AN ACTUARIAL INDEX OF THE RIGHT-TAIL RISK*

Shaun Wang†

ABSTRACT

A common characteristic for many insurance risks is the right-tail risk, representing low-frequency, large-loss events. In this paper I propose a measure of the right-tail risk by defining the right-tail deviation and the right-tail index. I explain how the right-tail deviation measures the right-tail risk and compare it to traditional measures such as standard deviation, the Gini mean, and the expected policyholder deficit. The right-tail index is applied to some common parametric families of loss distributions.

1. INTRODUCTION

In a broad sense, an insurance risk refers to the business, legal, or management aspects of transferring the economic impact of unforeseen mishaps. In this paper, the term "insurance risk" refers to a loss variable that quantifies the potential loss amount associated with an insurance contract or the whole book of an insurer's business, depending upon the intended application. With this narrow definition, a characteristic of many insurance risks (individual or aggregate) is the right-tail risk, which represents low-frequency and large-loss events. This characteristic can be observed from the following two aspects:

1. *Process Deviation*. In liability insurance, insurance loss amounts or loss developments often are highly skewed and have long right tails. In property insurance, the hazards of natural disasters (earthquake, hurricane, flood) often manifest themselves at the right tail in a property writer's aggregate claims distribution. In such situations, the large deviations due to the right-tail losses are a major concern to the insurer. Thus an indicator of the right-tail random deviation from the expected loss is desirable.
2. *Parameter Risk*. In practice, the probability distributions for losses or loss developments are seldom known with precision. There is always considerable uncertainty about the best-estimate probability distribution. In terms of statistical sampling error, the further at the right tail, the fewer are the available data and thus the higher the uncertainty regarding the best-estimate tail probabilities. For instance, in liability insurance, considerably greater uncertainty exists in increased-limits ratemaking than in the basic-limit ratemaking.

Traditionally, the most commonly used measures of risks are variance and standard deviation. Standard deviation is a "standard" measure of deviation from the mean if the underlying variable has a normal distribution. Even though standard deviation has been used to measure the deviation from the mean for other than normal distributions, it is not a good risk measure for large insurance risks with skewed distributions. The poor performance of standard deviation in measuring insurance risks has been reported by many authors, for example, Ramsay (1993) and Lowe and Stanard (1996).

In clarification of the concept of risk margins, Philbrick (1994) discusses various types of risk measures as below. Here I elaborate Philbrick's points with special emphasis on measures of the right-tail risk.

- "A risk margin based on a certainty equivalent concept." The certainty equivalent is a well-established concept in actuarial and economic theories of risk and uncertainty (for example, the expected utility theory and premium calculation principles). The certainty equivalent may represent the market price for transferring the risk. Ideally, the certainty equivalent for a risk automatically adjusts for the right-tail risk. Therefore, once a certainty equivalent theory is established, a measure of the right-tail deviation can be derived from the difference between the certainty equivalent and the expected loss.
- "A risk margin based on probability confidence intervals." This is a natural approach as long as the

*This research was supported by a grant from the Society of Actuaries Committee on Knowledge Extension Research.

†Shaun Wang, A.S.A., Ph.D., is an Associate Actuary at the SCOR Reinsurance Company, One Pierce Place, Itasca, Illinois 60143-4049, e-mail scorus/itasca/swang%5512559@mcimail.com.

parameter risk is of main concern. A measure of the right-tail risk should be capable of quantifying the parameter uncertainty.
- "A risk margin based on a theory of ruin." The probability of ruin is essentially a quantile concept: (1) if a tolerable (small) probability of ruin is stated, it gives the amount of capital needed to ensure this safety level; (2) if a threshold amount of loss is stated, it gives the probability that the actual loss will exceed this threshold. Butsic (1994) proposes the use of the expected policyholder deficits (EPD) in determining risk-based-capital requirements. Gerber and Shiu (1997) discuss the probability of ruin as well as the deficit at ruin. The EPD is an advancement of the probability of ruin concept because it takes into account not only the probability of ruin but also the magnitude of deficit and thus reveals more information about the right tail than simply the probability of ruin. A good measure of the right-tail risk should utilize/reveal as much information as possible about the tail probabilities.

In this paper I propose a new measure of the *right-tail deviation* for a non-negative random variable X:

$$\mathrm{D}[X] = \int_0^\infty \sqrt{\Pr\{X > t\}}\, dt - \mathrm{E}[X].$$

I explain how $D[X]$ measures the right-tail risk from various perspectives including (1) the certainty equivalent approach, (2) the parameter risk approach, and (3) distance between loss distributions. I show that $D[X]$ preserves the basic properties of the standard deviation, namely,
- $D[aX+b]=aD[X]$, for constants $a>0$ and b.
- $D[X+Y]\leq D[X]+D[Y]$ and equality holds for perfectly correlated risks.

However, the right-tail deviation differs from the standard deviation in the following aspects:
- The right-tail deviation $D[X]$ preserves the common ordering of risks such as first and second stochastic dominance, while the standard deviation does not. Indeed, the right-tail deviation is much more capable of differentiating risks than standard deviation.
- The right-tail deviation $D[X]$ is additive when a risk is divided into excess-of-loss layers, while the standard deviation is subadditive for layers. This distinctive behavior gives $D[X]$ a comparative advantage in calculating risk charges in reinsurance pricing.

2. PRELIMINARY CONCEPTS

2.1 Expected Loss

For an insurance risk X, a non-negative random variable, its cumulative distribution function is defined by $F_X(t)=\Pr(X\leq t)$, and its decumulative distribution function (ddf) is defined by $S_X(t)=\Pr(X>t)$. Whenever possible, we use the ddf representation, which has many advantages including a unified treatment of both continuous and discrete variables.

Lemma 2.1

For a non-negative random variable X, we have

$$\mathrm{E}[X] = \int_0^\infty S_X(t)dt.$$

This result can be found in Bowers et al. (1986), but here we give a simple yet general proof.

Proof

For $x\geq 0$ it is true that

$$x = \int_0^\infty I(x > t)dt,$$

where I is the indicator function. For a non-negative random variable, it holds that

$$X = \int_0^\infty I(X > t)dt.$$

By taking expectation on both sides of the equation, we get

$$\mathrm{E}[X] = \int_0^\infty \mathrm{E}[I(X > t)]dt = \int_0^\infty S_X(t)dt. \qquad \square$$

2.2 Insurance Layers

Most insurance contracts have some policy provisions such as deductibles and limits. A large risk is often divided into layers or quota shares among several insurers. Here we introduce a general term of layers. A *layer* $(a, a+h]$ of risk X is defined as an excess-of-loss cover:

$$X_{(a,a+h]} = \begin{cases} 0, & 0 \leq X < a \\ (X - a), & a \leq X < a + h, \\ h, & a + h \leq X \end{cases}$$

where a is the deductible (or retention), and h is the limit.

