td>-12.73 %</td>
<td>-5.25 %</td>
</tr>
<tr>
<td>400</td>
<td>$25,000$</td>
<td>11.596</td>
<td>0.02 %</td>
<td>0.02 %</td>
<td>0.06 %</td>
<td>0.02 %</td>
</tr>
<tr>
<td>400</td>
<td>$50,000$</td>
<td>2.524</td>
<td>0.03 %</td>
<td>0.13 %</td>
<td>0.17 %</td>
<td>0.03 %</td>
</tr>
<tr>
<td>400</td>
<td>$100,000$</td>
<td>0.549</td>
<td>0.10 %</td>
<td>0.12 %</td>
<td>0.22 %</td>
<td>0.32 %</td>
</tr>
<tr>
<td>400</td>
<td>$250,000$</td>
<td>0.073</td>
<td>0.05 %</td>
<td>2.13 %</td>
<td>2.02 %</td>
<td>2.87 %</td>
</tr>
<tr>
<td>400</td>
<td>$1,000,000$</td>
<td>0.003</td>
<td>-4.96 %</td>
<td>-3.07 %</td>
<td>6.01 %</td>
<td>62.97 %</td>
</tr>
</tbody>
</table>

The approximation to the optimal reinsurance retention appears to work very well in practical situations.
5 Summary

An insurance company faces a set of operational questions that relate to the size of the largest claim it retains. The company’s management needs to know the amount of surplus funds that they need to cover the prospect that losses and expenses may exceed premiums. They need to know how much profit they must achieve to provide their investors a fair return. Furthermore, they need to determine the reinsurance retention that is optimal for the company. A discussion was begun of the values and operational aspects that underlie those questions. Insurance companies’ capital needs were found to be linked to the difference between the amount needed to secure payment of all the claims within a specified confidence level, and the expected loss content that is built into their rates. Consequently, since the analysis relates solely to the losses “retained” by the company, this paper suggests that no profit markup is needed on the losses passed to a “reinsurer”. Further, this paper examined the long-running controversy as to whether insurance companies should be recompensed for the basic volatility of insurance losses in addition to being recompensed for their exposure to the systemic risk of the stock market as a whole. The data analyzed in this study suggest that some premium for non-systemic risk could be demanded by stockholders of insurance companies. However, the data also suggest that the markup should be a small portion of that non-systemic risk\textsuperscript{54}. The analysis indicates that, in the absence of correlation between the losses borne by the company and the stock market, the profit loaded into insurance rates should be the precise amount which, less the tax on underwriting profits, equals the tax on the risk free rate on surplus, plus the risk-free rate times the amount of capital deployed in non-investment income bearing assets, plus the small non-systemic risk loading on the standard deviation of the possible retained losses. Since first principles would indicate that there is typically no correlation between the volatility of insurance losses and the volatility of the market, the analysis makes a strong statement about the indicated profit need for an insurance company.

The resulting capital and profit needed to sustain operations, and optimal retentions across a comprehensive set of scenarios for the distribution of aggregate loss costs\textsuperscript{55} were analyzed within this paper. As part of the analysis, closed-form equations for the normal and lognormal approxima-

\textsuperscript{54}As measured by the square root of the portion of the variance of the aggregate losses that is independent of systemic market risk.

\textsuperscript{55}As regards the situation where claims are paid within the term of the policy giving rise to the claim.
tions to the upper percentiles of the aggregate distribution were developed. A specific distribution with high skew and kurtosis was chosen to test each scenario. Even though that distribution, by virtue of its skew, presented a challenge for the normal approximation, the quality of the approximations to the various values were nevertheless acceptable\textsuperscript{56}. The example indicates, though, that generally the lognormal approximation is closer to the true percentile value of $F$ in the range of caps or retentions that are likely to be used in practice. Importantly, the formulas provided are, to a very large extent, distribution free and subject to only a bare minimum of assumptions.

The analysis in this paper develops a general formula for the optimal retention (the point where individual large losses should be transferred to reinsurers) that uses basic cost, tax, and interest rate information, a special loading for non-systematic risk, and characteristics of the loss cost distribution. A specific formula for the optimal retention was developed. The formula was based on approximating the derivative of the $100(1-p)^{th}$ percentile with respect to the potential retentions with the derivative of the $100(1-p)^{th}$ percentile of the approximating normal distribution. Considering the possible error in that approximation, it is also important to recognize that its accuracy, and the impact of the approximation error, should be monitored by users of the formulas in this paper. To that end, a formula relating the error in approximating the derivative to the consequent error in the estimate of the retention was developed. When evaluated against the example distribution, the errors in the retentions for various limits and expected claim counts were found to be acceptable\textsuperscript{57}. The formulas developed throughout this paper were also, to a large extent, distribution free and subject to only a bare minimum of assumptions.

This paper also includes several results that are more qualitative in nature than purely quantitative. As noted earlier, an analysis was presented suggesting that that investors do demand a small premium for variance uncorrelated with the market. That contrasts significantly with the core theory of the Capital Asset Pricing Model as discussed by Sharpe [66] and others. It also suggests that some small load for the pure randomness of insurance claims is required to support the capital employed by an insurance company. Carrying that analysis further, the analysis showed that the profit load needed in the rates is independent of how an insurance company invests its assets. As such, the needed profit load in the rates is mostly

\textsuperscript{56}At least, acceptable in most relevant circumstances.

