An Economic Basis for Property-Casualty Insurance Risk-Based Capital Measures

(RBC Research Working Parties Report 5)

Robert P. Butsic

Abstract: A solvency measure is needed to consistently and fairly determine the level of an insurer's capital, which is needed for protection against defaulting on policyholder claims. There are several competing measures in current use, including VaR and the expected policyholder deficit. However, there is no published analytical method for selecting or calibrating any of these measures to produce a level of capital consistent with economic principles.

This paper develops an economic basis for selecting the solvency measure, and additionally determines how the measure can be calibrated to produce optimum capital. By maximizing policyholder welfare, a reasonable goal for regulation and corporate governance, I show that the optimal capital amount can be established by assessing the policyholders' perceived value of the expected default relative to the insurer's cost of holding capital. This optimality is achieved while allowing insurers a competitive rate of return.

The result is that the proper solvency measure is *adjusted ruin probability*, where the probability distribution of losses or assets is modified to reflect policyholders' *risk preferences*. The optimal level of the adjusted ruin probability is uniquely determined by the frictional cost of holding capital. With this foundation, I also show that the *subadditivity* property of a coherent risk measure is an unnecessary criterion for evaluating insurance solvency.

Under the policyholder welfare framework, the level of the adjusted ruin probability standard will vary by degree of policyholder risk aversion, interest rates, insurer income tax rates, amount of guaranty fund protection and other factors not considered in applying the above conventional solvency measures. I also discuss the relationship between the minimum regulatory level of capital and the insurer's optimal level.

Keywords: Solvency risk measures; policyholder welfare; optimal capital; adjusted probability distribution; certainty-equivalent losses; frictional capital costs; exponential utility; stochastic mean; subadditivity.

1. INTRODUCTION AND SUMMARY

The primary purpose of capital in an insurance organization is to protect policyholders, who in the event of insolvency, would not receive the full claim payment to which they are contractually entitled. Since there is an inverse relationship between the amount of an insurer's capital and the impact of insolvency on its policyholders, it is important to know (1) what kind of protection is desired, (2) how much protection is needed and (3) how much capital will provide the desired protection.

The first issue is addressed by selecting a *solvency measure*. The commonly used solvency measures are ruin probability, value-at-risk (VaR), expected policyholder deficit (EPD) and tail value-at-risk (TVaR). These solvency measures, which are discussed more thoroughly in section 5.4, use the probability distribution of losses and assets to characterize the harm to policyholders in an insolvency. The first two measures assess policyholder harm simply by whether or not an insolvency occurs. The latter two measures incorporate both the likelihood of insolvency and its average value provided that it occurs. Given a particular solvency measure, the amount of protection is addressed by choosing the *level* of the solvency measure (for example, with VaR a specific confidence level must be assumed). Selecting the solvency level is called *calibrating* the solvency measure. After calibration, the required capital follows directly using actuarial and statistical techniques applied to the probability distributions for the relevant balance sheet items.

There is much debate over the proper choice of risk measure: some adherents tout technical features, such as subadditivity (e.g., Artzner [1999]) or practical ones, such as ease of explanation or common use in other financial service industries. However, to my knowledge, there is no literature that establishes a particular solvency measure based upon economic principles. Furthermore, despite the widespread use of solvency measures, there has been no analytic basis for setting the level of the risk measure — calibration has been arbitrary, using judgment. Although there is a vast literature on implementing risk measures, especially VaR, each author inevitably assumes that the calibration level (say, 99% VaR over one year) is given. There is no discussion regarding how to determine the specific level.[1] This is surprising, since it is well known that there is a trade-off between the cost of having too much capital and the downside of not having enough capital. For example, few would believe that a 99.999% annual VaR standard is appropriate, since this level implies too much capital, which would be extremely costly to carry. Conversely, a 60% VaR standard would indicate an intolerable risk of insurer insolvency. Therefore, some intermediate value of the VaR standard must be best.

In this paper, I have addressed both the solvency measure and the calibration concerns by establishing an analytical framework that directly applies the above cost-benefit relationship for capital. Given that the economic objective in setting capital standards is to *maximize policyholder welfare* while allowing a fair return to the insurer's owners, I show that this goal implies that there is an optimal capital amount for each insurer. That amount depends on three key inputs: the probability distribution of losses and assets, the insurer's cost of holding capital and the risk preferences of the policyholders. If the values of these underlying variables are known, then the optimal capital is uniquely determined. The theoretical optimal capital amount then forms the basis for regulatory capital standards, internal insurer risk management and

---

[1] The 99.5% VaR standard of Solvency II is based on mimicking the annual default probability underlying a Standard & Poors BBB rating. However, this approach dodges the question, since there is no objective reason why this particular default rate (0.5%) is superior to that of any other rating (e.g., AA).

pricing applications.

The key result of this paper is that the appropriate solvency risk measure is *adjusted ruin probability* (or a simple function of it), using a transformed distribution of outcomes for each component risk of the insurer. The adjusted distribution incorporates policyholder risk preferences and has a much fatter tail than the original distribution. This gives a suitable heavy weight to the extreme outcomes and prevents engineering the tail shape to manipulate capital requirements. Therefore, on an economic basis, the best solvency measure is *none* of the conventional measures. I show that using a conventional risk measure, such as VaR or expected policyholder deficit, will overstate the capital for low-risk losses (and assets) while understating the amount of capital for high-risk components. The latter effect is more serious.

The level of the adjusted ruin probability standard is unique and is a function of the *frictional cost of capital*. Thus, the calibration is not arbitrary. However, the adjusted ruin probability is equivalent to a conventional ruin probability standard that varies by the volatility of the insurers' component risks and by its policyholder risk preferences. So, even though the adjusted ruin probability standard may be fixed for all insurers, the corresponding unadjusted ruin probability will vary by line of business and by insurer.

Under the policyholder welfare framework, the level of the adjusted ruin probability standard will differ by degree of policyholder risk aversion, interest rates, insurer income tax rates and other factors not considered when applying the above conventional solvency measures. Thus, the adjusted ruin probability standard is not static and will vary over time. Another consequence of the policyholder welfare basis is that the amount of guaranty fund protection will also influence insurer capital. This result is important and (to my knowledge) has been ignored in the previous insurance literature.

Although the optimal level of capital may be appropriate as a standard for internal insurer governance and for pricing applications, I discuss how the regulatory level of capital should be lower than the insurer's optimal level.

1.1 Outline

The remainder of the paper is summarized below:

Section 2 provides some historical background for the development of solvency risk measures as applied to insurance and other financial firms.

Section 3 develops the notion of consumer surplus (relabeled as consumer value) and the certainty-equivalent value for insurance losses. The basic idea here is that consumers are risk-averse and will pay more for insurance than the expected value of their losses. I have used these concepts, which are the economic foundation for insurance, to value insurer default from the policyholder's perspective. This section also relates the certainty-equivalent loss concept to utility theory.

Section 4 develops a simple one-period model of an insurer with risky losses and riskless assets and specifies the cost of holding capital. This section also formulates the premium charged to policyholders, which includes the frictional capital cost.

Section 5 shows how the consumer value of the insurance transaction is maximized by minimizing the cost of holding capital plus the value to the policyholder of the insurer's default. This section shows that the optimum amount of capital is determined from the adjusted (for policyholder risk preferences) ruin probability. It compares results from the adjusted ruin probability to those from conventional solvency measures and shows that the coherent risk measure property of subadditivity is not necessary for an economically valid insurance solvency risk measure.

Section 6 discusses how the results of section 5 can be extended to include asset risk, guaranty funds and multiple-period assets and liabilities.

Section 7 examines implementation issues in applying the above capital-setting methodology, including its use in regulatory risk-based capital.

Section 8 provides a brief conclusion.

2. HISTORY OF SOLVENCY RISK MEASURES

European actuaries have applied risk measures for decades. Ruin theory, also called collective risk theory, is a branch of actuarial science that studies an insurer's vulnerability to insolvency based on mathematical modeling of the insurer's surplus (capital). The theoretical foundation of ruin theory, known as the classical compound-Poisson risk model in the literature, was introduced in 1903 by the Swedish actuary Filip Lundberg.[2] Usually, the main objective of the classical model and its extensions was to calculate the probability of an insurer's ultimate ruin.

The ruin probability measure has seen some use for internal insurer risk management, but has not yet been directly used for solvency regulation (although the closely related VaR has).

The VaR measure was introduced in 1945, as a means of measuring bond portfolio risk.[3] In the 1970s, as leverage became widespread, securities firms sought more effective ways to manage portfolio risk. They wanted a single risk metric that could be applied consistently across asset categories, including derivatives, which were becoming increasingly complex. Concurrently, computing power became cheap enough to analyze large portfolios. However, VaR was still viewed as a theoretical tool.

During the early 1990s, concerns about the proliferation of derivative instruments, some well-publicized massive trading losses and the 1987 stock market crash spurred the field of financial risk management. Through its RiskMetrics service, JP Morgan introduced VaR to professionals at many financial institutions. Ultimately, the value of proprietary VaR measures was recognized by the Basle Committee, which authorized their use by banks for performing regulatory capital calculations.

VaR became common in the banking and finance industry in the 1990s onward. It is used to control the risk of the positions in investment portfolios or bank divisions for managers of these units. Supporters of VaR-based risk management claim that a major benefit of VaR is the improvement in systems and modeling it forces on an institution (see Jorion [2006]). For insurance, it is the measure used in Europe for the capital standards of the Solvency II regime.

After the 2008 financial crisis, VaR came under severe criticism (see Nocero [2009] and Einhorn [2008]), primarily because of abuses in its implementation. It has been argued that the

---

[2] See Lundberg [1903].
[3] For a more detailed discussion on the history of VaR, see Holton [2003].

2008 financial crisis was exacerbated by bankers misusing VaR. In order to reduce apparent risk levels (and thereby regulatory capital) for mortgage-backed derivatives, the banks engaged in "tail-stuffing," wherein the securities were purposely designed to increase the amount of risk in the tail, while keeping VaR at a low level. These abuses highlighted a technical weakness of VaR, in which very large extreme events are treated equally with events just large enough to breach the VaR confidence level.

The expected policyholder deficit (EPD) measure first appeared in the insurance financial literature in Butsic [1994]. This work arose from participation in the American Academy of Actuaries Property-Casualty Risk-Based Capital group, which advised the NAIC in its development of the current RBC method in the early 1990s. The concept developed as a response to a perceived deficiency in using ruin probability (or its VaR equivalent) as a solvency standard in that it did not incorporate the *depth* of an insurer's insolvency.

TVaR had a similar genesis in banking and investment management as the EPD in insurance. It was also a response to the same deficiencies in applying VaR. The above tail-stuffing abuses would have been severely mitigated under a TVaR metric. The TVaR concept saw implementation and became common in the 2000s. It is presently used as the solvency measure in Swiss insurance capital regulation.

Within the last decade or so, a new class of risk measures called *spectral measures* have been developed (see Acerbi [2002]). They are based on TVaR and include a risk-aversion component; i.e., extreme tail events are given weights that correspond to the investors' desire to avoid them. If the weights are large, more capital is required. I have used the concept of risk-aversion in this paper, although under a different context (i.e., optimization).

3. CONSUMER VALUE AND CERTAINTY EQUIVALENT LOSSES

The important concept of this section is that individuals (and organizations as well) will pay more than expected value to insure losses. The implied value placed on losses by the policyholder is called the *certainty equivalent* value. The economic gain from buying insurance is called the consumer surplus, which I have renamed as *consumer value*. We can measure the consumer value by using a modified version of the underlying probability distribution of losses. The adjusted distribution provides the means for determining the expected loss of consumer value due to the possibility that the insurer becomes insolvent. The adjusted distribution can be determined directly from a policyholder's utility function.

3.1 The Consumer Surplus Concept

The fundamental basis for insurance is that individuals are risk-averse: they are willing to pay more than the mathematical expected value of their potential loss in order to buy insurance.

As a simple example, suppose an individual is subject to loss on a home worth $100,000. There is a 1% chance of a total loss and a 99% probability of no loss. The mathematical expected value of the loss is $1,000 = 0.01 x 100,000. However, suppose that the homeowner is willing to pay up to $1,500 to an insurer to completely remove the risk of loss. Meanwhile, an insurer will charge only $1,100. The insurer is able to charge less than the policyholder (PH) is willing to pay because, through the law of large numbers, the risk to the insurer is reduced by pooling similar risks from other PHs.

The above three amounts ($1,000, $1,500, and $1,100) are important, and deserve a distinct nomenclature. The first is commonly called the *expected loss,* and in actuarial parlance the *pure premium*. The second is the *certainty equivalent expected loss* and the third the *premium.* In setting prices, actuaries normally include specific loadings for the insurer's expenses and for its bearing the risk, although the level of those loads is often limited by competition. In any event, the premium represents the insurer's price actually charged for bearing the risk.

The difference between the premium and the expected loss ($100 in the example above) is called the *provision for expenses and profit*. The difference between the certainty equivalent loss and the premium ($400) is called *consumer surplus* in the economics literature.[4] It is the difference between the total amount that consumers are willing and able to pay for a good or service and the total amount that they actually do pay (i.e., the market price for the product). If the consumer surplus is greater than zero, then the policyholder will buy insurance; otherwise the policyholder will self-insure.

The consumer surplus concept was introduced by Alfred Marshall in 1890,[5] and was designed to measure the welfare effects of economic policy. Standard microeconomics textbooks use the consumer surplus concept to equilibrate supply and demand. Consumer surplus applications are

[4] In finance, the equivalent concept is called the *risk premium* (See Panjer [1998]). I have not used this term here, since the term as used in insurance often represents the market price of risk (what the insurer charges for risk) and not what the policyholder is *willing* to pay.

[5] Marshall, Alfred (1947 [1890]): *Principles of Economics*. London, Macmillan.

common in welfare economics and government regulation, where cost-benefit analyses are needed.[6]

For insurance, consumer surplus represents the monetary benefit to the PH of having insurance. Since "surplus" commonly represents a different concept (capital) in insurance, for this paper I have renamed consumer surplus as *consumer value*.

I assume that in purchasing insurance, the PH will seek to maximize consumer value, which equals the difference between the certainty equivalent value of the potential loss, and the premium the PH must pay.

3.2 Mathematical Formulation of Certainty Equivalent Losses

For an *individual* PH having an exposure to an insurable risk, let y represent the size of a possible loss and $p(y)$ the probability that y occurs. The expected loss L is the summation of each possible loss times its probability: $L = \int_0^\infty yp(y)\,dy$.

Assume that for each possible loss amount y, there is a unique amount $\hat{y} = k(y)y$ representing the certainty equivalent (CE) loss. Call the term $k(y)$ the *certainty equivalent function*. Therefore, the CE expected loss (CEL), denoted by $\hat{L}$, will be the expectation

$$\hat{L} = \int_0^\infty \hat{y}p(y)\,dy = \int_0^\infty k(y)yp(y)\,dy. \tag{3.21}$$

Since policyholders are risk-averse, we have $\hat{L} \geq L$. The value $k = \hat{L} / L$, or the average of the certainty equivalent function, is a useful parameter. Notice that $k \geq 1$.

If the premium equals the expected loss, then the consumer value of the insurance equals the difference between the CEL and the expected loss, or $\hat{L} - L$.

The contribution to the CE expectation from loss size y in equation 3.21 can also be expressed as $k(y)yp(y) = y\hat{p}(y)$, where $\hat{p}(y) = k(y)p(y)$. Here, $\hat{p}(y)$ is a transformed probability that will give greater weight to large loss values and less weight to small values than $p(y)$, thus producing an expected value greater than L. An alternative version of equation 3.21 is then

$$\hat{L} = \int_0^\infty y\hat{p}(y)\,dy = \int_0^\infty k(y)yp(y)\,dy. \tag{3.22}$$

---

[6] For example, see Einav, Finkelstein, and Culleny [2010].

Notice that the certainty-equivalent probabilities $\hat{p}(y)$ are conceptually similar to the risk-neutral measures that form the cornerstone of modern finance theory.[7] Transformed probability measures have also been used to value insurance losses in pricing models (see Wang [1996] and Butsic [1999]). The CE probabilities can be considered as subjective weights attached by the PH to the various loss sizes. Since these weights are equivalent to probabilities, they must sum to 1. Thus, an important restriction on the CE function $k(y)$ is that $\int_0^{\infty} k(y)p(y)\,dy = 1$. Appendix A3 discusses this restriction further. It also explains why $k(y)$ depends not only on the particular loss size y, but on all the other possible loss sizes as well.

3.3 A Numerical Example

A numerical example will help to illustrate these concepts: suppose a PH faces a loss which can have three values {100, 400, 1200} with respective probabilities {0.60, 0.30, 0.10}. The corresponding CE function values are {0.80, 1.10, 1.90}, showing an increasing risk aversion with loss size. Table 3.3 below shows details of the CE expected loss calculation.

Table 3.3
Certainty-Equivalent Loss Calculation for Numerical Example

					Total
Loss Amount	y	100	400	1200	
Probability	$p(y)$	0.60	0.30	0.10	1.00
Expected Value Component	$y \cdot p(y)$	60	12	120	300
CE Function	$k(y)$	0.80	1.10	1.90	1.36
Certainty Equivalent Loss	$y \cdot k(y)$	80	440	2280	
CE Exp. Loss Component	$y \cdot k(y) \cdot p(y)$	48	132	228	408
CE Probability	$k(y) \cdot p(y)$	0.48	0.33	0.19	1.00

The expected loss L is 300 and the CE expected loss is 408, giving an average CE function value of k = 1.36 = 408/300. The CE probabilities have shifted from their unadjusted counterparts: the subjective chance of the small (100) loss drops from 60% to 48% and the subjective likelihood of the larger losses increases, from 30% and 10% to 33% and 19%.