The following result can be easily verified.

Lemma 2.2

The excess-of-loss cover $X_{(a,a+h]}$ has a ddf:

$$S_{X_{(a,a+h]}}(t) = \begin{cases} S_X(a+t), & t < h \\ 0, & t \geq h. \end{cases};$$

From Lemmas 2.1 and 2.2, the expected loss for the layer $(a, a+h]$ is

$$\mathrm{E}[X_{(a,a+h]}] = \int_0^\infty S_{X_{(a,a+h]}}(t)\,dt$$
$$= \int_0^h S_X(a+u)\,du = \int_a^{a+h} S_X(t)\,dt.$$

Remark

A layer $(a, a+h]$ is like a window. The payment of a layer $(a, a+h]$ is the part of insurance claims that is observed through this window. Also, the ddf for a layer $(a, a+h]$ is the part of the original ddf that falls into this window.

Remark

Note that $S_X(t)$ represents the frequency of hitting the layer $(t, t+dt]$, and $S_X(t)dt$ represents the expected loss for the infinitesimal layer $(t, t+dt]$. Furthermore, $X_{(t,t+dt]}$ has approximately a Bernoulli distribution with

$$\Pr\{X_{(t,t+dt]} = 0\} = 1 - S_X(t), \quad \Pr\{X_{(t,t+dt]} = dt\} = S_X(t).$$

2.3 Comonotonicity

Traditionally, the relationship of two random variables, X and Y, is usually measured by the Pearson correlation coefficient:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\,\sqrt{\mathrm{Var}(Y)}},$$

where $\mathrm{Cov}(X,Y)=\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]$ is the covariance of X and Y. If $\rho(X,Y)=0$, then X and Y are uncorrelated. If $\rho(X,Y)=1$, then X and Y are perfectly correlated; in this case it is necessary that $Y=aX+b$ with $a>0$.

The concept of perfect correlation is too restrictive because it holds only for random variables with linear relationships $Y=aX+b$ with $a>0$. As a generalization of perfect correlation, an important concept of *comonotonicity* introduced by Yaari (1987) and Schmeidler (1986) has played a very important role in decision theory under uncertainty.

Definition 2.1

Two risks X and Y are **comonotonic** if there exists a random variable Z and nondecreasing real functions u and v such that

$$X = u(Z), \quad Y = v(Z), \qquad \text{with probability one.}$$

Comonotonicity is a generalization of the concept of perfect correlation to random variables without linear relationships. Note that perfectly correlated risks are comonotonic, but its converse does not hold. Consider two layers $(a, a+h]$ and $(b, b+h]$ for a continuous variate X. The layer payments $X_{(a,a+h]}$ and $X_{(b,b+h]}$ are comonotonic because both are nondecreasing functions of the original risk X. They are bets on the same event, and neither of them is a *hedge* against the other. On the other hand, for $a \neq b$, $X_{(a,a+h]}$ and $X_{(b,b+h]}$ are *not* perfectly correlated because neither can be written as a linear function of the other.

Lemma 2.3

For two comonotonic risks X and Y, Cov(X,Y)≥0.

Proof

If X and Y are comonotonic, from Definition 2.1 there exists a random variable Z such that

$$X = u(Z), \quad Y = v(Z), \quad \text{with probability one,}$$

where the functions u and v are nondecreasing. Since u is nondecreasing, there exists a number t_0 such that $u(t) \geq \mathrm{E}[u(Z)]$ for $t \geq t_0$ and $u(t) \leq \mathrm{E}[u(Z)]$ for $t < t_0$. Now we have

$$\begin{aligned}
\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] &= \int_0^\infty u(t)v(t)\,dF_Z(t) \\
&\quad - \mathrm{E}[u(Z)] \int_0^\infty v(t)\,dF_Z(t) \\
&= \int_0^{t_0} \{u(t) - \mathrm{E}[u(Z)]\}\, v(t)\,dF_Z(t) \\
&\quad + \int_{t_0}^\infty \{u(t) - \mathrm{E}[u(Z)]\}\, v(t)\,dF_Z(t) \\
&\geq \int_0^{t_0} \{u(t) - \mathrm{E}[u(Z)]\}\, v(t_0)\,dF_Z(t) \\
&\quad + \int_{t_0}^\infty \{u(t) - \mathrm{E}[u(Z)]\}\, v(t_0)\,dF_Z(t) \\
&\geq 0. \qquad \square
\end{aligned}$$

2.4 Additivity

For a risk measure it is often desirable to have some sort of additivity criterion for aggregating/allocating the total risk measure of a risky portfolio.

In many situations, a natural requirement for a risk measure is subadditivity, that is,

$$\mathrm{R}[X_1 + X_2] \leq \mathrm{R}[X_1] + \mathrm{R}[X_2].$$

Artzner et al. (1996) argue for this subadditivity in the context of value at risk or capital requirement:

> If a risk measure, for example, were to fail to satisfy this property, then an individual would be motivated to open two accounts, one holding X_1 and the other X_2, incurring the smaller risk measure of $R[X_1]+R[X_2]$. Similarly, if a firm were required to meet a capital requirement which did not satisfy this property, it would be motivated to set up two subsidiaries.

While subadditivity may reflect risk reduction through risk pooling, additivity may be required if the risk pooling effect does not exist. Because there is no hedge among comonotonic risks, it would be desirable to require that a risk measure be additive for comonotonic risks. Recall that insurance layers of the same risk are comonotonic; thus it is desirable for risk measure to be additive for insurance layers.

2.5 Ordering of Risks

There are widely accepted methods for comparing risks in the economic and statistical theories of risk. The most basic concepts are the first and second stochastic dominance.