\textsuperscript{57}In light of insurance companies’ common buying practice (and reinsurers’ common selling practice) of purchasing reinsurance only at rounded figures such as $25,000, $100,000, $250,000, $500,000, and $1,000,000.
independent of the Capital Asset Pricing Model “beta” that relates the return required by investors for the correlation between the volatility in the company’s investment portfolio and the volatility of the stock market as a whole. As noted earlier, in the common case where the losses borne by an insurance company are independent of the return of the stock market, the profits generated by the insurance company need only recover the tax on the risk-free rate applied to invested surplus and the risk-free return on un-invested surplus. This paper also discusses the existence of a limit beyond which it is not optimal for a company to purchase reinsurance.

There is potential for further investigation of these ideas. They could conceivably be extended to situations where the average frequency and severity are uncertain; where the insurance company sells more than one line of business (and hence has more than one claims severity distribution); where the claims payout (and hence the capital commitment) takes several years; where there is some correlation between loss costs and stock market yields; and where guidance in assessing some of the key risk factors is desired by a company selling several different lines of insurance. Except for those issues, this paper presents a comprehensive analysis of funding percentiles, profit needs, retention selection, and the impact of various loss limits. It is hoped that professionals engaged in insurance ratemaking, reinsurance purchasing, and enterprise risk management will deploy them effectively, and to the public’s benefit.
A List of Special Symbols

To create a standard, easy-to-follow notation, the following symbols were introduced in this paper.

\( \nu \) — Indicated factor loading on the portion of the standard deviation that is uncorrelated with the market, used in determining the market-indicated excess return for volatility uncorrelated with the market as a whole.

\( A(L) \) — Coefficient of variation of \( T_c(L) \) squared, plus unity, \( (\text{Var}[T_c(L)]/E^2[T_c(L)]) + 1 \). Used as key value in computing lognormal estimate to percentiles \( F_{T_c(L)}^{-1}(1 - p) \). Also, a coefficient used in combining coefficients to simplify the analysis of the error in estimating the optimal retentions as a function of the error in estimating the derivative.

\( C \) — The estimated ratio of the error in the optimal retention to the error in estimating the derivative of the funding percentile with respect to the limit with the derivative of its normal approximation. Also, a matrix used in converting the \( L \) and \( M \) derivative errors due to use of the normal approximation to the errors in the optimal retention estimates.

\( F_{T_c(L)}(\cdot) \) — Cumulative distribution function of the aggregate output of a compound Poisson or similar random variable, with individual claims/jump sizes capped at \( L \), most commonly used in inverse form, the \( 100(1 - p)^{th} \) percentile of the aggregate output of the distribution, \( F_{T_c(L)}^{-1}(1 - p) \).

\( G(L,p) \) — Ratio of true percentile \( F_{T_c(L)}^{-1}(1 - p) \) to the normal approximation to it, \( E[T_c(L)] + \Phi^{-1}(1 - p)\text{Var}^{1/2}[T_c(L)] \). Also used for lognormal approximation.

\( l \) — Reinsurer’s expense and profit loading built into its price as a ratio to expected loss dollars transferred to the reinsurer for losses under the severity distribution of \( X \). Briefly, a placeholder for various loss caps between 0 and \( L \) in an integration of variance. Also, count of the number of different \( X \) distributions, or lines of business, in a scenarios where there are more than two lines of business.

\( L \) — A cap or limit imposed on individual samples (claims, or jump sizes) from the \( X \) distribution.

\( P, P(L,p) \) — The aggregate dollar profit earned or required to be earned by an insurance company on underwriting (specifically not passive investment of surplus), often the profit needed to sustain capital needed to reduce the probability of inability to pay all claims to \( p \) when each each individual claim is capped at \( L \) by reinsurance or policy limits.

\( Q(L, \infty) \) — Reinsurer’s aggregate profit load when assuming band between
$L$ and infinity.

$s(x)$ or $s_X(x)$ — loss severity (jump size, in the compound Poisson case) distribution for the $X$’s. Specially, the probability mass function of the size of an individual claim, $x$ representing possible claim sizes. In parameter variance situations, $s$ generally refers to the distribution of the baseline scenario where all parameter variance values are unity.

$s_T(t)$ — loss severity distribution for the aggregate losses $T$. Specially, the probability mass function of the total aggregate costs, $t$ representing possible aggregate costs.

$S(L, p)$ — $S$ or Surplus (“buffer fund”) of an insurance company, often the surplus needed to limit the probability of inability to pay all claims to $p$ when each each individual claim is capped at $L$ by reinsurance or policy limits.

$T$ — Random variable of total loss (sum of the unlimited loss on all claims combined) generated by a company over a time period.

$T_c(L)$ — Random variable of loss capped at a limit $L$ (sum of the loss, capped at $L$ per claim, on all claims combined) generated by a company over a time period.

$U$ — Dollar amount of assets owned by an insurance company, or the capital invested in those assets, which do not generate interest or dividend income and are not expected to generate capital gains.

$Var_c(L)$ — Variance of $T_c(L)$.

$Var_r(L)$ — Variance of $T_r(L)$.

$X_r(L)$ — Randomly sampled individual claim or jump size from first distribution restricted to $L$. Eg. $X_r(L) = X$ when $X \leq L$, and $X_r(L) = 0$ otherwise.

$X_c(L)$ — Randomly sampled individual claim or jump size from first distribution capped at $L$. Eg. $X_c(L) = X$ when $X \leq L$, and $X_c(L) = L$ otherwise.

$X_A$ — Average tax rate on income produced by invested assets.

$X_U$ — Tax rate on income produced by insurance underwriting operations.
References