---

[7] The risk-neutral concept was first introduced by Arrow and Debreu [1954]. The Black-Scholes option pricing model can be derived using risk-neutral valuation, as shown in Hull [2008] (pages 307-309).

3.4 Utility Theory and Certainty Equivalent Losses

By using basic principles from utility theory, we can derive some general properties for the CE function. There is a direct connection between utility theory and the certainty equivalent. As shown in Appendix A1, the CE function can be determined from the utility function and the loss distribution. The certainty equivalent and the expected utility formulations are dual processes; each can be determined using the inverse of the other.[8]

The first property is that the CE value function $k(y)$ *increases* with loss size y. This occurs because utility increases with wealth, as Appendix A2 shows.

Second, because of risk aversion, $k(y)$ increases at a growing rate: its second derivative with respect to loss size is positive. Appendix A2 discusses this property in more detail.

Third, as discussed below, the certainty-equivalent expected loss (and thus each $k(y)$ value) depends on the *variance* of the loss distribution.

If the variance is non-zero, then we apply the basic utility theory assumption that PHs are risk-averse. For $U(x)$ representing the utility of (wealth given) a loss x, this implies a downward-sloping utility function, (i.e., $U'(x) < 0$) and that the function is concave (i.e., $U''(x) \leq 0$). The *absolute risk aversion* function[9] is defined as

$$R_A(x) = U''(x) \,/\, U'(x). \tag{3.41}$$

Denoting the variance by σ^2, it is straightforward to show[10] that the certainty-equivalent loss is approximated by

$$\hat{L} \cong L + \tfrac{1}{2}\sigma^2 R_A(W_0 - L). \tag{3.42}$$

Here, W_0 is the initial wealth of the PH. This important result shows that $\hat{L} - L$ is (approximately) directly proportional to the *variance* of the loss distribution and also proportional to a measure of risk aversion. If the loss distribution is *normal* and the utility

---

[8] This duality is established by Yaari [1987], who determines that the CE value of a risky prospect can be found using an adjusted probability distribution.
[9] See Pratt [1964]. As shown in Appendix A1, the sign of the absolute risk aversion function is negative when utility is a function of wealth, but positive when a function of loss size.
[10] See Panjer et al. [1998], page 137.

function is *exponential* with risk aversion parameter a, then equation 3.42 is exact,[11] and as shown in Appendix A4, the CEL becomes

$$\hat{L} = L + \tfrac{1}{2}a\sigma^2. \tag{3.43}$$

To summarize the above results, we see that the CEL is a function of both consumer risk preferences and the variance of the loss distribution. Additionally, the certainty equivalent value function increases with loss size at an increasing rate. These properties are essential to determining the optimal capital for an insurer, as I develop in section 5.

Although the optimal capital results can be developed directly from the underlying utility function, I prefer the certainty equivalent approach to valuing risk aversion, since it is more direct than the utility method and provides a tangible, monetized conversion of the expected loss. Also, there is a *unique* CE value for each possible loss size, while under expected utility, any linear transformation of the utility function will give valid results. Thus, the expected utility of a particular loss size has no meaning by itself.

3.5 The Certainty Equivalent Expected Default Value

In order to determine the optimal capital amount for an insurer (in section 5) it is necessary to find the consumer value of the insurance contract. This entails knowing the certainty equivalent value of the expected default. Here, for simplicity, I assume an insurer with a single policyholder. The expected default, also known in the actuarial literature as the *expected policyholder deficit*, is

$$D = \int_A^\infty (y - A)p(y)\,dy\,. \tag{3.51}$$

Here, A represents the insurer's assets, which are assumed to be fixed (non-stochastic) for this application, and y is the individual policyholder loss size. The CE value of the expected default, denoted by $\hat{D}$, and abbreviated to CED, can be determined by finding the CE value of the loss actually paid by the insurer (allowing the possibility of default) and subtracting it from the CE of

[11] The exponential utility function can be expressed as $U(x) = -\exp[-a(W_0 - x)]$, where a is the risk-aversion parameter, W_0 is initial wealth and x is the loss value. Since $\exp(aW_0)$ is a constant and utility functions are invariant to scale transformations, the utility reduces simply to $U(x) = -\exp(ax)$. Thus the utility at any loss size x is independent of initial wealth.

the loss $\hat{L}$ without possibility of default. The CE value for losses limited to the amount *A* is $\hat{L}(A) = \int_0^A y\hat{p}(y)dy + A\int_A^\infty \hat{p}(y)\,dy$. Notice that, for losses larger than *A*, the amount of loss paid by the insurer is simply *A*. Here, to be consistent with the valuation of losses below the amount *A*, we must also use the subjective CE probability $\hat{p}(y)$ of the loss being greater than *A*. We cannot use the unadjusted probability *p*(*y*). To get the CED, we have $\hat{D} = \hat{L} - \hat{L}(A)$, or

$$\hat{D} = \int_0^\infty y\hat{p}(y)\,dy - \int_0^A y\hat{p}(y)\,dy - A\int_A^\infty \hat{p}(y)\,dy = \int_A^\infty (y-A)\hat{p}(y)\,dy. \qquad (3.52)$$

Here, the only difference from the equation 3.51 expected default calculation is the substitution of the CE probability $\hat{p}(y)$ or its equivalent *k*(*y*)*p*(*y*) for the unadjusted probability *p*(*y*). Notice that if *A* = 0, the CE default value equals the CE expected loss.

To illustrate the certainty equivalent default concept, assume a simple case where there is a 98% chance of a \$0 loss and a 2% chance of a \$1000 loss. The expected value of the loss is \$20 = 0.02(1000). Also assume that the PH has risk aversion governed by exponential utility: $U(x) = -e^{ax}$, where *x* is the loss size. The risk aversion parameter is *a* = 0.002. The utility of the 0 loss is –1 and the utility of the \$1000 loss is –7.389 = –exp[0.002(1000)]. The expected utility is –1.128 = 0.98(–1) + 0.02(–7.389). The certainty-equivalent loss is the loss size for which the actual utility equals the expected utility; thus the CEL is \$60.13: –exp[0.002(60.13)] = –1.128.

Now suppose that the loss to the PH is limited to \$900. The utility of this amount is –6.050 = –exp[0.002(900)]. The expected utility of the PH's retained loss is therefore –1.101 = 0.98(–1) + 0.02(–6.050). The certainty equivalent value for the retained loss is \$48.11. Therefore the CE value of the uppermost \$100 of protection is the difference between the CE of the entire loss and the CE of the retained loss:
\$12.02 = 60.13 – 48.11. Note that the expected value of the \$100 coverage is only \$2.00.

The \$12.02 also represents the *CE value of default* when only the first \$900 of loss is actually covered by the insurer. It is the difference between the CE expected value of the entire loss, minus the CE value of the coverage actually provided.

It is interesting to compare the result of covering the last \$100 of the loss (as above) with covering the first \$100 of loss.[12] If the insurer covers the amount *above* \$100 (i.e., the deductible), the utility of the retained loss is –1.221, with an expected utility of –1.004 and a CE of \$2.21. If the deductible is \$200, then the CE of the retained loss is 4.89, giving a \$2.69 = 4.89 – 2.21 CE value for the layer from \$100 to \$200. The CE value of each layer progressively increases as we move up to higher layers.

In this example, the ratio of $\hat{D}$ to D is 6.01 for assets of \$900, which is greater than 3.01, the ratio of the \$60.13 CEL to the \$20 expected loss. In general, the ratio of $\hat{D}$ to D will increase with the asset (and thus capital) amount.

This *tail leverage* is a consequence of PH risk aversion (which creates a fatter tail than for an unadjusted distribution), combined with the volatility of the loss distribution. To analyze the tail leverage in more detail, consider a normal distribution of losses for a PH with an exponential utility function. Appendix A4 provides a general method for determining the CE default value for a given loss distribution paired with a specific utility function. It also derives explicit results for the normal-exponential model.

Even with a small variance, the ratio of $\hat{D}$ to D can be large: assume a normal distribution with mean loss of 1000 and a 100 standard deviation. Suppose we have assets of 1100, which is 1 standard deviation above the mean, and that the risk aversion parameter is $a = 0.02$. From equation 3.43, the overall CEL is $1100 = 1000 + (.01)(200)^2$, giving k (the average CE function across all loss sizes) of 1.10. The straight expected default D is 8.33, but its certainty equivalent $\hat{D}$ is 57.39, a ratio of 6.89 to 1. For 2 standard deviations above the mean ($A = 1200$), we have $D = 0.85$ and $\hat{D} = 20.17$, for a ratio of 23.75. Table 3.5a shows results for these and other asset values:

[12] Notice that, given that the loss has *already occurred*, it doesn't matter whether the \$100 amount comes from retaining a \$100 deductible or sustaining a net \$100 loss on a \$1000 loss where the insurer pays the first \$900. The losses are completely equivalent to the PH. In fact, the certainty-equivalent concept makes no sense here, because *ex post*, the losses are certain. However, *ex ante*, the CE and expected utility concepts transform risky outcomes to fixed values that differ from the actual values that may occur.

Table 3.5a
Tail Leverage for Numerical Example;
Normal Loss Distribution, Exponential Utility (a = 0.02)

Assets	D	CED	Ratio CED/D	Ruin Probability (RP)	Adjusted Ruin Prob. (ARP)	Ratio ARP/RP
1100	8.33	57.39	6.9	15.866%	50.000%	3.2
1200	0.85	20.17	23.7	2.275%	25.161%	11.1
1300	0.04	4.44	116.2	0.135%	8.054%	59.7
1400	0.001	0.50	701.4	0.003%	1.291%	407.5

The adjusted ruin probability is the probability of default with the CE probability distribution used instead of its unadjusted counterpart. Figure A4 in Appendix A4 shows the probability densities for the normal distribution and its CE transformation. Here, it is clear that the tail of the adjusted distribution is much fatter than that of the parent normal distribution.

Table 3.5b gives results for a higher risk aversion (*a* = 0.04):

Table 3.5b
Tail Leverage for Numerical Example;
Normal Loss Distribution, Exponential Utility (a = 0.04)

Assets	D	CED	Ratio CED/D	Ruin Probability (RP)	Adjusted Ruin Prob. (ARP)	Ratio ARP/RP
1100	8.33	136.49	16.4	15.866%	68.281%	4.3
1200	0.85	77.25	91.0	2.275%	50.000%	22.0
1300	0.04	36.49	954.8	0.135%	31.719%	235.0
1400	0.001	13.01	18200.4	0.003%	15.883%	5015.0

Notice that even though the unadjusted probability and default values at the extreme tail are rather small, their CE equivalents may be meaningful. However, it is important to recognize that for a practical loss distribution and CE function (with extremely large losses truncated by policy limits and with CE factors restricted by wealth effects[13]) the CED values would not be as large as shown in this illustrative example.

[13] For example, bankruptcy laws limit the harm of an uninsured large third-party loss to an individual. An individual with $100,000 of net worth is likely to value a $1 million loss about the same as a $2 million loss.

3.6 Default Values with Multiple Policies

The previous sections have analyzed results for a single policyholder under the assumption that an insurer covers only that PH. Here, I extend the analysis to an insurer with multiple PHs. However, the analytical perspective remains that of the *individual* PH. But now if a default occurs, the default amount is *shared* among the individual policyholders in proportion to their respective loss amounts. Assume that the PHs are homogeneous, with the same loss distribution and risk preferences.

To illustrate CE valuation of multiple-policy default, I use the binary loss[14] example with exponential utility introduced in section 3.5. Appendix A5 develops the CE values for the binary loss model, including the CED and the adjusted ruin probability. Assume that we have two PHs with independent losses defined by the section 3.5 example: a loss of $1,000 with a 2% probability and zero otherwise. The CE values are determined from exponential utility with a risk aversion parameter a = 0.002.

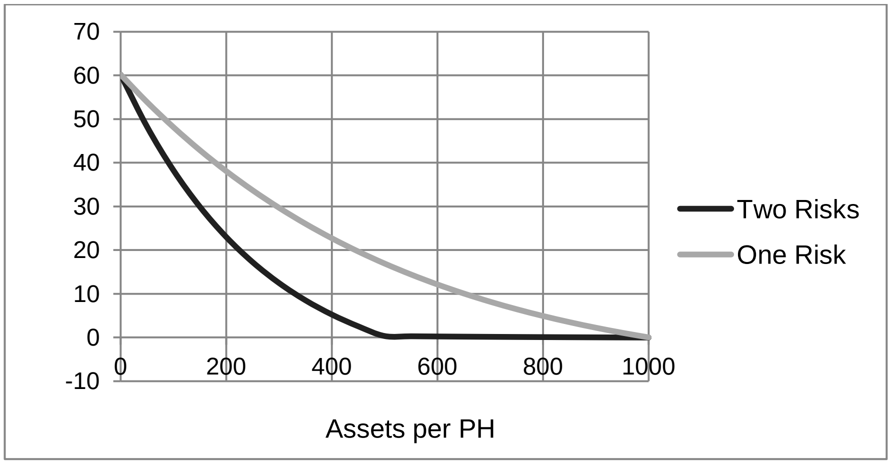
To illustrate the expected default calculation, suppose that an insurer covers these two PHs with an amount of assets equal to $200 per PH, or $400 in total. The default amount for a single PH (say, PH 1) depends on whether a loss occurs for PH 1 and whether a loss occurs for PH 2. If PH 1 doesn't have a loss, then there cannot be a default amount for PH 1 no matter what happens to PH 2. If PH 1 has a loss and PH 2 does not have a loss (with 0.0196 = 0.02 x 0.98 probability) then the default amount for PH 1 is $600 = 1000 – 400: all $400 of the insurer's assets cover the loss. If both PHs have a loss (with probability 0.0004 = 0.02 x 0.02) the total default amount is $1,600 = 2000 – 400, but it is shared equally, so PH 1 has a default amount of $800. Therefore, the expected default for PH 1 is $12.08 = 0.0196(600) + 0.0004(800). Notice that in the case where the insurer has only one PH (with $200 of assets), the expected default is larger: $16.00 = 0.02(1000 – 200).

To calculate the CE expected default value for PH 1, we use equation A5.5 from Appendix A5, getting $22.99. This compares to $38.05 for the single-risk insurer. Figure 3.6 compares, by asset value per PH, the CED per PH for the single-risk insurer to that of a two-risk insurer.

---

[14] This simple loss model with a general utility function is used in the influential paper on insurance market equilibrium by Rothschild and Stiglitz [1976]. Note that a one-year term life insurance contract has a binary loss distribution.

Figure 3.6
CE Default Values per PH by Asset Amount for a Single Risk and Two Risks
Binary Loss with Exponential Utility



Here, the CE expected default value per PH is lower when the risks are combined. This effect is a consequence of diversification, where the variance of losses per policy is reduced by adding risks to the insurer's portfolio.

In general, we can determine the per-PH CED for multiple risks by finding (or approximating) their joint CE probability distribution. This process is analogous to that of finding the unadjusted expected default for a portfolio of losses. In the case where the sum of the CE losses has the same distribution as a component CE loss (as in the normal distribution), the expected default calculation is straightforward. Otherwise, we must resort to approximation methods.[15]

The assumption of statistical independence will drive the expected default value to an extremely low level if the number of policies is large. However, in reality, insurance losses are correlated. They are subject to common factors such as inflation, regulation, the legal system and multi-loss events like catastrophes. Further, the mean of losses for a given line of business (or other subdivision of an insurer's risk portfolio) is not known; it must be determined empirically. This effect is an important case of parameter risk,[16] which adds to the uncertainty of standard

[15] One method is to assume that the distribution for the sum of the CE losses has the same distribution as the unadjusted losses, but with a different mean and/or variance. Manipulating these two parameters will generate a range of corresponding utility functions.
[16] For an example, see Meyers and Schenker [1983].

insurance risk models.

Appendix B develops a model of losses based upon a *stochastic mean*, where the expected value of the loss per policy is itself a random variable. For example, suppose the unconditional mean loss per policy is 1000, with a standard deviation of 300. This mean is considered a random variable: the 1000 amount is multiplied by a separate random variable with a mean of 1 and a 0.10 standard deviation. As the number of policies becomes large, the average policy will still have an expected loss of 1000, but will take on the risk of the stochastic mean variable, so it will have a standard deviation of 100 = (0.10)(1000). The influence of the original per-policy standard deviation of 300 vanishes. Thus, beyond a certain point the size of an insurer has little influence on the risk characteristics that determine the value of default.

4. A ONE-PERIOD PREMIUM MODEL

This section develops a basic insurer model and determines the premium, which includes the cost of holding capital. Extensions to the model are discussed in sections 6 and 7.

4.1 A Basic Insurer Model

To understand how optimal capital values can be determined, in this section I establish a simple, bare-bones model of an insurance company containing only a few necessary components.

I have assumed that the insurer's policyholders have the same individual loss distributions and the same risk aversion. Using the section 3.6 framework for multiple policies, this homogeneity implies that we can analyze portfolios of risks (even entire insurance companies) as if they were insurers having only a single policyholder.