A risk X is smaller than risk Y in the *first stochastic dominance* (notation $X \prec_{1st} Y$) if any of the following equivalent conditions is met (Hadar and Russell 1969):

1. For every decision-maker with an increasing utility function u: $\mathrm{E}[u(-X)] \geq \mathrm{E}[u(-Y)]$, *that is, a common ordering shared by all individuals who prefer more wealth to less.*
2. $S_X(t) \leq S_Y(t)$, for all $t \geq 0$, *that is, for every t, risk Y has a higher tail probability.*
3. Y is derived from X by the addition of a random variable that is non-negative with probability one, *that is, additional risk is added.*

A risk X precedes Y under *second stochastic dominance* (notation $X \prec_{2nd} Y$) if any of the following equivalent conditions holds (Rothschild and Stiglitz (1970):

1. For every decision-maker who has an increasing concave function u:
$$\mathrm{E}[u(-X)] \geq \mathrm{E}[u(-Y)],$$
that is, a common ordering shared by all risk-averse individuals.
2. $$\int_x^\infty S_X(t)\,dt \leq \int_x^\infty S_Y(t)\,dt, \quad \text{for all} \quad x \geq 0,$$
that is, the net expected loss for the layer $[x,\infty)$ is always higher with risk Y. For this reason, second stochastic dominance is also called stop-loss ordering in the actuarial literature (Kaas et al. 1994).
3. $Y \stackrel{D}{=} X+Z$ in which $\mathrm{E}[Z|X] \geq 0$ with probability one, *that is, Y is equal (in distribution) to X plus noise Z.*

Now we apply these concepts in comparing the relative riskiness of different layers of the same risk.

Lemma 2.4

Given a risk Y,

- For two layers $(a, a+h]$ and $(b, b+h]$ with the same limit h,
$$a < b \Rightarrow X_{(b,b+h]} \prec_{1st} X_{(a,a+h]}$$
- For two layers $(a, a+h_1]$ and $(b, b+h_2]$ with the same expected loss, that is, $\mathrm{E}[X_{(a,a+h_1]}] = \mathrm{E}[X_{(b,b+h_2]}]$, we have
$$X_{(a,a+h_1]} \prec_{2nd} X_{(b,b+h_2]}$$

Proof

See Wang (1996a, pp. 74–75). □

Now we are ready to move on the discussions of a right-tail risk measure.

3. THE CERTAINTY EQUIVALENT APPROACH

For a risk X, let

$$\mathrm{H}[X] = \mathrm{E}[X] + \mathrm{R}[X]$$

represent the certainty equivalent to risk X, or the price for transferring the risk X to other parties. We are hoping that a theory on the certainty equivalent $\mathrm{H}[X]$ may induce a measure of the right-tail risk, $\mathrm{R}[X]$.

Wang, Young, and Panjer (1997) discuss the axiomatic characterization of insurance prices. They propose five basic axioms for insurance prices and give interpretations for each of them.

Axiom 1. For a given market condition, the price of an insurance risk X depends only on its distribution. That is, if $S_X=S_Y$, then $\mathrm{H}[X]=\mathrm{H}[Y]$.

Axiom 2. If $X \leq Y$ with probability one, then $\mathrm{H}[X] \leq \mathrm{H}[Y]$.

Axiom 3. If X and Y are comonotonic, then $\mathrm{H}[X+Y]=\mathrm{H}[X]+\mathrm{H}[Y]$.

Axiom 4. For $d \geq 0$, $\lim_{d\to 0+} \mathrm{H}[X_{(d,\infty)}]=\mathrm{H}[X]$, and $\lim_{d\to\infty} \mathrm{H}[\min(X,d)]=\mathrm{H}[X]$.

Theorem 3.1 (Wang, Young, and Panjer 1997)

If the certainty equivalent $\mathrm{H}[X]$ satisfies Axioms 1–4 and $\mathrm{H}[1]=1$, then H has the following representation:

$$\mathrm{H}[X] = \int_0^\infty g[S_X(t)]\,dt,$$

where g is a *distortion* function (that is, increasing with $g(0)=0$ and $g(1)=1$).

Furthermore, the $\mathrm{H}[X]$ in Theorem 3.1 has the following properties

- $\mathrm{H}[aX+b]=a\mathrm{H}[X]+b$ for a, $b \geq 0$.
- $\mathrm{H}[X] \geq \mathrm{E}[X]$ if and only if $g(x) \geq x$ for all $x \in [0,1]$.
- H preserves the second stochastic dominance if and only if g is concave.
- If g is concave, then $\mathrm{H}[X+Y] \leq \mathrm{H}[X]+\mathrm{H}[Y]$.

As a by product of this certainty equivalent approach, we obtain a general class of risk measures for the right-tail deviation:

$$\mathrm{R}[X] = \int_0^\infty g[S_X(t)]\,dt - \mathrm{E}[X],$$

where g is increasing concave with $g(0)=0$ and $g(1)=1$.

In addition to the four basic axioms for insurance prices, Wang, Young, and Panjer (1997) also propose the following axiom for reduction of compound Bernoulli risks and give a no-artbitrage interpretation.

Axiom 5. Let $Y=BX$ be a compound Bernoulli risk, where the Bernoulli frequency random variable B is independent of the loss severity random variable $X=Y|Y>0$. Then the market prices for risks $Y=BX$ and $B\mathrm{H}[X]$ are equal.

Theorem 3.2 (Wang, Young, and Panjer 1997)

If the insurance price $\mathrm{H}[X]$ satisfies Axioms 1–5, then we have the following unique representation:

$$\mathrm{H}[X] = \int_0^\infty [S_X(t)]^r\,dt,$$

where $0 \leq r$. Furthermore, if $\mathrm{H}[X] \geq \mathrm{E}[X]$, then $0 \leq r \leq 1$.

Besides Theorem 3.2, there are many other reasons to use distortions g of the form $g(x)=x^r$, $0 \leq r \leq 1$; see Wang (1996a, b) for further details. In order to yield a sharp numerical indicator for the right-tail risk, in this paper we choose the square root function: $g(x)=\sqrt{x}$.

4. THE RIGHT-TAIL DEVIATION

As a consequence of the certainty equivalent theory in Theorems 3.1 and 3.2, we introduce a new risk-measure for the right-tail deviation.

Definition 4.1

For a non-negative random variable X with decumulative distribution function (*ddf* $S_X(t)=\Pr\{X>t\}$, we define the *right-tail deviation* as

$$\mathrm{D}[X] = \int_0^\infty \sqrt{S_X(t)}\,dt - \int_0^\infty S_X(t)\,dt,$$

and a *right-tail index* as

$$d(X) = \frac{\mathrm{D}[X]}{\mathrm{E}[X]} = \frac{\int_0^\infty \sqrt{S_X(t)}\,dt}{\int_0^\infty S_X(t)\,dt} - 1.$$

It is straightforward to show that the right-tail deviation $\mathrm{D}[X]$ satisfies:

- If $\Pr\{X=b\}=1$, then $\mathrm{D}[X]=0$.
- Scale invariant: $\mathrm{D}[cX]=c\mathrm{D}[X]$ for $c>0$.
- Shift invariant: $\mathrm{D}[X+b]=\mathrm{D}[X]$ for any constant b.
- Subadditivity: $\mathrm{D}[X+Y] \leq \mathrm{D}[X]+\mathrm{D}[Y]$.
- If X and Y are comonotonic (including perfect correlation), then $\mathrm{D}[X+Y]=\mathrm{D}[X]+\mathrm{D}[Y]$.

Proposition 4.1

For a small layer $(t, t+dt]$, we have

1. $\mathrm{D}[X_{(t,t+dt]}] \leq \sigma[X_{(t,t+dt]}]$
2. For a small layer $[t, t+dt)$, the ratio

$$\frac{\mathrm{D}[X_{(t,t+dt]}]}{\sigma[X_{(t,t+dt]}]}$$

is an increasing function of t.

3. If X is a continuous variable with support over $[0,\omega)$ (ω can be infinity),

$$\lim_{t\to\omega} \frac{\mathrm{D}[X_{(t,t+dt]}]}{\sigma[X_{(t,t+dt]}]} = 1.$$

4. For a non-negative random variable X, the right-tail deviation $D[X]$ is finite if, and only if, the standard deviation $\sigma(X)$ is finite.

Proof

Let $u=S_X(t)$ be the probability of hitting the layer $[t, t+dt]$. The layer payment from $[t, t+dt]$ has approximately a Bernoulli distribution:

$$\Pr\{X_{(t,t+dt]} = 0\} = 1 - u, \quad \Pr\{X_{(t,t+dt]} = dt\} = u.$$

Thus,

$$\sigma[X_{(t,t+dt]}] = \sqrt{u - u^2}\, dt.$$

On the other hand,

$$D[X_{(t,t+dt]}] = \left(\sqrt{u} - u\right) dt.$$

We can show that (1)

$$\sqrt{u} - u \le \sqrt{u - u^2} \text{ for } 0 \le u \le 1,$$

(2) the ratio

$$\left(\sqrt{u} - u\right)/\left(\sqrt{u - u^2}\right)$$

increases as u decreases, and (3)

$$\lim_{u\to 0} \frac{\sqrt{u} - u}{\sqrt{u - u^2}} = 1.$$

(4) This is based on a formula in Aebi/Embrechts/Mikosch (1992, pp. 147, Remarks (ii)). □

In other words, for a small layer at the right tail, the standard deviation and the right-tail deviation converge to each other, as demonstrated in the following example.

Example 4.1

Consider the claim distribution with a ddf:

$$S_X(t) = \left(\frac{1000}{1000 + t}\right)^2.$$

For different layers $(a, a+h]$ with fixed limit $h=1000$, we compare the standard deviation and the right-tail deviation in Table 1.

5. GINI INDEX

Historically, some long-tailed distributions have an origin in income distributions, for example, Pareto and lognormal distributions; see Arnold (1983). In social welfare studies, a celebrated measure for income inequality[1] is the Gini index.

Definition 5.1

Assume that level of wealth for all individuals in a country (community) can be summarized by a distribution, $S_X(u)$, representing the proportion of individuals with wealth in excess of u. As a measure of income inequality of a society, the Gini index is defined as

$$\text{gini}(X) = \frac{E[|X_1 - X_2|]}{E[X_1 + X_2]} = \frac{E[|X_1 - X_2|]}{2E[X]},$$

where X_1 and X_2 are independent and have the same distribution as X.

Proposition 5.1

The Gini index can be equivalently represented as

$$\text{gini}(X) = 1 - \frac{\int_0^\infty [S_X(u)]^2\, du}{\int_0^\infty S_X(u)\, du}.$$

The following proof is due to Dorfman (1979).

[1]Here "income inequality" refers to the polarization of the wealth distribution.

Table 1
Standard Deviation Versus Right-Tail Deviation: A Numerical Comparison

Layer, L	Expected Loss, $E[L]$	Standard Deviation, $\sigma(L)$	Right-tail Deviation, $D[L]$	Percentage Difference, $\sigma(L)/D[L] - 1$
(0, 1000]	500.	369.2	193.1	91.1%
(1000, 2000]	166.7	341.3	238.8	42.9
(10000, 11000]	7.576	85.43	79.44	7.55
(100000, 101000]	0.09707	9.836	9.755	0.83
(1000000, 1001000]	0.0009970	0.9983	0.9975	0.08
(10000000, 10001000]	0.9997×10^{-5}	0.09998	0.09998	0.01

Proof

Note that

$$|x - y| = 2\left[\frac{x + y}{2} - \min(x, y)\right].$$

For i.i.d. variables X and Y,

$$S_{\min(X,Y)}(t) = \Pr(X > t, Y > t) = S_X(t)^2, \quad 0 \leq t < \infty.$$

Therefore

$$\frac{1}{2}\,\mathrm{E}[|X - Y|] = \mathrm{E}\left[\frac{X + Y}{2}\right] - \mathrm{E}[\min(x, y)]$$

$$= \mathrm{E}[X] - \int_0^\infty S_X(t)^2\,dt \qquad \square$$

The higher the Gini index is, the more polarized a society is. As a measure of welfare inequality, the Gini index has the following properties:

- Each dollar transferred from the rich (higher than average) to the poor (lower than average) decreases the Gini index. Essentially this implies that the Gini index preserves the second stochastic dominance (or mean-preserving spread); see Rothschild and Stiglitz (1970) and Wang and Young (1998).
- Adding an equal amount to all persons decreases the Gini index.

For convenience, we define the *Gini mean* as

$$\mathrm{G}[X] = \frac{1}{2}\,\mathrm{E}[|X_1 - X_2|] = \int_0^\infty [S_X(t) - S_X(t)^2]\,dt,$$

or equivalently,

$$\mathrm{G}[X] = \mathrm{E}[X] - \mathrm{E}[\min(X_1, X_2)]$$

$$= \mathrm{E}[\max(X_1, X_2)] - \mathrm{E}[X],$$

where X_1 and X_2 are independent and have the same distribution as X.