In this model, I eliminate extraneous variables such as expenses, income taxes and investment returns by assuming that they are zero. In section 5, I include these components.

With no investment return, all assets are cash. I assume that the insurer is operating efficiently and thus the insurer's costs of holding capital can be passed on to policyholders as long as it improves their welfare. In fact, the *policyholders* determine the amount of capital and then pay for its associated costs through their premium, denoted by π. Thus, the owners of the insurer are indifferent to the amount of capital actually held by the insurer, since they are fairly compensated for its use.

The model is one-period: the premium and capital are determined at the beginning of the period and the actual[17] loss is determined at the end of the period. Further assume that there is no secondary insurer default protection for policyholders, such as a guaranty fund.

Its owners capitalize the company with an initial capital *C*. The initial assets of the insurer equal the premium from the policyholders plus the capital. Prior to the end of the period, the insurer's cost of holding its capital is expended, so that amount is not available to pay the PH's claims. The ending asset amount, denoted by *A*, thus equals the initial capital plus the premium, minus the capital cost. The loss amount is recognized at the end of the period and the default amount (if any) is determined accordingly.

If there is no cost to the insurer for holding the capital in the company, then the insurer will hold enough capital to exceed the largest possible loss, and thus there is no possibility of insurer default. In this case, the insureds simply pay a premium $\pi = L$, the expected value of the losses. With no default, their consumer value is maximized at an amount $\hat{L} - L$, the CEL minus the expected loss.

4.2 The Cost of Holding Capital

There is a cost to an insurer for holding capital to mitigate default risk. This cost is separate and distinctly different than the "cost of capital," which is the return expected by the capital suppliers (e.g., equity holders or bondholders) and is commensurate with the risk borne by these investors. To avoid the confusion created by the similar terminology, a useful name for the cost of holding capital is the *frictional cost of capital*, as defined by Hancock *et al* [2001]. The frictional capital cost (FCC) is the opportunity cost that accrues to the use of capital in an insurance firm, and which the investor would not incur if investing directly in financial markets. These costs include double taxation, financial distress, agency and regulatory restriction costs.

The primary component of the FCC for U.S. insurers is double taxation.[18] Of the above FCC components, it is also the easiest to determine empirically. To illustrate, assume that an investor provides \$100 of equity capital to an insurer, whose corporate income tax rate is 30%. The insurer invests the \$100 in assets *A* with an expected return of 6%. At the end of one year the expected return on the assets, after taxes, is \$4.20 = 100(0.06)(1 – 0.30). This amount is returned to the investor as a capital distribution, giving a net return on the capital of 4.2%. On the other

[17] For regulatory and other external party uses, a multiple-period model must address the fact that the insurer might not use an unbiased estimate of losses. Accordingly, the risk of under-reserving must also be assessed, with additional capital required beyond what is needed for the pure loss-variation risk addressed in this paper.
[18] See Harrington [1997].

hand, if the investor had invested directly in the assets *A*, rather than through the insurer, he/she would have received an expected return of $6.00.

The $1.80 difference must be made up by charging policyholders through additional premium. But the extra premium itself is taxed at the 30% corporate tax rate, so it must be grossed up to $2.57 = 1.80/(1 – 0.30). So, the double taxation component of FCC, as it applies to premium, equals $[rt/(1-t)]C$, where r is the insurer's investment return,[19] t is its income tax rate and C is the capital as defined in section 4.1. Notice that if the investment return is zero, then the double-taxation component of the FCC is also zero.

I assume that the FCC is at least equal to the above double taxation cost, and that all FCC components are proportional to the capital amount. Let z denote the FCC per unit of capital. Since the FCC must be borne by the policyholders in order to provide a fair market return[20] (the cost of capital) to investors, the premium must include an amount zC in addition to expected losses and other insurance expenses. Therefore, the *basic premium model* is

$$\pi = L + zC. \tag{4.21}$$

4.3 Fair Premium With Default

Since the frictional costs of capital must be passed on to PHs, they will not want the insurer to carry unlimited capital. Therefore, the insurer will have a non-zero probability of becoming insolvent. Then, in order to be actuarially fair, the basic premium ($L + zC$) must be reduced by the expected value of default D:

$$\pi = L - D + zC. \tag{4.31}$$

---

[19] In a competitive market, the investment return for the FCC will tend to equal the risk-free interest rate, despite the insurer's own expected return on investments. A higher return corresponds to greater risk and therefore requires a greater return to shareholders. Similarly, a high FCC cannot be passed on to policyholders if other insurers have lower investment returns and charge a smaller premium amount to cover double taxation costs.

[20] The expected return to investors, or the traditional "cost of capital" is built into the profit margin, another component of the fair premium. For simplicity, I have ignored it in the premium model. The profit margin can be directly embedded into the loss value by taking its present value at a risk-adjusted interest rate. The fair premium (with no default or expenses) will then equal the risk-adjusted PV of the expected loss plus the PV of the frictional capital costs. The cost of capital doesn't directly enter into the premium calculation.

So now the premium has three components: the base amount *L* is increased by the FCC and is reduced by *D*. Although, as I will discuss in section 5.1, the fair premium is approximated in practice by the basic premium. However, the basic model is not a competitive equilibrium[21] model, where the premium is a market-clearing price. Since I have assumed a zero interest rate and that market risk is captured by the loss value *L*, the market expected return on capital is zero. For equilibrium to occur, the premium must provide investors a zero expected return. Under the basic premium model, the expected return is *D*, since the policyholders' loss is the investors' gain. In a competitive market, this gain is reduced to zero by decreasing the basic premium by *D*. Thus, the fair premium satisfies both policyholders and investors, representing an equilibrium result.

[21] See Varian [1992], page 219.

5. DETERMINING OPTIMAL CAPITAL

Sections 3 and 4 have provided the ingredients to determine an insurer's optimal capital: we have a model for the value of policyholders' default as well as a specification for the insurer's cost of carrying capital. Since the capital amount governs the default value, we can balance these two factors to maximize consumer value.

5.1 General Model

For simplicity, I initially assume that the insurer does not deduct[22] the expected default from the premium, so we have the basic model $\pi = L + zC$. Since I have assumed that the interest rate is zero, the frictional capital cost rate z does not contain an income tax component (as discussed in section 4.2).

The value to the policyholders of their insurance is the certainty equivalent of the insured losses minus the cost of the insurance (which is a certain amount). Because insolvency is possible, the CE value of the actual coverage is the certainty-equivalent expected loss minus the certainty equivalent of the expected default. The consumer value (denoted by V) of the insurance transaction,[23] therefore, equals the certainty equivalent of the covered losses minus the premium:

$$V = \hat{L} - \hat{D} - \pi\,. \tag{5.11}$$

As we increase the amount of assets (by adding capital), the CE value of expected default decreases, while the premium (through the capital holding cost zC) increases with capital. This situation is a classic economics optimization problem, which can be solved by taking the derivative of V with respect to assets and setting it to zero.

Since $\hat{L}$ is constant with respect to a change in assets, taking the derivative of V with respect to A in 5.11 and setting the result to zero gives $-\partial\hat{D} / \partial A = \partial\pi / \partial A$. From Appendix C, which applies the derivative of an integral, we have $\partial\hat{D} / \partial A = -\hat{Q}(A) = -\int_A^\infty \hat{p}(x)\,dx$, so the general

---

[22] This is certainly the case in practice for the U.S. with regard to an *explicit* premium component for default. However, it can be argued that weaker insurers (with higher expected default amounts) will charge a lower premium to remain competitive.

[23] If the CE loss distribution is derived from some utility functions (such as the square root model), the CE value of a constant plus a random variable is not equal to the constant plus the CE of the random variable. In this case, equation 5.11 is an approximation. However, it is exact for CE values derived from exponential utility.

condition for optimum assets is

$$\hat{Q}(A) = \partial\pi / \partial A. \tag{5.12}$$

Here, $\hat{Q}(A)$ is the *adjusted ruin probability* (ARP), or the chance that losses exceed assets, under the transformed density $\hat{p}(x)$. The corresponding *unadjusted* ruin probability under $p(x)$ is denoted by $Q(A)$.

For the *basic premium* model, where the premium excludes the expected default, we have $\pi = L + zC$. Since L is constant with respect to assets, the derivative becomes $\partial\pi / \partial A = z(\partial C / \partial A)$. The assets available to pay claims equals the initial capital C plus the premium, minus the frictional capital cost, so we have $A = C + (L - zC) - zC = C + L$. Thus, $\partial C / \partial A = 1$, with the result

$$\hat{Q}(A) = z. \tag{5.13}$$

This rather simple result establishes that the optimal level of assets (and thereby capital) is determined by a risk measure that is an *adjusted ruin probability*. The ARP is a function of the probability distribution of losses and the policyholder risk aversion, as incorporated into the transformed density function. The risk measure is calibrated to the frictional capital cost rate z.

Given the optimal asset level from equation 5.13, the optimal capital is readily found by using the above relationship $A = L + C$.

For a *fair premium*, where the insurer deducts the expected default from the premium, we have $\pi = L - D + zC$ and $A = L + C - D$. Taking derivatives of these two expressions, and noting that $\partial D / \partial A = -Q(A)$, we get $\partial\pi / \partial A = Q(A) + z(\partial C / \partial A)$ and $\partial C / \partial A = 1 - Q(A)$. Thus, $\partial\pi / \partial A = z[1 - Q(A)] + Q(A)$ and equation 5.12 gives

$$\theta(A) = \frac{\hat{Q}(A) - Q(A)}{1 - Q(A)} = z \tag{5.14}$$

as the condition for optimal capital.

From section 3.5, tables 3.5a and 3.5b show that the transformed ruin probability $\hat{Q}(A)$ is much larger than the unadjusted ruin probability $Q(A)$, particularly at lower ruin probabilities (which correspond to the high safety levels that would be required in practice). Thus, $Q(A)$ can be set to zero and equation 5.13 may be considered as an approximation[24] to the optimal capital condition for a true fair premium.

Notice that the level of insurer expenses doesn't affect the optimal capital, as long as the expenses are a function of the expected losses and not capital. Let the premium be $\pi = L - D + zC + e_0 + e_1 L$, where e_0 and e_1 are constants that determine expenses. The derivative of V in equation 5.11 will be the same with or without expenses, since the derivatives of L, e_0 and e_1 with respect to capital are all equal to zero.

Equation 5.12 is general in scope and can be used for alternative premium formulations. For example, if the frictional capital cost zC is not consumed prior to default, then the optimal capital is determined from $\hat{Q}(A) = z / (1 + z)$, which approximates equation 5.13.

If an insurer's policyholders have *heterogeneous risk preferences*, (but with the same loss distribution) the optimal capital can still be calculated using equation 5.11. However, the premium for each PH will differ. For a given capital amount, each PH will have a specific CED value (based on their risk aversion) along with a share of the joint capital cost. The share is allocated to each PH via their willingness to pay (determined from their respective consumer values). This gives a higher premium for the more risk-averse PHs: in effect, they pay the low risk-aversion PHs in order to use a high capital amount. However, this result is theoretical, since normally an insurer does not charge different premiums for policyholders with identical loss characteristics. Accordingly, in practice the optimal capital for a group of insureds must be based on an *average* (weighted in some fashion) of their risk preferences. In this case, low and high-risk aversion PHs will have consumer values that are less than the theoretical optimum.

Consequently, they might improve their consumer values by moving to an insurer whose other PHs have similar risk preferences to their own; i.e., to the extent that insurers incorporate capital costs into their premium, PHs are best served by choosing an insurer whose capital strength suits their needs. In finance, this grouping behavior is called the *clientele effect*. For insurance, it has implications for pricing (section 7.1) and regulation (section 7.2). However, further analysis of this topic is beyond the scope of this paper.

---

[24] Using z = 2%, and the normal-exponential model from section 3.5 (with risk aversion of 0.02), the optimal capital under equation 5.13 is 379.73. Under equation 5.14 it is 379.56, a difference of only 0.045%.

5.2 The Effect of Income Taxes

Another variation to the premium model includes the effect of income taxes, which, as discussed in section 4.2, can be the major component of the frictional cost of holding capital. In order to pursue practical applications, the impact of income taxes on optimal capital must be addressed. Appendix D develops the result for optimal capital in this case. As in the general model with a constant frictional cost of capital, the optimal asset amount is found by setting the CE ruin probability equal to a constant value:

$$\hat{Q}(A) = \frac{rt}{1+r-t}. \tag{5.21}$$

Here, *A* is the end-of-period assets (before subtracting the loss and income taxes), *t* is the income tax rate and *r* is the riskless investment rate of return.

Equation 5.21 is important because it establishes a benchmark for practical applications. The effective corporate income tax rate for insurers will be less than the current nominal 35% highest marginal rate, due to the ability to defer taxes on capital gains, shelter income using municipal bonds and other measures. Assume that the effective rate is 30%. As discussed in section 4.2, the appropriate investment benchmark is the Treasury rate (which should be matched to the average liability duration: about 3 years for U.S. property-casualty insurers). The 3-year rate has varied from about 1.5% to 6% over the past 10 years. Consequently, the optimal adjusted ruin probability has been in the range of about 0.4% to 1.8% over this period.

The corresponding optimal *unadjusted* ruin probabilities can be smaller than the 0.4% to 1.8% range, as indicated in the section 3.5 examples (tables 3.5a and 3.5b). However, several factors (e.g., using a more realistic loss distribution,[25] incorporating the regulatory constraints in section 7 and including the effects of guaranty funds) will increase the unadjusted ruin probabilities. Consequently, the above range of frictional capital costs appears to be broadly consistent with more subjective solvency measures such as the Solvency II 99.5% VaR standard, which translates to a 0.5% unadjusted ruin probability. In other words, the overall required capital for

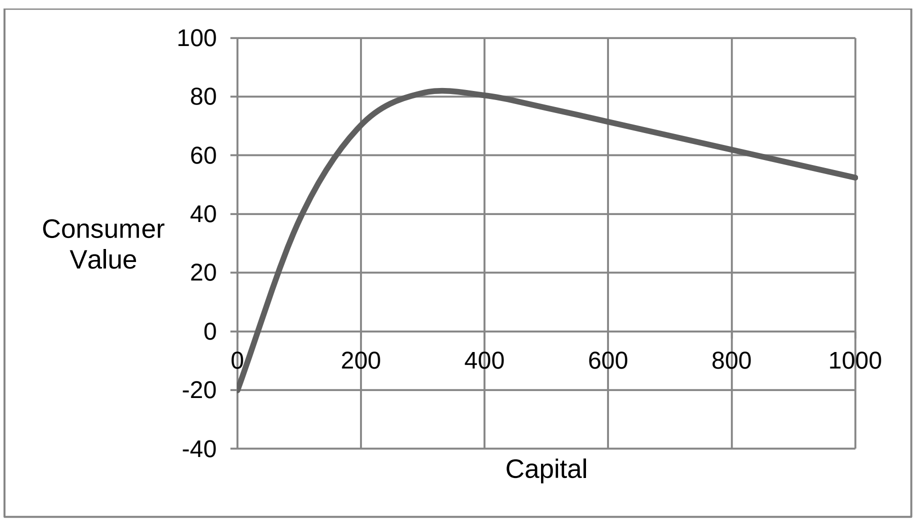
[25] The normal and lognormal distributions that I've used for illustration have unlimited losses. In practice, insurance coverage imposes policy limits. Consequently, policyholders must absorb the high end of their losses, which carry the greatest CE values. This effect reduces the CE default value and increases both the ARP and the unadjusted ruin probability corresponding to the optimal capital.

an average insurer under the ARP risk model is not inconsistent with current practice.

5.3 Numerical Examples

Here, I've used the normal distribution with exponential utility from section 3.5. For the same parameters (1000 mean, 100 standard deviation, 0.02 risk aversion) and with a capital cost rate $z = 0.05$, we get the optimal ARP of 5% and optimal capital of 330.66 using equation 5.13 and equation A4.9 in Appendix A4. The CE expected loss is 1100 and CED value is 2.46 from equation A4.8. The premium is 1016.53 = 1000 + 0.05(330.66), and so the consumer value of the insurance contract is 81.00 = 1100 – 1016.53 – 2.46. Figure 5.3 below shows how the consumer value of the insurance varies by amount of capital.

Figure 5.3
Consumer Value of Insurance by Capital Amount in Numerical Example
Normal Loss/Exponential Utility Model



Notice how the shape of the CV curve is steep at low levels of capital and flattens with higher capital. This behavior indicates that, beyond the optimal capital level, adding more capital has only a slight impact on PH welfare. Thus, the relative insensitivity of the optimal capital might be exploited in practical applications, where it could be necessary to use approximate values for some of the underlying variables.

Table 5.3 shows how the optimal capital varies by standard deviation (SD) and risk aversion:

Table 5.3
Optimal Capital by Standard Deviation and Risk Aversion in Numerical Example

Risk Aversion	SD = 25	SD = 50	SD = 100	SD = 200
0.005	44	92	205	493
0.010	46	103	247	661
0.020	51	123	331	1007
0.040	62	165	504	1720
0.080	83	252	860	3186

As we would expect, the optimal capital increases with both risk aversion and the volatility of the losses.

5.4 Comparison to Other Risk Measures

The conventional solvency risk measures can be considered as equal to or as simple functions of the *tail moments* of the loss distribution. Here I define the *n*th tail moment as

$$MT(n) = \int_A^\infty (x - A)^n p(x) dx \,, \qquad (5.41)$$

where $p(x)$ is the density function and A is the assets, as defined earlier. Notice that if assets are zero, the tail moment equals the regular moment of the entire distribution.