Motivated by the explanations for the Gini index, now we give an intuitive interpretation of the right-tail deviation:

> Let X be an observable random variable. Suppose that there are two hidden underlying i.i.d. variables Y_1 and Y_2 such that $X = \min(Y_1, Y_2)$; then the right-tail deviation of X is just the Gini mean of Y_1. In other words, right-tail deviation of X is half the expected absolute difference between Y_1 and Y_2, or the expected difference between Y_1 and X.

The right-tail index and the Gini index are similar in their definition formula. Later we show that the role of the right-tail index $d(X)$ in measuring the right-tail risk is parallel to the role of the Gini index in measuring income inequalities.

6. PROBABILITY METRICS

In the study of various limit theorems in probability theory, *probability metrics* have been developed to measure the distance between probability distributions; see Zolotarev (1979) and Rachev (1991). Here we introduce one of the most popular probability metrics, namely, the *Kantorovich metric* as given by Rachev (1991, pp. 27–28), or *Mallows metric* as given in the actuarial literature by Aebi, Embrechts and Mikosch (1992, pp. 144–145).

Definition 6.1

For two decumulative distribution functions S_1 and S_2 on **R**, the first order (Kantorovich-) Mallows metric is defined as

$$\mathrm{M}(S_1, S_2) = \inf\{\mathrm{E}[|X - Y|]\colon S_X = S_1, \quad S_Y = S_2\}.$$

Aebi, Embrechts and Mikosch (1992, pp. 144–145) indicate that the notion of Mallows metric has many applications in insurance mathematics. They show the following results:

Proposition 6.1

Let U be a uniformly distributed random variable on [0,1]. For two decumulative distribution functions S_1 and S_2, the first-order Mallows metric can be calculated as:

$$\mathrm{M}(S_1, S_2) = \mathrm{E}[|S_1^{-1}(U) - S_2^{-1}(U)|]$$

$$= \int_{-\infty}^{\infty} |S_1(t) - S_2(t)|\,dt.$$

In other words, for any two distributions, their Mallows metric is the expected absolute difference of two comonotonic variables with the respective distributions. [For a detailed discussion on comonotonicity, see Wang and Dhaene (1997).]

For a non-negative random variable X, we can easily see that the right-tail deviation has a simple representation in terms of Mallows metric:

$$\mathrm{D}[X] = \mathrm{M}(S_X, \sqrt{S_X}),$$

and so does the Gini mean:

$$\mathrm{G}[X] = \mathrm{M}(S_X, S_X^2).$$

7. TRADITIONAL METHODS

In the actuarial literature there have been many discussions on risk measures in the context of premium calculation principles (for example, Goovaerts et al. 1984). However, most traditional methods do *not* satisfy layer additivity. In this section, we analyze how those risk measures behave when layering a risk. Interestingly, by *forcing* them to be additive for layers, we recover the right-tail deviation and Gini mean.

7.1 Variance and Standard Deviation

Variance and standard deviation are the most widely used risk measures (Bowers et al. 1986). One criticism of the variance or standard-deviation based risk measures is that they do not preserve the first stochastic dominance; see Kaas et al. (1994, p. 17).

Here we give a detailed analysis of how they behave when dividing a risk into layers. Neither variance nor standard deviation is additive for comonotonic risks. Consider the case for which X is divided into two nonoverlapping layers $X_1=(0,d]$ and $X_2=(d,\infty)$. We have in general that

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(X_1 + X_2)\\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2).\end{aligned}$$

From Lemma 2.3, we have $0 \le \mathrm{Cov}(X_1, X_2) \le \sigma(X_1)\sigma(X_2)$, and thus

$$\mathrm{Var}(X) \ge \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

but

$$\sigma(X) \le \sigma(X_1) + \sigma(X_2).$$

Consider a risk X with a finite second moment. A risk $X \ge 0$ can be divided into many small layers

$$X = \sum_{j=0}^{\infty} X_{(jh,(j+1)h]}, \quad h > 0.$$

As the division refines (that is, $h \to 0$), the small layer $X_{(jh,(j+1)h]}$ has approximately a Bernoulli distribution:

$$\Pr\{X_{(jh,(j+1)h]} = 0\} \approx 1 - S_X(jh),$$

$$\Pr\{X_{(jh,(j+1)h]} = h\} \approx S_X(jh).$$

Thus,

$$\mathrm{Var}(X_{(jh,(j+1)h]}) \approx h^2\, S_X(jh)[1 - S_X(jh)].$$

Therefore, the sum of the variances of layers approaches zero, indeed,

$$\sum_{j=0}^{\infty} \mathrm{Var}(X_{(jh,(j+1)h]}) \sim h\int_0^{\infty} [S_X(t) - S_X(t)^2]\,dt = h\mathrm{G}(X).$$

On the other hand, the standard deviation exhibits subadditivity. By dividing a risk into pieces of smaller risks, the total standard deviation would generally increase:

$$\sigma(X_1 + X_2) \le \sigma(X_1) + \sigma(X_2).$$

It is easy to verify that a small layer $X_{(jh,(j+1)h]}$ has a standard deviation

$$\sigma(X_{(jh,(j+1)h]}) \approx h\sqrt{S_X(jh)[1 - S_X(jh)]}.$$

Therefore, by dividing a risk X into many small layers $X_{(jh,(j+1)h]}$ and taking the limit as the length of the layer h goes to zero, the sum of the standard deviations of layer payments approaches to a maximum

$$\sum_{j=0}^{\infty} \sigma(\mathrm{X}_{(jh,(j+1)h]}) \to \int_0^{\infty} \sqrt{S_X(t)[1 - S_X(t)]}\,dt.$$

Definition 7.1

For any non-negative random variable X with $S_X(t)=\Pr\{X>t\}$, we define the *maximal standard deviation* as

$$\mathrm{MSD}[X] = \int_0^{\infty} \sqrt{S_X(t)[1 - S_X(t)]}\,dt.$$

The following results can be easily verified:

- The maximal standard deviation is additive for layers.
- $\mathrm{D}[X] \le \mathrm{MSD}[X]$.
- If X is a continuous variable with support over $[0, \omega)$ (ω can be infinity),

$$\lim_{t\to\omega} \frac{\mathrm{D}[X_{(t,\infty]}]}{\mathrm{MSD}[X_{(t,\infty]}]} = 1.$$

- Unlike the right-tail deviation, the maximal standard deviation does not preserve first or second stochastic dominance.

The some extent, for the proponents of the standard deviation, the right-tail deviation can be viewed as a simplified or improved version of the maximum standard deviation.

7.2 Quantile Concept

Sometimes management is concerned only with some threshold amount of loss, and quantiles (say, 95th or 99th percentile) can serve as a simple risk indicator.

It does have some drawbacks because it fails to reveal more information at other percentiles. Some authors point out that subadditivity does not always hold for quantiles; see Artzner et al. (1996) and Gerchak and Mossman (1992). As an advancement from the quantile concept, Butsic (1994) proposes the use of expected policyholder deficit in deciding risk-based-capital requirements.

7.3 Expected Policyholder Deficit

Let the threshold amount of loss be $\beta \mathrm{E}[X]$, $\beta > 0$. The expected policyholder deficit (EPD) in excess of $\beta \mathrm{E}[X]$ is defined as

$$\mathrm{EPD}_\beta[X] = \mathrm{E}[X_{(\beta \mathrm{E}[X],\infty]}] = \int_{\beta \mathrm{E}[X]}^{\infty} S_X(t)\,dt.$$

Van Heerwaarden and Kaas (1992) propose a Dutch premium calculation principle given by

$$\mathrm{H}[X] = \mathrm{E}[X] + \alpha\,\mathrm{EPD}_\beta[X], \quad 0 \le \alpha \le 1.$$

They show that $\mathrm{H}[X]$ is subadditive and preserves the second stochastic dominance. However, Wang (1996c, p. 111) gives an example that shows that $\mathrm{EPD}(\beta)$ is not additive for layers.

Now we divide a risk $X \ge 0$ into many small layers, $X_{(jh,(j+1)h]}$, $h>0$, and consider the EPD with $\beta=1$. Note that for the small layer $(t, t+dt]$ we have

$$\mathrm{E}[X_{(t,t+dt]}] = S_X(t)\,dt,$$
$$\mathrm{EPD}_1[X_{(t,t+dt]}] = [1 - S_X(t)]S_X(t)dt.$$

Therefore, by dividing a risk $X \ge 0$ into many small layers and taking a limit as the length of the layer goes to zero, we get a maximal (total) EPD_1 at

$$\int_0^{\infty} [1 - S_X(t)]\,S_X(t)\,dt = \mathrm{G}[X].$$

7.4 The *p*-th Mean Value

Goovaerts et al. (1984) discuss the method of calculating insurance premiums using the p-th mean value

$$\mathrm{E}_p[X] = \{\mathrm{E}[X^p]\}^{1/p}, \quad p \ge 1.$$

They also show that $\mathrm{E}_p[X]$ is subadditive.

By dividing a risk $X \ge 0$ into many small layers, $X_{(jh,(j+1)h]}$, we can show that the total p-th mean value approaches a maximum

$$\sum_{j=0}^{\infty} \mathrm{E}_p[X_{(jh,(j+1)h]}] \to \int_0^{\infty} [S_X(t)]^{1/p}\,dt,$$

which gives the PH-transform principle of Wang (1995). This induces a measure of deviation from the mean:

$$\mathrm{R}_p[X] = \int_0^{\infty} [S_X(t)]^{1/p}\,dt - \mathrm{E}[X],$$

where $p=2$ corresponds to our right-tail deviation.

8. EXAMPLES

Now we give some examples of how to calculate the right-tail index and the Gini index.

Example 8.1

For a Bernoulli risk with $\Pr\{X=0\}=1-q$ and $\Pr\{X=1\}=q$, the right-tail deviation is $\mathrm{D}[X]=\sqrt{q}-q$ and the Gini mean is $\mathrm{G}[X]=q-q^2$. Accordingly, the right-tail index is $d(X)=1/\sqrt{q}-1$ and the Gini index $\mathrm{gini}(X)=1-q$.

Similarly, for an arbitrary risk X, the small layer $X_{(t,t+dt]}$ can be approximated as a Bernoulli variable. Thus,

$$d(X_{(t,t+dt]}) = 1/\sqrt{S_X(t)} - 1$$

and

$$\mathrm{gini}(X_{(t,t+dt]}) = 1 - S_X(t). \qquad \square$$

Example 8.2

Assume that X has an exponential distribution with mean λ, that is, with ddf:

$$S_X(t) = e^{-t/\lambda}.$$

(a) It can be easily verified that both the right-tail index and the coefficient of variation are equal to 1 and the Gini index is 1/2.

(b) For a layer $(0, x]$ of X, the right-tail index is

$$d(X_{(0,x]}) = [1 - e^{-x/2\lambda}]^2.$$

Example 8.3

Assume that X is Pareto distributed with ddf:

$$S_X(t) = \left(\frac{\lambda}{\lambda + t}\right)^a, \qquad t \ge 0.$$

The right-tail index and the Gini index are (respectively):

$$d(X) = \begin{cases} \dfrac{\alpha}{(\alpha-1)(\alpha-2)}, & \alpha > 2 \\ \infty, & \alpha \le 2. \end{cases}$$

$$\text{gini}(X) = \begin{cases} \dfrac{\alpha}{2\alpha-1}, & \alpha > 1 \\ 1, & \alpha \le 1. \end{cases}$$

Example 8.4

Consider the exponential-inverse gaussian (E-IG) distribution with ddf:

$$S(x) = \exp\left\{\frac{\mu}{\beta}[1-(1+2\beta x)^{1/2}]\right\}, \quad x > 0.$$

The E-IG distribution has the following mean and standard deviation (Hesselager, Wang, and Willmot 1997):

$$E[X] = \frac{\beta+\mu}{\mu^2}, \quad \sigma(X) = \frac{\sqrt{5\beta^2+4\beta\mu+\mu^2}}{\mu^2}.$$

It can be easily verified that

$$D[X] = \frac{3\beta+\mu}{\mu^2}, \quad d(X) = 1 + \frac{2\beta}{\beta+\mu}.$$

Note that the right-tail deviation is greater than the standard deviation. The Gini index is

$$\text{gini}(X) = \frac{1}{2} + \frac{\beta}{4(\beta+\mu)}.$$

Example 8.5

As we have seen in Example 8.2, for an exponential distribution, the right-tail index and the coefficient of variation are all equal to 1, while the Gini index is one half. Now we want to investigate how the right-tail index behaves for probability distributions with thicker or thinner tails than an exponential distribution. To do so, we investigate the gamma(α) distribution with pdf

$$f(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \quad x \ge 0,$$

which has a mean of α and a coefficient of variation at $1/\sqrt{\alpha}$.

- When $\alpha>1$, it has a thinner tail than an exponential distribution (asymptotically).
- When $\alpha=1$, it is an exponential distribution.
- When $\alpha<1$, it has a thicker tail than an exponential distribution.