Observe that the ruin probability is the 0^{th} tail moment and the expected default (policyholder deficit) is the *first* tail moment. Define the *valuation level* as the predetermined numerical value of the tail moment, such as 1% or 5%, that produces the desired level of assets. In other words, if the risk measure is ruin probability (RP) and the valuation level is 1%, then $MT(0) = 0.01$ and we solve equation 5.41 for A.

VaR is the amount of assets such that $\alpha = 1 - MT(0)$ or 1 minus the RP, where α is the VaR *confidence level*. Tail value-at-risk, or TVaR,[26] is the amount of assets equal to VaR + $MT(1)/MT(0)$ at the α confidence level. Thus, the conventional risk measures are simple functions of the tail moments with n equal to 0 or 1.[27] In the following discussion, I use the ruin probability and the expected default ratio to loss (D/L) to characterize the tail-moment based

[26] Notice that TVaR and the EPD are not equivalent and will not necessarily produce the same capital amount. EPD includes only the amount of loss *exceeding* the asset threshold *A*, while TVaR also includes the portion of the loss *below A* for losses exceeding *A*.
[27] The value of n need not be an integer. For example, with $n = 0.5$, the weight of the tail losses will be somewhere between that of a ruin probability and an expected default measure.

solvency measures.

It is noteworthy to compare the adjusted ruin probability (ARP) measure to the common tail-moment risk measures. Assume that both a straight (unadjusted) ruin probability (RP = 1 – VaR) and an unadjusted EPD measure are also used to determine capital. Further assume that all three measures (ARP, RP and EPD) are calibrated to give the same capital for a typical insurer. We observe the valuation level for each risk measure implied by the capital and keep it fixed as we change the variance of the loss distribution.

For this exercise, let the typical insurer have the characteristics of the section 5.3 example, with a 5% ARP providing optimum capital of 330.66. This capital amount implies that RP = 0.047% and the EPD/Loss ratio = 0.0012%. We fix all three measures and consider two other insurers, also having normally distributed losses. One insurer has low-risk policyholders with a standard deviation (SD) of 50; the other has high-risk PHs with a 200 SD. We apply each risk measure to these insurers, and compare to the typical insurer:

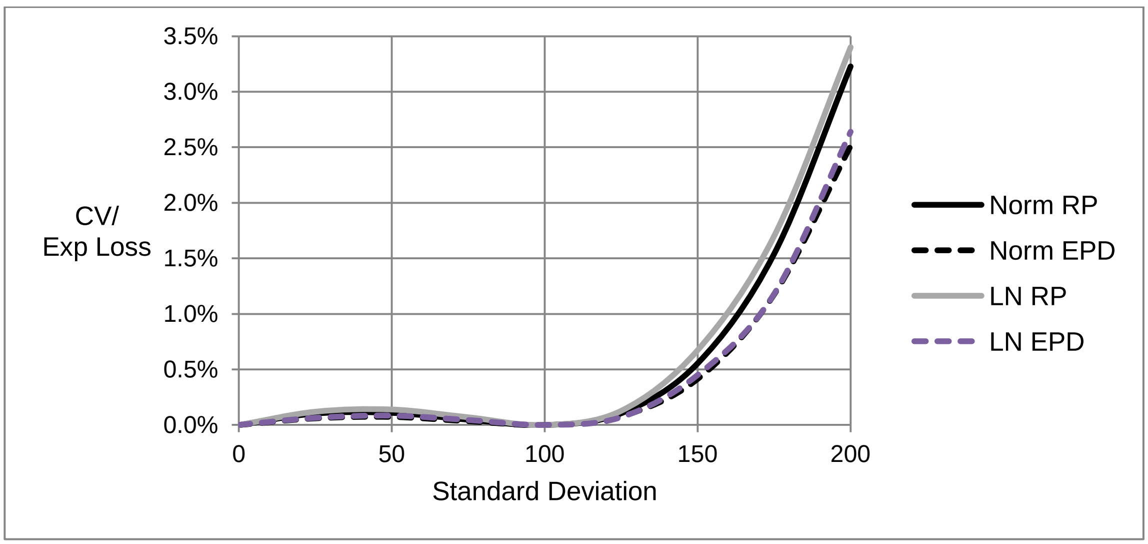
Table 5.4a
Comparison of Risk Measures Calibrated to Normal SD = 100
Adjusted Ruin Probability = 5%

SD	Optimal Capital	Implied RP Using Optimal Capital	Implied EPD Ratio Using Optimal Capital	A Capital: RP Standard	B Capital: EPD Standard	A % Diff.	B % Diff.
50	123.3	0.685%	0.0112%	165.3	156.0	34.1%	26.6%
100	330.7	0.047%	0.0012%	330.7	330.7	0.0%	0.0%
200	1007.0	0.00002%	0.000001%	661.3	697.0	-34.3%	-30.8%

Notice that the low-risk insurer has more capital under RP and EPD, with the high-risk insurer having less capital, compared to the optimal capital under ARP. Because the resulting capital for the low and high-risk insurers is not optimal, using a conventional risk measure reduces *consumer value*. For the high-risk insurer, using RP (or its VaR equivalent) lowers the CV from 344.07 to 311.79, a decrease equaling 3.2% of the expected loss. For the low-risk insurer, the CV drops from 17.70 to 16.60, a reduction of only 0.1% of the expected loss. Similarly, applying the EPD ratio reduces CV by 2.5% of expected loss for the high-risk insurer and 0.1% for the low-risk insurer. Note that this disparity between the high and low-risk insurers is due to the fact that, for the same level of risk-aversion, high-risk policyholders gain more consumer value from insurance than low-risk PHs.[28]

We get similar results with a lognormal[29] distribution paired with exponential utility. Here the typical insurer again has a mean loss of 1000, standard deviation of 100 and a risk aversion of 0.02. The initial calibration gives RP = 0.037% and the EPD ratio = 0.0013%. Varying the SD and keeping the risk measures constant, we get similar results to the normal case. Figure 5.4 shows the loss in CV/ expected loss for the lognormal case compared to the optimal ARP capital. For comparison, it also provides the results for the normal distribution.

Figure 5.4
Loss of Consumer Value by Using RP and EPD Measures
Compared to Optimal ARP Result; Percentage of Expected Loss
Normal and Lognormal Distributions



For both distributions, the CV loss is slightly less (for the high-risk PHs) using the EPD, compared to the RP measure. To summarize the above results, we see that, compared to the ARP

[28] The CE value of the loss (being related to the variance) is lower with respect to its expected value as variance decreases. At the extreme, with a zero variance, there is no difference between the CEL and the expected loss and so the difference in CV from using any risk measure must be zero.

[29] Here I've approximated the lognormal distribution using the binomial option pricing method as described by Panjer [1998], page 246. The CE ruin probabilities and default values are directly determined using the method of Appendix A4. Also, I have adjusted the risk aversion parameter to produce, for each standard deviation value, the same CE loss as with the normal distribution.

standard, the conventional risk measures overstate the optimal capital for low risk lines and *understate the capital for above-average risk lines*. The latter effect is more serious, because the under-capitalization can produce a meaningful loss in consumer value. For the low-risk lines, the loss in CV appears to be negligible.

To gain further insight regarding the deficiencies of the conventional measures, assume that we have calibrated all three risk measures to produce the same capital for an insurer, as in the above example (e.g., Table 5.4 with a 100 standard deviation). Now we increase the size of a particular possible loss $x > A$ in the tail by an amount Δ.

If RP or VaR is the standard, there is no change in capital since the loss increment does not affect the RP. Yet policyholders are worse off. This effect was exploited in a perverse way during the 2008 financial crisis, when financial firms "stuffed the tails" to keep their apparent risk low (see section 2.1). To manipulate the tail probabilities, the firms designed securities to have a low probability of loss, but with an extreme loss size when the loss occurred.

Under an EPD or TVaR standard, the effect of a loss increment is more subtle. Suppose that we take two "slices" of the tail, one with smaller losses and the other with larger losses. The widths of the slices (i.e., the probability that the losses in the intervals will occur) are selected so that the probability of losses being in each interval are equal. Let x_1 be the average loss in the lower interval and x_2 be the average loss in the upper interval. If we simultaneously adjust losses so that x_2 increases by Δ and x_1 decreases by Δ, the expected default amount remains the same (as does the default probability). Note, however, that this operation increases the variance of the tail losses. Since the certainty equivalent value function $k(x)$ is concave upward, we have $k(x_2) > k(x_1)$, and thus this adjustment will increase the CE value of the default. Therefore, more capital is needed, even though the EPD stays constant.

The preceding discussion shows that, given a particular loss distribution, under the ARP method it is not possible to "engineer" the tail to produce an artificially low capital amount. If the tail is altered by reinsurance or other financial techniques, the CE value function will automatically produce the proper capital as long as the firm uses the correct loss distribution (adjusted for PH risk preferences).

A practical disadvantage of using the ARP measure is that it does not translate to any fixed conventional standard. For example, to get the correct optimal capital under the 5% ARP standard, the appropriate unadjusted RP in table 5.4 ranges from 0.00002% to 0.685%. The EPD ratio has a similar large range. Although this ARP feature presents no difficulty in calculating capital, it may create problems in comparing results to conventional solvency measures.

5.5 Subadditivity

To clarify risk definition in financial economics, theoreticians have described several properties for a risk measure. A *coherent risk measure*[30] is a function that satisfies monotonicity, subadditivity, homogeneity, and translational invariance. *Subadditivity* has become the most important of these properties when applied to risk measures used in practice.

The subadditivity (SA) property requires that the value of the risk measure for the combination of two risks is less than or equal to the sum of the risk measure values taken separately. For insurance capital requirements, this means that when two risks (or risk portfolios) are combined, the required assets derived under the risk measure must be less than or equal to the sum of the assets derived from applying the risk measure to the risks individually. To be consistent with the individual PH focus of our analysis, the subadditivity requirement can be restated: if two risks are combined, the assets *per risk* under the risk measure cannot be greater than assets for either risk taken separately under the risk measure.

Assume that two PHs have identically distributed losses. Let $RM_1(A)$ represent a risk measure that is a function of assets per policyholder A for a PH of an insurer with one risk and $RM_2(A)$ for a PH of an insurer with two risks combined. For two different asset amounts $A_2 > A_1$, we have $RM_1(A_2) \leq RM_1(A_1)$ and $RM_2(A_2) \leq RM_2(A_1)$; i.e., increasing assets decreases the value of the risk measure (this is the monotonicity property of a coherent risk measure). A *subadditivity violation* will occur when

$$RM_2(A) > v > RM_1(A), \tag{5.51}$$

where v is the valuation level of the risk measure and A is any asset amount. Under this inequality, a SA violation occurs because, for both measures to equal v, assets in the combined-risk insurer must *increase* and assets in the single-risk insurer must *decrease* (this effect is shown graphically in figure 5.51). Therefore, assets per PH in the combined-risk insurer will be greater than for the single-risk insurer.

---

[30] See Artzner [1999] for a discussion of coherent risk measures with insurance applications.

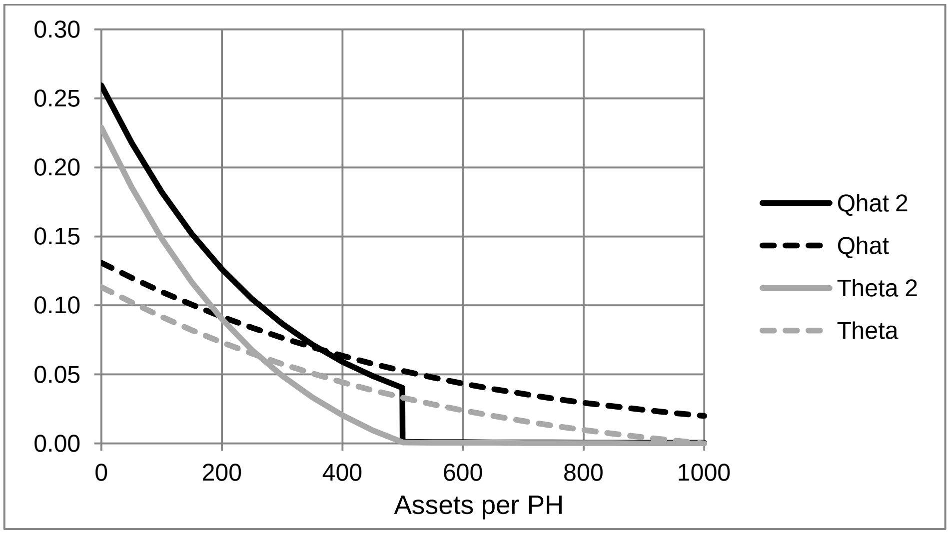
A classic bond risk example,[31] which has a *binary loss distribution*, is used to illustrate SA violation. I have modified this investment illustration to represent insurance by using the section 3.6 example with two independent binary risks each with a 2% probability of a $1000 loss. Suppose that the VaR measure is set at 97%. This means that a single risk must have at least a 3% chance of loss in order to require assets (and thereby capital). Otherwise no assets are required to back the loss. Thus, with a 2% chance of loss, no assets or capital are required. However, if two independent risks are combined, the probability of a loss is 3.72% = 2(0.02)(0.98) + 0.02(0.02) and therefore $1,000 of total assets (for both PHs) is required. This reduces the default probability to 0.04% and satisfies the 97% VaR valuation level. Subadditivity is violated here since more assets are required per PH for the combined risks ($500 each) than for either of the separate risks (zero).

Using the ruin probability counterpart to VaR as the risk measure, we have $v = 0.03$. Let $Q_1(A)$ denote the RP for the single risk and $Q_2(A)$ the RP for the combined risks. Applying equation 5.51, we see that for A (assets per PH) from 0 to $500, $[Q_2(A) = 0.036] > 0.03 > [Q_1(A) = 0.02]$. Thus, subadditivity is violated. For assets above $500, we have $0.03 > [Q_1(A) = 0.02] > [Q_2(A) = 0.004]$, so there is no SA violation. Also, for a valuation level $v > 3.68\%$ (required assets are zero for both the single or combined risks) or $v < 2\%$ (required assets are $1000 per PH for the single risk and $500 for the combined risks), there is no SA violation.

For the *adjusted ruin probability* measures $\hat{Q}(A)$ or $\theta(A)$ in equations 5.13 and 5.14, there also is SA violation for a range of valuation levels. Applying equations A5.4 and A5.6 in Appendix A5 to the parameters for this example, we get Figure 5.51, which graphically shows the pairs $\hat{Q}_2(A)$, $\hat{Q}(A)$ and $\theta_2(A)$, $\theta(A)$. These are labeled respectively as Qhat 2, Qhat, Theta 2 and Theta (for simplicity, I have dropped the subscript 1 denoting a single risk).

[31] This example is from Albanese [1997] and has been used by Artzner and others.

Figure 5.51
Adjusted Risk Measures for Section 3.6 Example



Notice that a risk measure based on PH risk preferences can be a *continuous* function of the asset amount, even if the underlying loss distribution is discrete. This occurs because the values of the CE expected default and its derivatives are continuous with respect to the amount of assets if the underlying CE function $k(y)$ or the equivalent utility function is continuous.

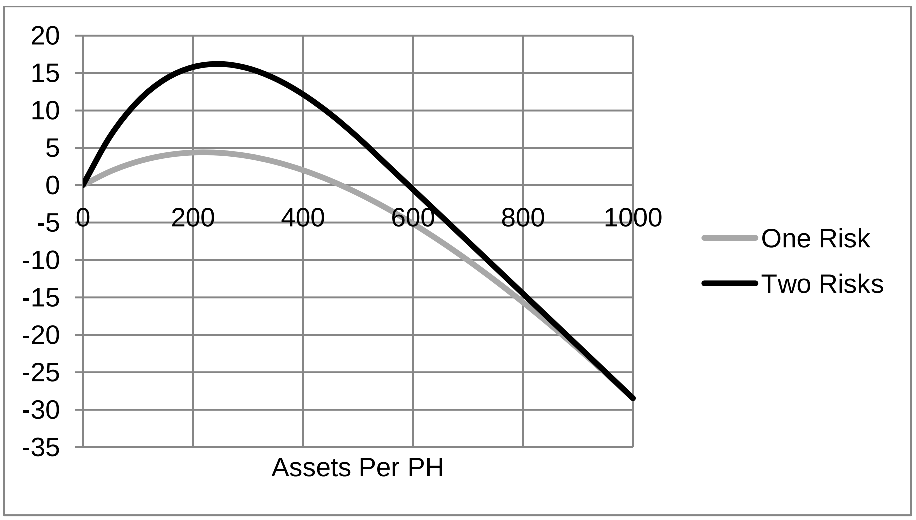
The graph shows that $\hat{Q}_2(A) > \hat{Q}(A)$ for $\hat{Q}(A) \geq 6.78\%$, where assets per PH are less than \$364.44. Subadditivity can be violated in this region. Similarly, $\theta_2(A) > \theta(A)$ when $\theta(A) >$ 6.34%, corresponding to $A < 260.80$.

Does the SA violation create any negative effects for the policyholders in this example? To answer this, suppose that the frictional cost of capital is $z = 7\%$, which is the valuation level for the adjusted risk measures. It exceeds the above critical values of 6.78% and 6.34%, so there will be a SA violation for each measure. From section 5.1, we have the premium and from section 3.6 the CED values by asset amount. Thus, the consumer value for any asset value can be readily found.