In Table 2, we compute the coefficient of variation, the right-tail index, and the Gini index for different values of α.

Table 2
Different Measures of Variation for the Gamma Distribution

Gamma (α) Distribution	Coefficient of Variation	Right-Tail Index	Gini Index
$\alpha = 5$	0.447	0.381	0.246
$\alpha = 3$	0.577	0.514	0.313
$\alpha = 2$	0.707	0.656	0.375
$\alpha = 1$	1.000	1.000	0.500
$\alpha = 1/2$	1.414	1.532	0.637
$\alpha = 1/3$	1.732	1.963	0.713
$\alpha = 1/5$	2.236	2.671	0.798

We can see that, as the right-tail becomes thicker (thinner), our right-tail index $d(X)$ increases (decreases) faster than the coefficient of variation, which in turn is faster than the Gini index.

Note that the sum of k independent gamma(1), that is, exponential, variables has a gamma(k) distribution. From Table 2 we observe that the pooling of independent exponential variables results in more reduction in the right-tail index than the coefficient of variation.

8.1 Comparison of Continuous Distributions

Consider the following loss distributions each with the same mean (=1) and variance (=3).

- A Pareto distribution with $\lambda=2$ and $\alpha=3$ (see Example 8.3).
- An E-IG distribution with $\mu=(3+\sqrt{5})/2$ and $\beta=2+\sqrt{5}$ (see Example 8.4).
- A lognormal distribution with ddf

$$S(t) = \Psi\left(\frac{\ln t - \mu}{\sigma}\right), \quad t > 0,$$

where $\mu=-\ln 2$, $\sigma=\sqrt{\ln 4}$ and $\Psi(.)$ is the ddf for a standard normal distribution.
- A gamma distribution with pdf

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1}e^{-\lambda t}, \quad t > 0,$$

where $\alpha=\lambda=1/3$.
- An inverse-Gaussian distribution with pdf

$$f(t) = \mu(2\pi\beta t^3)^{-1/2}\exp\left\{-\frac{(t-\mu)^2}{2\beta t}\right\}, \quad t > 0,$$

where $\mu=1$ and $\beta=3$.

- A Weibull distribution with a ddf

$$S(t) = \exp\{-at^b\},$$

where $a=0.607248$ and $b=1.26957$.

Without referring to higher moments, we can order them by the right-tail index $d(X)$.

Table 3
Ranking of Some Continuous Probability Distributions by the Right-Tail Index

Continuous Distributions	Mean	Coefficient of Variation	Right-Tail Index	Gini Index
Pareto	1	$\sqrt{3}$	3.00	0.600
Lognormal	1	$\sqrt{3}$	2.59	0.595
E-IG	1	$\sqrt{3}$	2.24	0.655
Inverse-Gaussian	1	$\sqrt{3}$	2.17	0.632
Weibull	1	$\sqrt{3}$	2.13	0.681
Gamma	1	$\sqrt{3}$	1.96	0.713

For these parametric distributions, the ranking by the right-tail index is in agreement with our knowledge of their relative tail thickness; see Embrechts and Veraverbeke (1982) and Panjer and Willmot (1992, p. 350). However, the Gini indices are not in agreement with the commonly perceived ranking of tail thickness.

In summary, as its name may suggest, the right-tail deviation measures the right-tail risk, as opposed to the standard deviation, which measures the deviation about the mean, and as opposed to the Gini index, which measures the polarization of the wealth distribution.

8.2 Comparison of Counting Distributions

In the same way that the right-tail index $d(X)$ is used for claim severity distributions, we can compare claim frequency distributions. Here we consider the most popular two-parameter counting distributions, namely:

- A two-point Bernoulli distribution with a probability function $f(0)=3/4$, and $f(4)=1/4$.
- Negative binomial distribution with a probability function:

$$p_n = \frac{\Gamma(r+n)}{\Gamma(r)\, n!}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^n, \quad n = 0, 1, 2, \ldots$$

  where $r=0.5$ and $\beta=2$.
- Poisson inverse gaussian distribution with a probability generating function:

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = \exp\left\{\frac{\mu}{\beta} - \frac{\mu}{\beta}\sqrt{1 - 2\beta(z-1)}\right\},$$

  where $\mu=1.0$ and $\beta=2.0$. Note that the probabilities can be evaluated recursively, see Willmot (1987, pp. 114–115).
- Generalized Poisson distribution with a probability function (see Consul 1990):

$$p_n = \theta(\theta + n\lambda)^{n-1}\frac{e^{-\theta-n\lambda}}{n!}, \quad n = 0, 1, 2, \ldots$$

where $\theta=1/\sqrt{3}$ and $\lambda=1-1/\sqrt{3}$.

The parameters were chosen so that all three distributions have the same mean (=1) and variance (=3). Their right-tail indices and Gini indices can be easily computed, as listed in Table 4.

Table 4
Ranking of Some Discrete Probability Distributions by the Right-Tail Index

Discrete Distributions	Mean	Coefficient of Variation	Right-Tail Index	Gini Index
Poisson Inverse Gaussian	1	$\sqrt{3}$	2.01	0.72
Generalized Poisson	1	$\sqrt{3}$	1.94	0.73
Negative Binomial	1	$\sqrt{3}$	1.87	0.74
Bernoulli	1	$\sqrt{3}$	1.00	0.75

The right-tail index ranks P-IG as having a fatter tail than generalized Poisson, which in turn has a fatter tail than the negative binomial distribution. This ranking is consistent with their asymptotic behavior; see Willmot (1997) and Consul (1990). However, the Gini index indicates a reversed ordering. This suggests that the Gini index is not a good measure for the right-tail risk.

9. APPLICATIONS IN RISK CHARGES

Similar to the standard deviation principle, we can use the right-tail deviation in deciding risk-charges:

$$\mathrm{E}[X] + \beta \mathrm{D}[X].$$

As shown in Wang (1997), $\mathrm{E}[X]+\beta\mathrm{D}[X]$ can also be rewritten as a two-point mixture of PH-means:

$$(1-\beta)\mathrm{H}_1[X] + \beta\mathrm{H}_{0.5}[X], \qquad 0 < \beta < 1,$$

where $\mathrm{H}_r[X]=\int_0^\infty [S_X(t)]^r dt$ for $X \geq 0$.