For a *fair premium*, when assets are zero, the expected default equals the expected loss of \$20. The premium and capital are also zero. Since the CED equals the \$60.13 CE loss, the consumer value of the insurance is zero. When assets equal the \$1,000 loss value, the expected default is zero, but the premium equals $L + zC$ and capital equals assets minus the premium (the capital cost zC is not an available asset to pay losses). Thus, the capital is \$980 and the premium is

\$88.60 = 20 + 0.07(980). The consumer value is negative: –\$28.47 = 60.13 – 88.60. Between these asset value extremes, the CV will have an optimal value. Figure 5.52 shows the per-PH CV by assets for a single risk and for two combined risks.

Figure 5.52
Consumer Value per PH by Asset Amount for Single Risk and Two Risks
Fair Premium Model; z = 7%

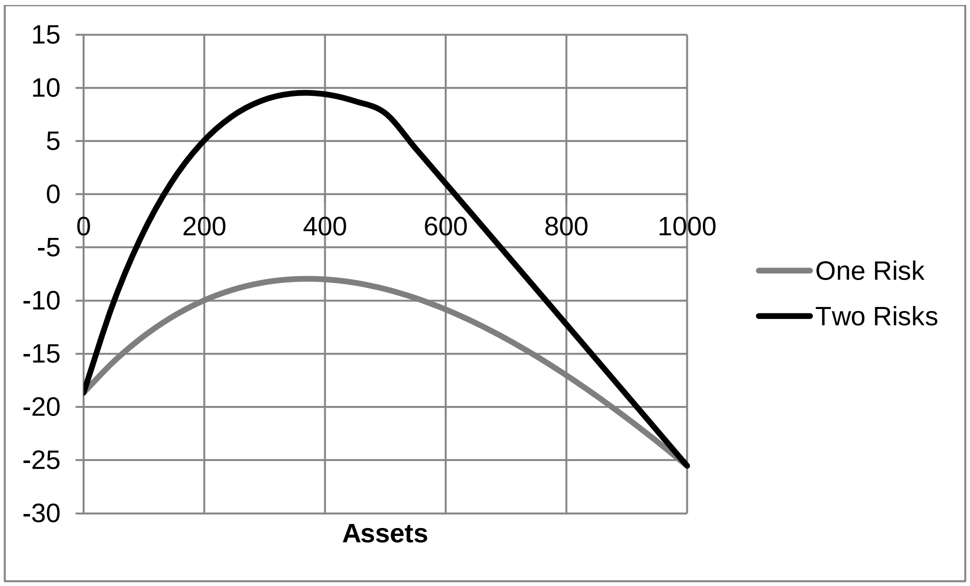


Here, for both cases, the optimal CV is positive. This is achieved with a per-PH asset value of \$219.51 for one risk and \$244.59 for the two combined risks. The corresponding respective optimal capital amounts are \$215.12 and \$234.90. These optimal amounts are derived directly by solving for $\theta(A) = 0.07$ in equation 5.14, using the Appendix A5 relationships.

This example clearly shows that the subadditivity criterion is violated, since *more* assets (or capital) are required per policyholder for the combination of two risks than for the single risk. It is also clear that the PHs are *better off* with the SA violation under the ARP risk measure, since their optimal consumer value is higher when the risks are combined.

For the *basic* (non-fair) *premium* case, where the risk measure is $\hat{Q}(A)$, we get similar results, with the optimal assets for a single risk being \$347.43, which is less than the \$356.15 optimal per-PH assets for the combined risks. However, since the premium is not actuarially fair, the CV is lower than for the fair premium case, for both the single risk and the combined risks:

Figure 5.53
Consumer Value per PH by Asset Amount for Single Risk and Two Risks
Basic (Non-Fair Premium) Model; z = 7%



Notice that the CV for the single risk here is negative for all asset values, indicating that the risk is not insurable — the PH is better off without insurance. However, when the two risks are combined, they become insurable.

With lower values of z in this binary loss example, subadditivity is not violated. For instance, with fair premium and z = 2%, optimal per-PH assets for the single-risk insurer are $648.35, compared to only $401.65 for the two-risk insurer.

In the financial economics literature an economic justification for the subadditivity constraint is that "if a firm were forced to meet a requirement of extra capital that did not satisfy this property, the firm might be motivated to break up into two separately incorporated affiliates, a matter of concern for the regulator."[32] But the above examples show that PHs are clearly better off being combined — with the consequent subadditivity violation requiring extra capital — than being insured separately. In fact, for the basic model, violating subadditivity turns uninsurable risks into insurable ones.

[32] See Artzner [1999], page 14.

Nevertheless, given that a major purpose of the subadditivity constraint is to promote aggregation of risks, the underlying economic basis of the adjusted risk measures used here will always indicate that PHs are better off[33] (or no worse off, if the risks co-vary) when risks are combined. Therefore, these measures promote the *spirit* of the SA constraint.

To summarize this section, I have shown that risk measures based on PH risk-preference can violate subadditivity, but when they do, the result makes perfect sense economically. Further, policyholders are never worse off when risks are combined — a fact that does not depend on the risk measure used to determine capital. Therefore, we must conclude that *subadditivity is an unnecessary criterion for an insurance solvency risk measure.*

6. EXTENSIONS OF RESULTS

The analysis in the preceding sections is based on a simplified model of an insurer, and concentrates on estimating optimal capital for insurance losses only. I have omitted some important elements that must be addressed before implementing the concepts for regulation, internal insurer risk management, pricing or other applications.

This section discusses some of these important missing pieces. The scope of this paper does not permit a full development of the topics, so for each of them I have stated the issue and outlined the general direction of the analysis. Although these areas present some difficulties, they can be attacked using the major idea of section 5: optimal capital can be determined by trading off the cost of holding the capital and the value to policyholders of having the capital.

6.1 Asset Risk

The treatment of asset risk adds another dimension to the section 3 formulation of default risk for losses, where I assumed that assets were riskless, with a zero return. Now assume that the insurer has a portion of its investments in risky securities. For initial assets of A_0, the ending asset value will be random, with an expected value of $A \geq A_0$ (the reward for bearing market risk is an expected return exceeding the risk-free rate). Further assume that the insurer keeps the total asset and loss risk constant, so that it varies its asset (and capital) level by changing the amount of riskless assets.

[33] For the basic premium model, at any per-PH asset level A, the difference between the combined-risk CV and the single-risk CV equals the difference between the single-risk CED and the combined-risk CED. This is so because the premiums for the two cases are identical. Thus the result will be non-negative. For a fair premium, the difference is the basic premium differential *minus* the difference between the unadjusted expected default amounts. This too will always be non-negative.

The certainty equivalent value of a risky asset is the converse of that for a risky loss: it is *less* than the expected value of its possible payoffs. Therefore, with a zero risk-free investment return, the beginning of period certainty-equivalent value of the ending asset amount equals the expected value of the ending amount, where the expectation is taken over an adjusted probability distribution. For an average investor, the expected value must equal the market value of the assets. Thus, this calculation *removes the expected return* from the ending asset values. The adjusted, or risk-neutral distribution (see Hull [2008]), reflects the concept that the *ex ante* perceived value of a particular asset outcome will depend on the *economic scenario* that generated the asset value.[34] For example, a low asset value may correspond to an unfavorable economy where a dollar is worth more; in that case the investor will weigh the result more heavily than the symmetrical high asset value.

Assume that the policyholder has the same risk preferences as the typical investor. For each ending asset value A_i, the CE expected default is $\hat{D}_i = \hat{L} - \hat{L}(A_i)$, where from section 3, $\hat{L}(A_i)$ is the CE amount of loss limited to A_i. Therefore, the unconditional CE of default over all asset values is the sum of the conditional values $\hat{D}_i$ weighted by the risk-neutral probabilities $p_A(A_i)$ for the asset values occurring:

$$\hat{D} = \hat{L} - \int_0^\infty p_A(x)\hat{L}(x)dx. \qquad (6.11)$$

Note that, if losses and assets are correlated, each $\hat{L}(A_i)$ will derive from a different expected loss corresponding to each A_i. For most applications using continuous distributions, the above integral can be evaluated with numerical techniques.[35] The result gives the CED for a particular initial asset amount A_0. The optimal capital occurs when $-\partial\hat{D} / \partial A_0 = z$.

To illustrate this calculation, I return to the normal-exponential example from section 5.3. Here, the expected loss is 1000 with a standard deviation of 100, the risk aversion is 0.02 and the capital cost rate is 5%. Assume that the insurer has 400 of risky assets, also with a normal distribution, with a 5% expected return and a SD of 80 (the *volatility*, or SD per unit of risky assets is 20%). The remaining assets are riskless, with a zero return. Also, asset risk is

[34] Note that a risky asset whose return is uncorrelated with market returns will generally not command a positive expected return above the risk-free rate.

[35] For the calculations here, I have used a discrete binomial approximation to the normally distributed ending assets. Each asset value determines the CE value of the loss limited to the asset amount using the conditional bivariate normal loss distribution (where the mean loss is a linear function of the asset/loss correlation). The overall CED was determined by inverting the expected utility of the conditional CE values using the risk-neutral asset probabilities.

independent of loss risk.

Assuming that the risk-neutral distribution is also normal, its expected value equals the initial asset amount.[36] The optimal capital with the risky assets becomes 356.89, which is 26.23 greater than the 330.66 for the case with riskless assets. The optimal CV at 78.88, is 2.13 lower than with the riskless assets.

Generally, under the policyholder welfare framework, the capital is always higher and the CV is lower when an insurer has a risky investment portfolio. Since on average, insurers' investment managers cannot beat the market, financial theory indicates that there is no benefit to PHs for holding the risky assets. In other words, the optimal investment portfolio has *only riskless assets*. Then why do insurers in practice have risky assets? To resolve this puzzle, there are several hypotheses, including: the insurers may believe that their investment managers can individually beat the market (although collectively they cannot); management compensation schemes reward positive income without penalizing negative income; and insurers may build an above risk-free investment return into their pricing models.

However, even though risky assets may lower the consumer value for PHs, the reduction may be small enough so that it is not material. For example, in the above calculation, the loss in CV is only about 0.2% of the expected loss. So the risky asset conundrum is theoretically interesting but generally may not be a practical issue. Regulators, rating agencies and insurance management recognize that a large amount of risky assets is imprudent. Especially large risky investment portfolios require additional capital whose costs cannot be passed on to PHs in a competitive market.

An important point to make here is that the risk-neutral probability distribution removes the positive expected excess market return from the CE default calculation. If an insurer increases its asset risk through securities whose return is uncorrelated with the market, the expected default will rise and more capital is required. Thus it is essential for an insurer to maintain a diversified investment portfolio.

Also worth observing is that both insurance losses and investment returns are not considered to be normally distributed; often a lognormal model or some other skewed distribution is used to approximate these variables. It may not be possible to represent the joint distribution in a tractable form. A more realistic application of the CE approach for total asset and loss risk will require a more elaborate method, such as a simulation model.

---

[36] If one does not reduce the expectation to the beginning asset level, the result can be an optimal capital amount that is *less* than that for the riskless assets. In this case the additional capital required for risky assets is *negative*.

6.2 Guaranty Funds

This topic deserves a full treatment in a separate paper. There is much academic literature[37] on the economic basis and design of guaranty funds, but I have found none that analyzes the effect of the funds on insurers' capital requirements.

Under the policyholder welfare concept, guaranty funds will substantially reduce the optimal capital for an insurer. To see why this is so, consider an economy in which all policyholders of all insurers are completely covered by a single guaranty fund (GF). Clearly, no policyholder will suffer an uncovered loss unless the entire industry defaults. Thus, the aggregate capital for *all insurers* is used to protect any individual policyholder. Contrast this situation with the opposite extreme, where no policyholder has GF protection. Here, the policyholder has access only to the capital of his/her own insurer. In this case, the insurer needs much more capital than in the full GF situation.

Under a GF within the U.S., essentially all of the capital for each insurer in a particular state is pooled to provide default protection for policyholders. The coverage is limited (usually $300,000 per policyholder for most lines of business), but some lines, such as workers compensation, have unlimited protection and others, such as surety, have no protection. For lines protected by the $300,000 limit, the GF coverage can vary significantly. For example, assuming a lognormal distribution with a 5.0 coefficient of variation, policyholders in a line with an average loss per policy of $1,000 (e.g., personal insurance) will have 99.53% of their expected losses covered by the GF, with only 0.47% exposure to the insurer's default. However, those policyholders in a line with an average loss per policy of $5,000 (say, commercial insurance) will only have 95.96% covered, with a 4.04% exposure. Relative to their expected loss, the ratio of non-covered losses for the two lines is 8.6 to 1. So, for this example the presence of GF protection is a major factor in assessing the optimal capital for the two lines.

Also, the GFs themselves can become exhausted[38] in extreme events, since there is an annual limit to the amount they can assess the solvent insurers. Thus, in order to estimate optimal capital for a particular insurer, the *risk of GF exhaustion* must be analyzed. When this threat is considered, a much higher portion of default risk becomes attributed to *extreme events*.

[37] Cummins [1988] is one of the most often cited references. He argues that a pre-funded GF system is superior to the predominant post-failure assessment model in current use. However, based on the analysis here, a properly constructed RBC implementation might produce equivalent results.

[38] The term guaranty *fund* is somewhat of a misnomer. The vast majority of the state GFs merely assess other solvent insurers; they have no "fund" to pay claims. Thus, the GFs themselves cannot become insolvent.

Consequently, modeling these becomes paramount. The extreme events can be national or world-wide in scope (e.g., a financial crisis or a deep pricing down-cycle) or regional (such as a natural catastrophe).

An important implication for analysis is that, with GF protection the optimal capital depends not only on the risk of a policyholder's own insurer's default, but also the *default risk of the other insurers* covered by the fund. Therefore, in analyzing the effect, say, of catastrophes on capital, one must also estimate the effect of the same catastrophes on the other insurers. This modeling might be simplified by using a *default correlation* parameter (DCP) for the insurer, where the parameter measures the correlation between the insurer's default and that of the remainder of the insurers in a particular state. A value of zero for the DCP would mean the insurer' capital can be modeled as a stand-alone entity. At the other extreme, a value of 1 would mean that whenever the insurer defaults, the GF is exhausted due to the simultaneous defaults of other insurers.

Although the protection afforded by a GF is considerable (the expected loss above a $300,000 threshold is a small fraction of the total expected loss), the certainty equivalent value of the above-threshold amount is large relative to its expectation. Consequently, the value of the GF protection for the policyholder is reduced somewhat in comparison to its straight expected value.

Other considerations in modeling the effect of GFs on optimal capital are that there may be a degradation of service (e.g., a delay in settlement) when, upon insolvency, a policyholder's claim is transferred to another claims management firm or that there may be market disruption from the insolvency of a large insurer. These effects can be incorporated into the model by modifying equation 5.11 to include a coefficient greater than 1 for the CE of the default.

The analysis of optimal capital under a GF should feature an additional term in the premium calculation: the expected GF assessment for the failure of other insurers. This is an unavoidable cost to the policyholder that is paid *ex post*, so its value is stochastic at the time of the policy is purchased. Note that, for a specific insurer, the expected GF assessment depends on the capital levels of the other insurers, so the optimal capital level for that insurer is influenced by both the GF assessment and the ability of the other insurers to provide GF protection for the insurer.

The presence of guaranty funds adds another element to the regulator's role of solvency protection. By monitoring capital for a particular insurer, the regulator must not only protect the interests of that insurer's policyholders, but also the interests of the *policyholders of the other* *insurers* who would be assessed in the event of the particular insurer's demise.

Summarizing this section, GF protection adds two important variables to incorporate into optimal capital determination. The first is the degree of GF coverage, which varies by line of business. More capital is required for lines with less GF coverage. Second, the optimum capital for an insurer depends on the default risk of other insurers covered by the GF. Thus, the correlation of the default risk with other insurers will affect capital: the higher the correlation, the higher the capital amount. Including these variables requires analysis of an insurer's data by state: for example, to properly determine catastrophe risk capital, the effect of GF exhaustion must be estimated for each state where there is material exposure.

6.3 Multiple Periods

This topic is the subject of much debate in the actuarial and insurance finance literature.[39] There are two camps: one side advocates using an annual time horizon, wherein capital is only needed to offset default risk based on market values over the *upcoming year*. The other side argues that capital is needed to offset the risk that cash flow will not be sufficient to pay claims over the *entire duration* (the runoff horizon) required to settle the liability. This topic also deserves a separate paper, so here I have only outlined a procedure that will establish optimal multi-period capital.

The single period model, with the above and other extensions, should suffice to determine optimal capital for lines of business, such as property, whose claims are paid over a short duration. For liability insurance, workers' compensation and life insurance, we need to expand the model to encompass long-duration claims. Although the long-duration contracts can be modeled in continuous time, it makes sense to use a discrete, multi-period time frame. This is because accounting time frames determine the valuation of insurer assets and liabilities and hence capital. The annual time period is especially important, so for practical purposes, we need to examine long time-horizon asset and liability risk over one-year time increments. For shorter time periods (e.g., quarterly), a similar analysis will apply.

With long-horizon risks, we can use the same fundamental assumptions that drive optimal capital for a single period. The main point is that the optimal capital over several periods still depends on the balance between capital costs and the certainty equivalent value of default.

A key component of the analysis is that the value of a long-horizon risk element (e.g., losses

[39] For a good discussion of this topic, see Lowe et al. [2011].

or assets) is stochastic (i.e., random) at the end of every period. Assume that we know the probability distribution for the evolution[40] of the risk element value and there are *n* periods. Suppose the risk element is a loss with expected present value *L*. So, if the insurer begins the first year with an optimal capital level C_1 and the loss value at the end of the year happens to be $L_1 \neq L$, then the risk of default (either in the next period or ultimately) will change if the original capital remains the same. Thus, the capital must be changed accordingly to regain the optimal position.

This process leads to a sequence of capital amounts $\{C_1, C_2, \cdots C_n\}$ corresponding to the sequence of loss values $\{L, L_1, \ldots L_{n-1}\}$. It will also produce a series of CE expected default amounts at the end of each year driven by the loss values: $\{\hat{D}_1, \hat{D}_2, \ldots \hat{D}_n\}$.

For each of these sequences of value realizations, we can determine the present value of the consumer value. However, we need a rule or strategy to determine the capital amount at each period C_t (based on L_t) that optimizes the expected present value over *all possible realizations* of the $\{L, L_1, \ldots L_{n-1}\}$ sequence. If we find a single strategy that does this, then we have settled the issue of setting capital for a long-horizon risk element.

This type of problem can be solved by a process called *discrete time stochastic dynamic programming*.[41] One of the techniques used in this method is backward induction, where one starts at time *n* – 1, finds the optimal decision rule, then steps backward to time *n* – 2, finds the optimal decision rule at that stage, and so forth, all the way back to the beginning of the first period. If the stochastic process is regular (such a random walk with a constant drift), then the decision rule at each stage will likely be the same.

7. APPLICATIONS

The preceding sections have developed a theoretical framework for determining optimal capital for insurers. This section discusses several issues involved in applying the theory in a practical setting.

There will be some applications of these results that are not directly related to setting the level

[40] This process is analogous to a discrete model of interest rate evolution, where the value at any period will generate multiple possibilities for the next period. Graphically, the structure will look like a tree, with each successive period having more branches.

[41] See Birge and Louveaux [1997].

of an insurer's capital, such as capital allocation,[42] but I will leave those topics for further research. Also, since capital is an essential ingredient in pricing models, the optimal capital results will be relevant to that application;[43] however, this work is outside the scope of this paper.

7.1 Implementation of Results

The results here are new and somewhat contrary to current practice. In my view, there are three major obstacles to implementing them.

The first is that there is little empirical work, especially for insurance, in quantifying policyholders' risk preferences. All we know for certain is that insurance consumers are risk-averse, and will pay more than expected value for their coverage. In the absence of empirical evidence the best we can do is to assume a functional form for the risk aversion process, such as the exponential utility used in sections 3 through 5. (I do not necessarily advocate this model; I have used it because it is familiar and provides mathematically tractable results). This can be calibrated to a presumed certainty equivalent factor (k) for an individual PH based on judgment.

Second, the adjusted ruin probability risk measure is not as easy to understand as the conventional ruin probability measure. That fact that a constant ARP translates into different conventional ruin probability standards for different risk elements (e.g., lines of business) may be difficult for some to comprehend, and may undermine acceptance of the results.

Third, the analysis has unearthed several currently unrecognized variables (e.g., the frictional cost of capital, which reflects interest rates and income tax rates; also, the level of guaranty fund protection) that should be considered in setting capital. Incorporating them will require considerably more data-gathering and analysis than is presently done. Based on the analysis of sections 3 through 6, Table 7.1 shows the key variables that should be considered in establishing optimal capital levels:

---

[42] The Myers and Read [2001] capital allocation method uses an expected default that is not adjusted for policyholder risk preferences. Incorporating this element will allocate relatively more capital to lines with more risk-averse policyholders.

[43] Section 5.1 has shown that, to the extent that PHs tend to select insurers having capital levels based on their risk preferences, then the insurer's *actual capital* level (rather than an industry standard) will be relevant to setting prices.

Table 7.1
Summary of Key Variables for Setting Capital Requirements
N = None; L = Low; M = Moderate; H = High

	Inputs				
	Risk-free investment rate	Effective tax rate	Risk-aversion parameter	Fraction of default not covered by GF	Insurer-GF default correlation
Variation by:					
Time	M	L	N	L	L
Insurer	L	L/M	L/M	L/M	M
Line of Business	L	L/M	M	H	H
State	N	N	N	L/M	H
Econ. Scenario/ Extreme Event	M	L	M	M	H
Difficulty of Estimation:	L	L	H	L/M	M/H

Here I have provided my subjective estimates of the importance of each relevant variable introduced in this paper, and how much each variable varies by time, insurer, line of business and other factors. I have also indicated the difficulty of estimating the parameters (in modeling the capital requirements) for each variable. I have assumed that incorporating the risk aversion component is done simply, perhaps with a single parameter. Similarly, the correlation between insurer default and guaranty fund exhaustion is modeled with a single parameter. Notice that other variables, such as the loss distribution, are quite important but are currently considered when assessing risk-based capital.

Although it may appear that the conventional risk measures are better than the ARP because they are simpler (needing fewer variables to evaluate), this is not the case if one accepts the policyholder risk-preference basis of this paper: these variables were always important, but simply were not recognized by the conventional risk-based capital methodology.

7.2 Regulatory Role in Capital Standards

The preceding sections have addressed finding the optimal capital for an insurer. In an efficient market, insurers will gravitate toward these optimal levels without regulatory intervention. However, the market is far from efficient from the perspective of maximizing policyholder welfare, and the involvement of regulators is often necessary. An important role of the regulator is to mimic the *outcome* of an efficient market, or at least to mitigate the effects of the market imperfections.

Consequently, this means attempting to maximize policyholder welfare while maintaining a competitive market for insurance. The goal of this process is to approximate the optimal capital generated in an efficient market. Using risk-based capital standards, the regulator has the authority to force an insurer to maintain a minimum level of capital. If the insurer fails to achieve the desired capital, the regulator can impose various restrictions on the insurer's operations,[44] including shutting down the insurer. So, the regulator will want to set the intervention thresholds at levels that will tend to produce optimal capital levels.

This means that the more severe (i.e., shut-down) thresholds will be *lower* than the optimal capital amount for an insurer. For example, if the regulator sets the shut-down level at the optimum level, the insurer's management will need to carry *more* than that amount of capital. There are several reasons why the stringent thresholds should be lower than either the optimal capital level:

(1) An insurer operating above, but near the shut-down level would have a strong chance of being forced out of business if the business does not perform well over the following year. Therefore, its management will try to maintain a sufficient clearance above the threshold to minimize this possibility.

(2) With a high threshold, there is strong possibility of misidentifying companies that are actually strong as weak. Harrington (in H. Scott, ed., 2005) discusses this problem. Regulators will tend to value this type of error (Type 2) more than the converse (Type 1 error) where weak companies are incorrectly identified as being strong. This type of forbearance will lower the stringent threshold levels.

(3) An insurer cannot operate near a stringent threshold without the market knowing about it. Operating near the threshold will signal that the insurer is weak, resulting in loss of business. Consequently, the insurer will become weaker, and the result will tend to be a self-fulfilling prophecy.

Additionally, the regulator cannot be certain that an insurer with a low capital level is truly undercapitalized or the insurer's policyholders have low risk aversion. In the latter instance, the low capital amount could be appropriate for those PHs. This possibility requires a lower stringent

---

[44] Under the current U.S. risk-based capital framework designed by the National Association of Insurance Commissioners (NAIC), there are five control levels (thresholds), ranging from no regulatory action when capital exceeds 200% of the base RBC amount, to mandatory takeover of the insurer at 70% of the base RBC.

threshold than if all PHs had the same risk preferences.

Since the insurers will tend to carry more capital than the stringent intervention thresholds, these must be set low enough to induce insurers to generally carry the optimum level of capital. It will be difficult to quantify the relationship between the optimal capital and the regulatory thresholds that will produce the optimal capital. So, as is currently done, it may be necessary to use expert judgment to establish the threshold levels,[45] even if the optimal capital itself can be estimated reasonably well.

I observe here that it is quite possible that insurers will in practice carry more capital than the optimal amount required to benefit policyholders. This will imply a wide gap between the stringent-threshold regulatory amount and the amount typically held by insurers. Because of incentive conflicts, the interests of insurance management, shareholders, regulators and rating agencies may differ from those of policyholders. For example, shareholders may be interested in protecting the franchise value of the insurer and may shortchange the interests of current policyholders to obtain future profits. Insurance management may desire capital sufficient to protect their private interests (e.g., future employment prospects) and may care more about the *chance* of insolvency than the *expected amount* of the default. Regulators and rating agencies have a vested interest in limiting the frequency of insurer insolvencies, since the failures can be viewed as a breakdown of supervision or of the rating system.

The above incentive conflicts are exacerbated by the presence of guaranty funds, since the GFs allow for a somewhat painless insolvency experience from the policyholder perspective, but not painless to the other parties such as regulators or insurance management.

8. CONCLUSION

Based on maximizing policyholder welfare, it is possible to determine the optimal capital that an insurer should carry. To accomplish this, the appropriate solvency risk measure is ruin probability, using an adjusted probability distribution that reflects policyholders' risk aversion. The level of the adjusted ruin probability standard depends only on the insurer's frictional cost of holding capital. The assessment of underlying probability distributions of losses and assets, however difficult, is a standard actuarial problem. Determining the frictional cost of capital is a straightforward financial economics problem. On the other hand, estimates for the policyholder

---

[45] Setting the proper threshold level is conceptually another optimization problem: find the level that will create the best overall policyholder welfare, recognizing the above market-disrupting effects.

risk preferences are presently not available. This presents a ripe new area for empirical research.

The results of the analysis here establish that a number of variables, which are not considered in conventional risk measures, are important to properly establish an insurer's optimal capital. These features are absent when applying conventional solvency risk measures such as VaR or expected policyholder deficit. Incorporating these new factors is also a rich opportunity for further study.

Finally, although I have focused on property-casualty insurers in particular, the underlying principles will apply to other financial institutions as well. These entities have primary stakeholders such policyholders, depositors and investors. As with property-casualty insurance, the welfare of these parties is governed by the same general relationship between consumer value and the cost of carrying capital.

APPENDIX A: UTILITY THEORY AND CERTAINTY EQUIVALENT LOSSES

In this appendix I show the relationship between utility theory under risk and the certainty equivalent valuation of insurance losses. The utility of a wealth amount W is designated by $u(W)$ and the initial wealth of the PH by W_0. Accordingly, the utility of wealth given a loss y is $u(W_0 - y)$. However, since we are concerned with insurance losses here, it is convenient to redefine the utility to be a function of the loss amount:
$U(y) = u(W_0 - y)$.

Since the utility theory axioms have $u'(W) > 0$ and $u''(W) \leq 0$, with the derivatives taken with respect to wealth, when we convert to the utility of loss basis, we get
$U'(y) < 0$ and $U''(y) \leq 0$. Here, the derivative is taken with respect to the loss size y. These results are developed in Appendix A2. On the utility of loss basis, the relative risk aversion function $R_A(W) = -u''(W)\,/\,u'(W)$ becomes $R_A(y) = U''(y)\,/\,U'(y)$.

A1: Finding CE Values From a Utility Function

The *expected utility* (of wealth) is

$$EU = \int_0^\infty U(y)p(y)\,dy\,. \qquad \text{(A1.1)}$$

The certainty-equivalent wealth is the amount of wealth that gives as actual utility, the same amount as the expected utility. Thus, $CEW = U^{-1}(EU)$, where U^{-1} is the inverse of the utility function. The certainty-equivalent wealth, in turn, equals the actual wealth minus the certainty equivalent of the expected loss, or

$$CEW = U^{-1}(EU) = W_0 - \hat{L}\,. \qquad \text{(A1.2)}$$

Suppose that a policyholder faces a loss of size y with probability p, or no loss with probability $1 - p$. We want to determine the *certainty equivalent* amount corresponding to y, or $k(y)y$, where $k(y)$ is the CE function defined in section 3.2. The CE loss is

$$\hat{L} = k(0)(1-p)\cdot 0 + k(y)p\cdot y = k(y)py\,. \qquad \text{(A1.3)}$$

From equations A1.2 and A1.3, we can determine the CE function value:

$$k(y) = \frac{W_0 - U^{-1}(EU)}{py} \ . \qquad \text{(A1.4)}$$

Thus, it is possible to determine the CE value of an individual loss amount directly from a utility function and the initial wealth.

To illustrate, consider the utility function $U(y) = \sqrt{W_0 - y}$. The initial wealth is 1600 and a loss of 1200 has a 10% probability. The expected wealth is 1600 – 0.1(1200) = 1480. The utility of the initial wealth is $40 = \sqrt{1600}$. The utility if the loss occurs is $20 = \sqrt{1600 - 1200}$, so the expected utility is 38 = 0.9(40) + 0.1(20).

The certain wealth corresponding to the expected utility of 38 is $1444 = U^{-1}(38) = 38^2$. Therefore, the certainty equivalent value of the expected loss is 156 = 1600 – 1444. From equation A1.4, the CE function value for the 1200 loss amount is *k*(1200) = [1600 – 1444]/[0.1(1200)] = 1.3. Notice that this value corresponds to a CE probability of 13% for the 1200 loss amount.

A2: The Shape of the CE Value Function

Because utility theory axioms impose constraints on the shape of the utility function, these restrictions will be reflected in the shape of the corresponding certainty-equivalent function. From Appendix A1, the CE function is related to the *inverse* of the utility function, so the properties of inverse functions will govern the translation from utility to certainty equivalence.

The first utility axiom is that utility increases with wealth: the derivative of utility with respect to wealth is $u'(W) > 0$. This means that utility declines as the loss size *y* becomes larger (i.e., $U'(y) < 0$) and thus the certainty-equivalent function value *increases* with *y*: $k'(y) > 0$.

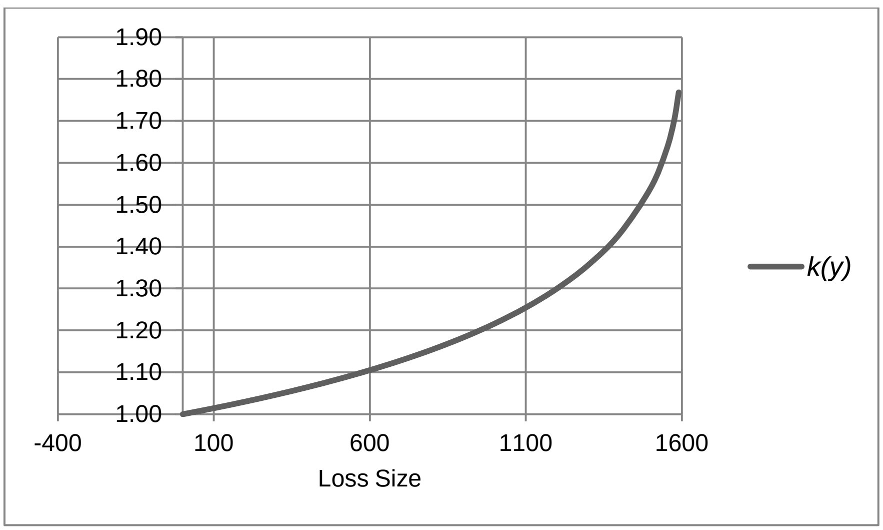
The second utility property is that, because individuals are assumed to be risk-averse, the second derivative of the utility function with respect to wealth is negative:. $u''(W) \leq 0$. This means that utility declines as the loss size *y* increases, but at an *increasing* rate. This property translates to a CE function that increases at an increasing rate: $k''(y) \geq 0$.

Returning to the Appendix A1 example, we can vary the loss size from 0 to 1600, keeping the other parameters the same (e.g., *p* = 0.1). Thus the CE function value is

$$k(y) = 1.8\frac{1600 - 40\sqrt{1600 - y}}{y} - 0.1. \qquad \text{(A2.1)}$$

Values for this function are shown graphically in Figure A2:

Figure A2
CE Function for Appendix A2 Numerical Example



A3: The CE Loss Distribution

An important restriction on the CE function $k(y)$ is that $\int_0^\infty k(y)p(y)dy = 1$. This constraint means that the *average* value of the CE function equals 1, and thus will *be less than* 1 for losses that are small. This result seems anomalous, since the PH will be averse to the risk of small losses as well as for larger ones (albeit less risk-averse for the small ones), since the small losses are also random. However, it makes sense when we consider the entire loss distribution: since losses are mutually exclusive, two different loss values are negatively correlated. The negative co-variation will reduce the CE function value if the loss amounts are simultaneously considered, as compared to a situation where some loss sizes are considered selectively.

To illustrate the effect of negative covariance, consider a PH facing a loss of $y - \varepsilon$ with probability ½ or another loss of $y + \varepsilon$ also with probability ½. The amount ε is very small. One of the two amounts will occur, but not both. If the PH insures against the first event and not the second, the CE expected loss will be greater than its expected value. This is because there is apparent risk: the loss will be zero or $y - \varepsilon$ with equal probability. The same is true if the second event is insured but not the first. However, if both events are insured (i.e., the entire loss distribution), then the CE expected loss will equal the expected value *y*, since essentially, the entire distribution is a single point *y* and there is no variance. The value *y* (plus or minus ε) is certain to occur. Thus the CE expectation over the entire distribution will be less than the sum of the CE values of the individual loss sizes taken in isolation.

The effect of the negative loss co-variation is that the *k*(*y*) values (such as in Appendix A2) will be reduced somewhat when other loss values are considered simultaneously. Extending the Appendix A2 example, suppose that (in addition to a loss amount of 1200 with a 10% probability) another loss of 1500 can occur, also with a 10% probability. Either may occur, but not both. Thus, there is an 80% chance that no loss will occur.

The utility if the 1200 loss happens is $20 = \sqrt{1600 - 1200}$, and the utility if the 1500 loss happens is $10 = \sqrt{1600 - 1500}$. So, the expected utility is 35 = 0.8(40) + 0.1(20) + 0.1(10). This gives a CE wealth of 1225 = 35^2 and the CE of the expected losses is 375 = 1600 – 1225. Thus the *joint* CE of the two loss amounts is 3750 = 375/0.1. However, taken separately, the CE of the 1200 loss is 1560 and the CE of the 1500 amount is 2310 (determined as in the Appendix A1 example). This gives a total CE for the separate losses of 3870, which is 130 more than their joint CE.

In this example, adding more possible loss values to fill out the entire probability distribution will reduce the CE values for all of the loss amounts even further.

To summarize, the particular value *k*(*y*) of the CE factor for a loss size *y* depends not only on *y* but also on *all other loss values* in the loss distribution, and their respective probabilities.

A4: Finding CE Default and Ruin Probability from a Utility Function

If we know the probability distribution of losses and the utility function, the certainty equivalent loss can be determined by inverting the expected utility, as shown in section A1. The expected utility for losses limited to an amount of assets *A* is

$$EUL(A) = \int_0^A U(y)p(y)\,dy + U(A)Q(A)\,, \qquad \text{(A4.1)}$$

where $U(y)$ is the utility if loss size y occurs and $Q(A)$ is the ruin probability, or chance that the loss exceeds assets. The expected utility for the entire loss distribution equals equation A4.1 with A set to infinity. The CE value of the limited loss is determined from the inverse utility function: $\hat{L}(A) = U^{-1}[EUL(A)]$. The CE of default is the difference between the CE of the entire loss and the loss limited to assets, just as the expected default is the difference between the expected loss and the limited expected value. Thus $\hat{D} = \hat{L} - \hat{L}(A)$. Since (from appendix C) the CE ruin probability $\hat{Q}(A)$ equals $-\partial\hat{D} / \partial A$, we have $\hat{Q}(A) = \partial[\hat{L}(A)] / \partial A$. Note that the derivative of $\hat{L}$ is zero, since it is not a function of A. Equation A4.1 thus provides a method for determining $\hat{D}$ and $\hat{Q}(A)$ given $U(y)$.

This method for getting the CE values for default and ruin probability can be illustrated using a general loss distribution with *exponential utility*. Here, I define the utility of wealth for loss size y as $U(y) = -e^{ay}$, with a being the risk aversion parameter. The expected utility of the limited loss is

$$EUL(A) = \int_0^A -e^{ay} p(y)\, dy - e^{aA} Q(A). \tag{A4.2}$$

To find $\hat{Q}(A)$, we first take the derivative of equation A4.1 with respect to A, getting

$$\begin{aligned} \frac{\partial EU[L(A)]}{\partial A} &= [U(A)p(A) - U(0)p(0)] + U(A)[-p(A)] + Q(A)\frac{\partial U(A)}{\partial A} \\ &= Q(A)\frac{\partial U(A)}{\partial A}. \end{aligned} \tag{A4.3}$$

The terms involving $p(A)$ and $p(0)$ vanish since $p(0) = 0$ and the derivative of $Q(A)$ is $-p(A)$. The inverse of the exponential utility function for a utility value X is $U^{-1}(X) = \ln(-X) / a$. Next, we take the derivative of $\hat{L}(A)$. Note that since $U(A) = -e^{aA}$, its derivative equals $aU(A)$:

$$\hat{Q}(A)=\frac{\partial\hat{L}(A)}{\partial A}=\frac{\partial}{\partial A}\frac{\ln[-EUL(A)]}{a}=\frac{-1}{aEUL(A)}\frac{\partial[-EUL(A)]}{\partial A}$$
$$=\frac{Q(A)}{aEUL(A)}\frac{\partial U(A)}{\partial A}=\frac{Q(A)U(A)}{EUL(A)} \tag{A4.4}$$

or $\hat{Q}(A)\,/\,Q(A)=U(A)\,/\,EUL(A)$. With a normal loss distribution, we can directly determine *EUL*(*A*). The normal density is

$$p(y)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-(y-L)^2}{2\sigma^2}\right]. \tag{A4.5}$$

Since the density in equation A4.1 is multiplied by –exp(*ay*), the product *U*(*y*)*p*(*y*) for the normal *p*(*y*) becomes another normal density $p_s(y)$ with a variate (shifted from the mean *L*) of $z_s=(y-L-a\sigma)\,/\,\sigma$, multiplied by the constant $-\exp[aL+\frac{1}{2}a^2\sigma^2]$. Thus, the expected utility equals this constant and its CE is the inverse, equal to $L+\frac{1}{2}a\sigma^2=\hat{L}$. This derives the result in equation 3.43. For the normal distribution, equation A4.1 becomes

$$EUL(A)=-e^{a\hat{L}}\int_0^A p_s(y)\,dy-e^{aA}Q(A)=-e^{a\hat{L}}P_s(A)-e^{aA}Q(A). \tag{A4.6}$$

Here P_s is the cumulative normal probability with the shifted variate z_s. Converting EUL(*A*) to a certainty equivalent, we get

$$\hat{L}(A)=\frac{\ln[e^{a\hat{L}}P_s(A)+e^{aA}Q(A)}{a}. \tag{A4.7}$$

Since $\hat{D}=\hat{L}-\hat{L}(A)$, we finally have

$$\hat{D}=\frac{-\ln[P_s(A)+e^{a(A-\hat{L})}Q(A)}{a}. \tag{A4.8}$$

To determine the adjusted ruin probability $\hat{Q}(A)$ for the normal distribution, we use equations

A4.4 and A4.6, getting:

$$\hat{Q}(A) = \frac{Q(A)}{Q(A) + e^{-a(A-\hat{L})}P_s(A)} . \qquad \text{(A4.9)}$$

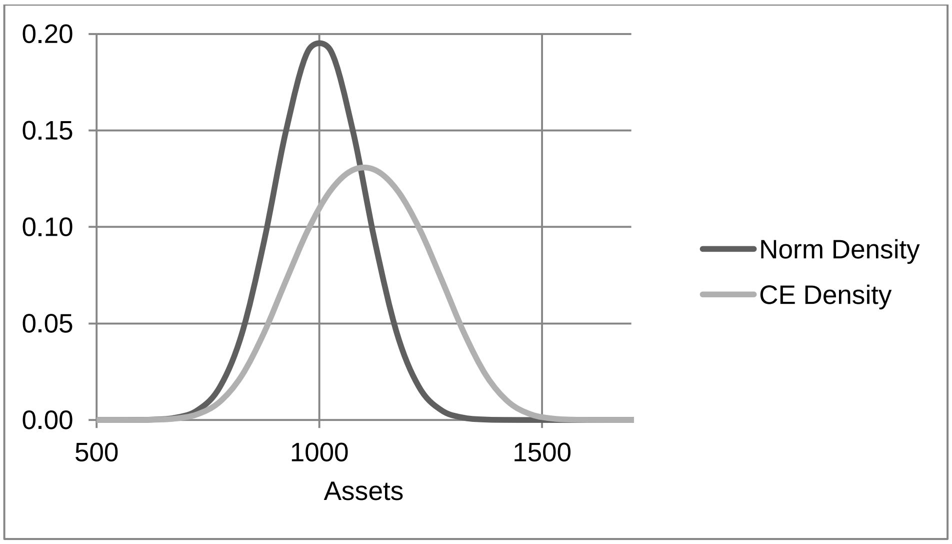
Notice that the distribution for $P_s(A)$ has a mean of $L + a\sigma^2$, while for $Q(A)$ the mean is L. The mean underlying the $\hat{Q}(A)$ distribution is $\hat{L} = L + \frac{1}{2}a\sigma^2$. Letting $A = \hat{L} + x$, we see that $P_s(\hat{L} + x) = Q(\hat{L} - x)$. From equation A4.9 it is straightforward to show that $\hat{Q}(\hat{L} + x) + \hat{Q}(\hat{L} - x) = 1$ and therefore the CE distribution is symmetric around $\hat{L}$. However, as shown below, the CE distribution is not normal.

To show the non-normality, I use the section 3.5 example with L = 1000, σ = 100 and a = 0.02. With assets of 1300 (3 standard deviations above the mean) the unadjusted expected default D is 0.038 and the ruin probability is Q(1300) = 0.135%. The CE expected loss is 1100 = 1000 + 0.5(0.02)(100)^2. The shifted variate is z_s = 1.00 = [1300 – 100 – (0.02)(100)]/100, which is one standard deviation above the mean. Thus, the shifted cumulative probability is P_s(1300) = 0.8413. The factor $\exp[a(A-\hat{L})]$ equals exp[0.02(1300 – 1100)] = 54.598, so we get $\hat{D}$ = 4.439 = {–ln[0.8413 + (54.598)(0.00135)]}/0.02. The CE ruin probability is $\hat{Q}(1300)$ = 8.05% = 0.00135/[0.00135 + 0.8413/54.598]. Notice that the denominator of equation A4.9 contains an exponential factor that is the reciprocal of the one in equation A4.8.

If the CE distribution were normal, then its ruin probability of 8.05% would imply a standard deviation of 142.71. Then, if we change the assets to 1200, we would get a CE ruin probability of 24.17%. However, following the above calculation, the true $\hat{Q}(1200)$ is 25.16%. Consequently, the CE distribution for the normal-exponential model is not normal.

It is interesting to compare the CE density $\hat{p}(y)$ with that of the underlying normal distribution $p(y)$. The approximate $\hat{p}(y)$ values can be calculated taking the difference of successive $\hat{Q}(A)$ values. Figure A4 below shows the two densities.

Figure A4
Probability Densities for Normal-Exponential Model
Illustrative Example



Notice that the CE density is symmetric, centered at the CE loss value of 1100. It also has a greater variance than its normal parent distribution. Also observe that, for a given asset amount (above the mean), the tail area of the adjusted distribution is much greater than that of the unadjusted distribution.

A5: CE Values for Binary Loss Model with Exponential Utility

Assume that an individual faces a loss of amount B with probability $p > 0$, and amount 0 with probability $q = 1 - p$. This is called a binary model, since there are two possible loss values: B or zero. The individual has risk preferences defined by exponential utility with risk aversion parameter a and has initial wealth of W_0 before considering the loss prospect. The utility of the initial wealth is $-e^{-aW_0}$ and the expected utility of wealth considering the loss is

$$EU(0) = -qe^{-a(W_0 - 0)} - pe^{-a(W_0 - B)} = -e^{-aW_0}[q + pe^{aB}]. \qquad \text{(A5.1)}$$

Letting $\hat{W}_0$ denote the CE value of the wealth considering the loss, we have $\hat{W}_0 = W_0 - \hat{L}$, where $\hat{L}$ is the CE expected loss. Since $-e^{-a\hat{W}_0} = -e^{-aW_0}[q + pe^{aB}]$, we get

$$\hat{L} = \ln[q + pe^{aB}] / a. \qquad \text{(A5.2)}$$

If the individual buys insurance for a premium π and the insurer has assets A, then the amount of loss absorbed by the individual is $B - A$ and the expected utility is $EU(A) = -qe^{-a(W_0-\pi)} - pe^{-a(W_0-B-A-\pi)}$. The certainty equivalent wealth after the insurance purchase is the initial wealth minus the premium minus the CE expected default, or $W_0 - \pi - \hat{D}$. Since $-e^{-a(W_0-\pi-\hat{D})} = EU(A)$, we solve for $\hat{D}$:

$$\hat{D} = \ln[q + pe^{a(B-A)}] / a. \qquad \text{(A5.3)}$$

Notice that if $A = 0$, then $\hat{D} = \hat{L}$ and if $A = B$, then $\hat{D} = 0$. The CE ruin probability $\hat{Q}(A)$ equals the negative derivative of the CED, so we get

$$\hat{Q}(A) = \frac{pe^{a(B-A)}}{q + pe^{a(B-A)}}. \qquad \text{(A5.4)}$$

If $A = 0$, we have $\hat{Q}(0) = pe^{aB} / (q + pe^{aB})$ and at $A = B$, $\hat{Q}(B) = p$.

For *two combined* independent binary risks, the development is similar (the values of variables for the combined risks are denoted with a subscript 2). Here we set the initial wealth and premium to zero, since they do not influence the CED and hence the CE ruin probability. Following section 3.6, assets are A per PH, for a total of $2A$. For $A < B/2$, the expected utility per PH is $EU(A) = -q - pqe^{a(B-2A)} - p^2e^{a(B-A)}$. For $B/2 \leq A < B$, $EU(A) = -q - pq - p^2e^{a(B-A)}$. Then the respective CED values are

$$\hat{D}_2 = \ln(Z_1) / a \qquad \text{for } A < B/2 \qquad \text{(A5.5)}$$
$$\hat{D}_2 = \ln(Z_2) / a \qquad \text{for } B/2 \leq A < B,$$

where $Z_1 = q + pqe^{a(B-2A)} + p^2e^{a(B-A)}$ and $Z_2 = q + pq + p^2e^{a(B-A)}$. Here, if $A = 0$, we again get $\hat{D}_2 = \hat{L}$ and if $A = B$, then $\hat{D}_2 = 0$. Taking the derivative of the CED with respect to A, we get the CE ruin probabilities:

$$\hat{Q}_2(A) = [2pqe^{a(B-2A)} + p^2 e^{a(B-A)}] \,/\, Z_1 \qquad \text{for } A < B/2 \tag{A5.6}$$

$$\hat{Q}_2(A) = p^2 e^{a(B-A)} \,/\, Z_2 \qquad \text{for } B/2 \le A < B.$$

For $A = 0$, we get $\hat{Q}_2(0) = (1-q^2)e^{aB} \,/\, (q + pe^{aB})$ and for $A = B$, $\hat{Q}_2(B) = p^2$. Notice that $\hat{Q}_2(0) \,/\, \hat{Q}(0) = 2 - p > 1$ and that $\hat{Q}_2(B) \,/\, \hat{Q}(B) = p < 1$. Thus, the CE ruin probability for the combined risks has a higher maximum value (at $A = 0$) and a lower minimum value (at $A = B$) than for a single risk. Thus, based on equation 5.51, there is a region where a subadditivity (SA) violation may exist and another region where a SA violation cannot happen.

For a fair premium, equation 5.14 defines the risk measure as $\theta(A) = [\hat{Q}(A) - Q(A)] \,/\, [1 - Q(A)]$. For a single binary risk, the unadjusted ruin probability is $Q(A) = p$, for $A \le B$. Consequently,

$$\theta(A) = [\hat{Q}(A) - p] \,/\, q, \tag{A5.7}$$

so the fair premium risk measure $\theta(A)$ is a linear function of the basic premium risk measure $\hat{Q}(A)$.

For a combination of *two independent binary risks*, and for $A < B/2$, the expected default for a single PH is $D_2 = pq(B - 2A) + p^2(B - A)$. By taking the negative of the derivative of D_2 with respect to A we get $Q_2(A) = 1 - q^2$. For $B/2 \le A < B$, we have $Q_2(A) = p^2$. Therefore, equation 5.14 gives

$$\theta_2(A) = [\hat{Q}_2(A) + q^2 - 1] \,/\, q^2 \qquad \text{for } A < B/2 \tag{A5.8}$$
$$\theta_2(A) = [\hat{Q}_2(A) - p^2] \,/\, [1 - p^2] \qquad \text{for } B/2 \le A < B.$$

For $A = 0$, with some manipulation, we get $\theta_2(0) \,/\, \theta(0) = (2 - p) \,/\, q > 1$ and for $A = B$ (as a limit), $\theta_2(B) \,/\, \theta(B) = p < 1$. Just as in the basic premium case, there is a region where a subadditivity violation may exist and another region where it cannot occur.

APPENDIX B: STOCHASTIC MEAN LOSS MODEL

The classic aggregate loss model from risk theory (see Lundberg [1903]) is the compound

Poisson process, where the number of losses is Poisson and each individual loss has the same distribution with prescribed parameters and thus a stable mean. The individual losses are independent. However, in practice, the losses are not independent (they are subject to common factors such as inflation, regulation and the court system). Further, the mean of losses for a given line of business (or other subdivision of an insurer's risk portfolio) is not known; it must be determined empirically.

Relaxing the independence assumption, we introduce another random variable that governs the mean from which all the individual losses are drawn (in this section, random variables are indicated with a tilde).

Let $\tilde{I}$ represent the stochastic mean variable, which itself has a mean of 1 and a variance γ^2. We assume that there are *N* policyholders. Each PH *i* has losses (we allow for multiple claims in the one-period model) denoted by $\tilde{x}_i$. The $\tilde{x}_i$ are measured before applying the mean variable $\tilde{I}$.

Let $\tilde{S}$ denote the sum of aggregate losses before applying the stochastic mean. The unconditional aggregate losses are

$$\tilde{X} = \tilde{I} \cdot \tilde{S} = \tilde{I}[\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_N]. \tag{B.1}$$

Here the $\{\tilde{x}_i\}$ are the individual losses and N is the number of losses. $\tilde{I}$ and $\tilde{S}$ are independent, so the covariance between losses is due to the stochastic mean. Let *M* be the mean of the individual loss $\tilde{x}_i$ and σ^2 its variance. Let ρ be the correlation between the individual losses. The variance of $\tilde{S}$ is

$$\mathrm{Var}(\tilde{S}) = \sigma^2[N + \rho(N^2 - N)] \tag{B.2}$$

and its mean is *NM.*

The variance of the product of two independent random variables $\tilde{I}$ and $\tilde{S}$ is

$$\mathrm{Var}(\tilde{I} \cdot \tilde{S}) = \mathrm{Var}(\tilde{I}) \cdot \mathrm{Var}(\tilde{S}) + \mathrm{E}(\tilde{I})^2 \cdot \mathrm{Var}(\tilde{S}) + \mathrm{E}(\tilde{S})^2 \cdot \mathrm{Var}(\tilde{I}). \tag{B.3}$$

Here, E( ) denotes the expectation. Thus the variance of the aggregate losses $\tilde{X}$ is

$$\mathrm{Var}(\tilde{X}) = [N(1-\rho) + \rho N^2]\,\sigma^2[1+\gamma^2] + \gamma^2 M^2 N^2. \tag{B.4}$$

Let $\tilde{Y} = \tilde{X} / N$ denote the share of aggregate losses for an individual PH. Then $\mathrm{Var}(\tilde{Y}) = \mathrm{Var}(\tilde{X}) / N^2$. Thus, we have

$$\mathrm{Var}(\tilde{Y}) = \gamma^2 M^2 + \rho\sigma^2(1+\gamma^2) + (1-\rho)\sigma^2(1+\gamma^2) / N. \tag{B.5}$$

As *N* becomes large, the variance of the individual PH losses tends toward

$$\mathrm{Var}(\tilde{Y}) = \gamma^2 M^2 + \rho\sigma^2(1+\gamma^2). \tag{B.6}$$

So for large *N*, the variance of the individual PH losses tends to a constant limiting value. Also, if $\gamma = 0$ and $\rho > 0$, the limit is $\rho\sigma^2$. If $\gamma > 0$ and $\rho = 0$, then the limit is $\gamma^2 M^2$. Consequently, if either the losses are subject to a stochastic mean or they are correlated, then with a large number of policyholders, the variance of individual PH losses will reach a limit.

An example will illustrate the convergence. Assume that *M* = 1,000, $\sigma = 300$, $\gamma = 0.1$ and $\rho = 0.2$. For *N* = 1,000, equation B.5 gives an individual PH variance of 28,253, compared to the asymptotic value of 28,180, a difference of only 0.26%. Increasing *N* to 10,000 policies moves the variance to 28,187, cutting the difference to 0.03% from the asymptotic value. Notice that the asymptotic standard deviation is 167.87, which compares to 300.00 from the individual loss distribution.

APPENDIX C: DERIVATIVE OF THE EXPECTED DEFAULT

To determine the derivative of the expected default, we use the general method for the derivative of an integral, with the upper limit a constant *b* and the lower limit a function of the variable whose derivative is taken:

$$\frac{\partial}{\partial y}\int_{g(y)}^{b} F(x,y)dx = \int_{g(y)}^{b} \frac{\partial}{\partial y} F(x,y)dx - F[g(y),y]\frac{d[g(y)]}{dy}. \tag{C.1}$$

Thus,

$$\frac{\partial D}{\partial A} = \int_A^\infty \frac{\partial}{\partial A}[x - A]p(x)\,dx - (A - A)p(A)\frac{\partial A}{\partial A}. \quad \text{(C.2)}$$

The right-hand term in equation C.2 equals zero and the derivative becomes

$$\frac{\partial D}{\partial A} = \int_A^\infty -p(x)\,dx = -Q(A), \quad \text{(C.3)}$$

where $Q(A) = \int_A^\infty p(x)\,dx$ is the tail, or ruin, probability.

For an adjusted probability density $\hat{p}(x)$, we have in parallel fashion,

$$\frac{\partial \hat{D}}{\partial A} = \int_A^\infty -\hat{p}(x)\,dx = -\hat{Q}(A). \quad \text{(C.4)}$$

APPENDIX D: OPTIMAL CAPITAL WITH INCOME TAXES

Expanding the basic model to include income taxes, we also need to introduce an investment component. Assume that all cash is invested in riskless investments at a one-period rate *r*, and that all income is taxed at the end of the period at a rate *t*. Premium is collected at the beginning of the period and losses are paid at the end of the period. We further assume that the losses contain no market risk, so that the expected return to investors[46] in the insurer is also *r*. Alternatively, we can assume that the value of the loss is adjusted to include the market risk.

In order to attract capital from the insurer's investors, the expected rate of return *after taxes* must also equal *r*.

If the premium equals the present value of the expected loss, then initial assets are $A_0 = C + L/(1+r)$. The expected value of the ending assets, prior to income tax, is $A_0(1+r) - L = C + rC$. The investment income *rC* is taxed, leaving *C* + *rC* – *trC*. However, for a fair return to investors, the ending assets must be *C* + *rC*. The amount *trC* must be made up

[46] In this formulation, the expected default is not subtracted from premium, so the result approximates a true equilibrium optimum, which is a more complex version of equation 5.14.

by charging an extra premium amount *zC* at some rate *z* proportional to capital, so the premium is $\pi = L/(1+r) + zC$. The extra premium is itself taxed as underwriting profit, so the amount *zC* will grow to $zC(1+r) - tzC - trzC$ after taxes. Notice that the investment income *rzC* on *zC* is also taxed. Equating the ending after-tax value of the additional premium with the double-taxation burden *trC*, we solve for *z*:

$$z = \frac{tr}{(1+r)(1-t)}\,. \tag{D.1}$$

Let *A* represent the amount of assets prior to payment of the loss and income taxes. If the loss is larger than *A*, the insurer will default and no tax is paid. I assume here that a negative income tax liability arising from a large loss does *not* increase the assets available to pay the loss. Thus we have

$$A = [C + \frac{L}{1+r} + zC](1+r) = L + C(1+z)(1+r). \tag{D.2}$$

In parallel fashion to equation 5.11, the consumer value *V* of the insurance equals the present value of the CE of the covered losses, minus the premium:

$$V = \frac{\hat{L}}{1+r} - \frac{\hat{D}}{1+r} - \frac{L}{1+r} - zC. \tag{D.3}$$

Taking derivatives and equating to zero, we have

$$z\frac{\partial C}{\partial A} = \frac{1}{1+r}\frac{-\partial \hat{D}}{\partial A} = \frac{1}{1+r}\hat{Q}(A). \tag{D.4}$$

From equation D.2, we get $\partial C / \partial A = 1/[(1+z)(1+r)]$. From equation D.4 we get

$$\hat{Q}(A) = \frac{z}{1+z}. \tag{D.5}$$

Using the value of z in equation D.1, we get the optimal CE ruin probability in terms of the interest rate and the income tax rate:

$$\hat{Q}(A) = \frac{rt}{1+r-t} \,. \tag{D.6}$$

ACKNOWLEDGMENTS

The ideas behind this paper arose as a result of my work on the American Academy of Actuaries Property-Casualty Risk-Based Capital Committee, headed by Alex Krutov. After the 2008 financial crisis, I began to model extreme events for the property-casualty industry and developed the policyholder welfare approach to optimal capital. I am grateful to Alex for his encouragement as I attempted to advance these concepts.

In 2011, the American Academy of Actuaries committee sought help from the Casualty Actuarial Society (CAS) in preparing risk-based capital proposals for the National Association of Insurance Commissioners. I joined the CAS RBC Dependency and Correlation Working Party, led by Allan Kaufman. As my contribution to this effort, I began a project to determine the best solvency risk measure for property-casualty insurers. The assignment greatly expanded my earlier work, and this paper is the result. In fact, it serves as the report for my subcommittee on the working party. I am deeply thankful for Allan's stewardship in guiding me along and keeping a clear focus throughout the project. His innumerable astute comments, sharp critique and editorial suggestions were invaluable; they forced me to explain results more clearly — the paper is much better for his involvement.

I also thank Glenn Myers, a fellow member of the working party, for his helpful comments and insight. In particular, he suggested a comparison of utility theory and the certainty equivalent formulation. Additionally, I thank Alex Krutov and Tom Struppeck for spurring the discussion of subadditivity, a topic I had not previously considered.

REFERENCES

[1] Acerbi, C. (2002) "Spectral measures of risk: A coherent representation of subjective risk aversion," Journal of Banking and Finance, 26 (7), 1487–1503.
[2] Albanese, C. (1997), Credit Exposure, Diversification Risk and Coherent VaR, Working Paper, Department of Mathematics, University of Toronto, September.
[3] Arrow, K. J. and G. Debreu. 1954. Existence of an equilibrium for a competitive economy. Econometrica 22:265–290.
[4] Artzner, P. (1999). "Application of coherent capital requirements", North American Actuarial Journal, vol. 3(2), 11–25.
[5] Birge, John R. and François V. Louveaux (1997). Introduction to Stochastic Programming. Springer Verlag, New York.
[6] Butsic, Robert P., Solvency Measurement for Property-Liability Risk-Based Capital Applications, Journal of Risk and Insurance, 1994 Vol. 61, (4), 656-690.
[7] Butsic, Robert P., CAS Forum 1999, Capital Allocation for Property-Liability Insurers: A Catastrophe Reinsurance Application, 1-70.
[8] Cummins, J. David, 1988, Risk-Based Premiums for Insurance Guaranty Funds, Journal of Finance, 43: 823-839.
[9] Einav, Liran, Amy Finkelstein, and Mark R. Culleny. 2010. Estimating welfare in insurance markets using variation in prices. Q. J. Econ. 125**.**
[10] Einhorn, David (June/July 2008), *Private Profits and Socialized Risk*, GARP Risk Review.
[11] Hancock, John, Paul Huber, and Pablo Koch (2001), "The Economics of Insurance: How Insurers Create Value for Shareholders," Zurich: Swiss Re Technical Publishing.
[12] Harrington, S., "Insurance Derivatives, Tax Policy, and the Future of the Insurance Industry," *Journal of Risk and Insurance*, 1997, 64:719-725.
[13] Harrington, Scott E., Edited version published in H. Scott, ed., Capital Adequacy Beyond Basel: Banking, Securities, and Insurance, Oxford University Press, 2005 Capital Adequacy in Insurance and Reinsurance.
[14] Holton, Glyn (2003). *Value-at-Risk: Theory and Practice*. Academic Press.
[15] Hull, John C. (2008). Options, Futures and Other Derivatives (7 ed.). Prentice Hall. ISBN 0-13-505283-1
[16] Jorion, Philippe (2006). Value at Risk*: The New Benchmark for Managing Financial Risk* (3rd ed.). McGraw-Hill. ISBN 978-0-07-146495-6.
[17] Lowe, Stephen, François Morin, and Dean Swallow, *Risk Horizon and the Measurement of Economic Capital for General Insurers*, Towers Watson, 2011.
[18] Lundberg F. (1903) Approximerad Framställning av Sannolikehetsfunktionen, Återförsäkering av Kollektivrisker, Almqvist & Wiksell, Uppsala.
[19] Marshall, Alfred (1947 [1890]): *Principles of Economics*. London, Macmillan.
[20] Meyers, G.G., Schenker, N., "Parameter Uncertainty in the Collective Risk Model", Proceedings of the Casualty Actuarial Society, LXX, 1983, 111-143.
[21] Myers, S. C., and J. A. Read, Jr. 2001. Capital Allocation for Insurance Companies. *Journal of Risk and Insurance* 68(4): 545–80.
[22] Nocero, Joe (January 4, 2009), *Risk Mismanagement*, The New York Times Magazine.
[23] Panjer, Harry H., ed., (1998) Financial Economics With Applications to Investments, Insurance, and Pensions, Society of Actuaries.
[24] Pratt, J. W., "Risk aversion in the small and in the large," Econometrica 32, January–April 1964, 122–136.
[25] Rothschild, M. and J. Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics, pages 629–649, 1976.
[26] Varian, Hal R., Microeconomic Analysis, Third edition. Norton, New York 1992.
[27] Wang, S. S. 1996. Premium Calculation by Transforming the Layer Premium Density. *ASTIN Bulletin* 26(1): 71–92.
[28] Yaari, M.E. (1987). "The dual theory of choice under risk", Econometrica 55, 95–115.

GLOSSARY OF ABBREVIATIONS AND NOTATIONS

Abbreviation	Meaning	Section Where Defined
ARP	Adjusted ruin probability	5.1
CE	Certainty equivalent	3.2
CEL	Certainty equivalent expected loss	3.2
CED	Certainty equivalent expected default	3.5
CV	Consumer value	3.6
DCP	Default Correlation Parameter	6.2
EPD	Expected policyholder deficit	1
FCC	Frictional capital cost	4.2
GF	Guaranty fund	6.2
NAIC	National Association of Insurance Commissioners	7.2
PV	Present value	4.3
PH	Policyholder	3.1
RBC	Risk-based capital	2
RP	Ruin probability	5.4
SA	Subadditivity	5.5
SD	Standard deviation	5.3
TVaR	Tail value-at-risk	1
VaR	Value-at-risk	1

Notation	Meaning	Section Where Defined
a	Exponential utility risk aversion parameter	3.4
A	Assets	3.5
A_0	Initial assets	App. D
b	Upper integration limit	App. C
B	Binary loss size	5.5
C	Capital	4.1
CEW	Certainty-equivalent wealth	App. A1
D	Expected default	3.5
$\hat{D}$	Certainty-equivalent expected default	3.5
e_0 , e_1	Expense coefficients	5.1
$E(\ )$	Expectation operator	App. B
EU	Expected utility	App. A1
$EUL(\ )$	Limited expected utility	App. A4
$f(\ )$	General function	3.6
$F(\ )$	General function	App. C
$g(\ )$	General function	App. C
$\tilde{I}$	Stochastic mean variable	App. B
$k(\ )$	Certainty-equivalent function	3.2
k	Average value of the CE function	3.2
K	Variable in normal-exponential model	App. A4
L	Expected loss	3.2
$\hat{L}$	Certainty-equivalent loss	3.2
$\hat{L}(\ )$	Limited CE expected loss	3.5
M	Mean of individual loss	App. B
$MT(\ )$	Tail moment	5.4
n	Degree of the tail moment	5.4
N	Number of policies	App. B
p	Binary loss probability	App. A1
$p(\ )$	Probability density	3.2
$\hat{p}(\ )$	Probability density, adjusted for risk aversion	3.2
$p_A(\ )$	Risk-neutral probability density for asset size	6.1
$p_s(\ )$	Probability density with shifted mean	App. A4
$P_s(\ )$	Cumulative probability with shifted mean	App. A4
q	Probability of zero loss	App. A5
$Q(\ )$	Ruin probability	5.1
$\hat{Q}(\ )$	Ruin probability, adjusted for risk aversion	5.1
r	Investment return	4.2
$R_A(\ )$	Absolute risk aversion function	3.4
$RM_1(\ )$, $RM_2(\ )$	Risk Measure	5.5
$\tilde{S}$	Sum of aggregate losses, without stochastic mean	App. B

Notation	Meaning	Section Where Defined
t	Income tax rate	4.2
$u(\ )$	Utility function, based on wealth	App. A1
$U(\ )$	Utility function, based on loss size	3.4
v	Valuation level of risk measure	5.5
V	Consumer value of insurance contract	5.1
Var( )	Variance operator	App. B
W	Wealth	App. A1
W_0	Initial wealth	3.4
$\hat{W}_0$	CE wealth given a potential loss	App. A5
x	Loss size; also a general variable	3.4
$\tilde{x}_i$	Random individual loss size	App. B
$\tilde{X}$	Unconditional aggregate losses	App. B
y	Individual policy loss size	3.2
$\hat{y}$	Certainty equivalent loss size	3.2
$\tilde{Y}$	Policyholder share of aggregate losses	App. B
z	Frictional cost of capital rate	4.2
z_s	Shifted normal variate	App. A4
Z_1, Z_2	Intermediate variables	App. A5
α	VaR confidence level	5.4
γ^2	Variance of stochastic mean	App. B
Δ	Change in loss size	5.4
ε	Small change in loss size	App. A3
π	Premium	4.1
$\theta(\ )$	Fair premium risk measure	5.1
ρ	Correlation between losses	App. B
σ^2	Variance of loss	3.4

BIOGRAPHY OF THE AUTHOR

Robert P. Butsic is a retired actuary currently residing in San Francisco. He is a member of the American Academy of Actuaries Property-Casualty Risk Based Capital Committee and the Casualty Actuarial Society's Risk-Based Capital Dependency and Calibration Working Group. He previously worked for Fireman's Fund Insurance and CNA Insurance. He is an Associate in the Society of Actuaries, has a B.A. in mathematics and an MBA in finance, both from the University of Chicago. He has won the Casualty Actuarial Society's Michelbacher Award (for best Discussion Paper) five times. In the early 1990s he was a member of the American Academy of Actuaries working group that advised the NAIC in developing the current property-casualty RBC methodology. Since the 2008 financial crisis he has enjoyed reading economics blogs, which have stimulated the development of this paper.