For ratemaking purposes, the right-tail deviation has some definite advantages over the standard deviation:

1. It automatically takes account of all moments without explicitly calculating them.
2. It preserves the second stochastic dominance and thus is more discriminative than the standard deviation. For layers with fixed width, the premium value decreases at higher layers, but the relative risk loading increases at higher layers; see Venter (1991).
3. It is additive for excess-of-loss layers and thus yields the same premium no matter how the risk is divided into sublayers. This is a desirable property for increased limits ratemaking.

For a layer $(a, a+h]$ of risk X, the *rate-on-line* (ROL) is defined as the ratio of expected layer payments $X_{(a,a+h]}$ to the limit h:

$$\mathrm{ROL}_{(a,a+h]} = \frac{\mathrm{E}[X_{(a,a+h]}]}{h}.$$

It can be easily verified that

$$\lim_{h\to 0+} \mathrm{ROL}_{(a,a+h]} = S_X(a).$$

In practice, the reinsurer can estimate the empirical rate-on-line using the ratio of the average observed layer payment to the limit. An empirical rate-on-line function over different layers serves as an approximation of the ddf S_X.

As a pragmatic method for deciding risk loads, some reinsurers (for example Swiss Re) use a multiple of the square root of the empirical or estimated ROL. As the layers (h becomes small) are refined, this pragmatic method yields an approximation of a theoretical formula:

$$(1 + \beta)\mathrm{E}[X_{(a,a+h]}] + \beta\mathrm{D}[X_{(a,a+h]}].$$

10. A Bridge to Measuring Parameter Risk

Many actuaries agree on the importance of parameter uncertainty in pricing insurance risks. However, there is little agreement on how to measure the parameter risk. This situation is evident from the recent discussions on risk loads by fellows of the Casualty Actuarial Society, for example, Miccolis (1977), Feldblum (1990), and Meyers (1991). In this section we show that the right-tail deviation can be used as a bridge to modeling parameter uncertainties.

Assume that we have a finite sample of n observations from a class of identical insurance policies. An empirical estimate for the loss distribution is

$$\hat{S}(t) = \frac{\#\text{ of observations} > t}{n}, \quad t \geq 0.$$

Let $S(t)$ represent the *true* underlying loss distribution, which is generally unknown and different from the empirical distribution $\hat{S}(t)$.

From statistical estimation theory (for example, Lawless 1982, p. 402), for some specified value of t, we can treat the quantity $[\hat{S}(t)-S(t)]/[\sigma(\hat{S}(t))]$ as having a standard normal distribution for large values of n, where

$$\sigma(\hat{S}(t)) \approx \frac{\sqrt{\hat{S}(t)[1 - \hat{S}(t)]}}{\sqrt{n}}.$$

The 100η% upper confidence limit for the true underlying distribution $S(t)$ can be approximated by

$$S^U(t) = \hat{S}(t) + \beta\sqrt{\hat{S}(t) - [\hat{S}(t)]^2}, \quad \beta = \frac{q_\eta}{\sqrt{n}},$$

where q_η is a quantile of the standard normal distribution: $\Pr\{N(0,1)\leq q_\eta\}=\eta$.

As a means of dealing with parameter risk regarding the best-estimate $\hat{S}_X(t)$, we use

$$\hat{S}_Y(t) = \hat{S}_X(t) + \beta\left\{\sqrt{\hat{S}_X(t)} - \hat{S}_X(t)\right\}, \quad \beta = \frac{q_\eta}{\sqrt{n}},$$

as an approximation of the 100η% upper confidence limit $S^U(t)$.

Note that $\hat{S}_Y(t)$ always lies within the interval [0,1], while $S^U(t)$ can be greater than one. The ddf $\hat{S}_Y(t)$ implies a risk-load formula:

$$\mathrm{E}[X] + \beta\mathrm{D}[X].$$

11. Conclusion

We have discussed the new concept of right-tail deviation for a random variable $X\geq 0$:

$$\mathrm{D}[X] = \int_0^\infty \sqrt{\Pr\{X > t\}}\, dt - \mathrm{E}[X].$$

We have shown that it is a simple yet powerful measure of the right-tail risk. It outperforms the standard deviation and Gini mean and has a great potential for actuarial applications.

ACKNOWLEDGMENT

I wish to thank John A. Beekman and Virginia R. Young for numerous helpful comments and suggestions. I also thank Paul Embrechts for pointing out a couple of useful references.

REFERENCES

AEBI, M., EMBRECHTS, P., AND MIKOSCH, T. 1992. "A Large Claim Index," *Bulletin of the Association of Swiss Actuaries* 2: 143–56.

ARNOLD, B.C. 1983. *Pareto Distributions*. Fairland, Md.: International Co-operative Publishing House.

ARTZNER, P., DELBAEN, F., EBER, J-M., AND HEATH D. 1996. "A Characterization of Measures of Risk," working paper, February.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. 1986. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries.

BUTSIC, R.P. 1994. "Solvency Measurement for Property-Liability Risk-Based Capital Applications," *Journal of Risk and Insurance* 61, no. 4:656–90.

CONSUL, P.C. 1990. "A Model for Distributions of Injuries in Auto-Accidents," *Journal of the Swiss Association of Actuaries*, 161–68.

DORFMAN, R. 1979. "A Formula for the Gini Coefficient," *Review of Economics and Statistics* 61:146–49.

EMBRECHTS, P., AND VERAVERBEKE, N. 1982. "Estimates for the Probability of Ruin with Special Emphasis on the Possibility of Large Claims," *Insurance: Mathematics and Economics* 1:55–72.

FELDBLUM, S. 1990. "Risk Loads for Insurers," *Proceedings of the Casualty Actuarial Society* LXXVII:160–95.

GERBER, H.U., AND SHIU, E.S.W. 1997. "The Joint Distribution of the Time of Ruin, the Surplus Immediately before Ruin, and the Deficit at Ruin," *Insurance: Mathematics and Economics*, 21:129–37.

GERCHAK, Y., AND MOSSMAN, D. 1992 "On the Effect of Demand Randomness on Inventories and Costs, *Operations Research* 40:804–807.

GOOVAERTS, M.J., DE VYLDER, F., AND HAEZENDONCK, J. 1984. *Insurance Premiums: Theory and Applications*. Amsterdam: North-Holland.

HADAR, J., AND RUSSELL, W.R. 1969. "Rules for Ordering Uncertain Prospects," *American Economic Review* 59, no. 1: 25–34.

HESSELAGER, O., WANG, S., AND WILLMOT, G.E. 1997. "Exponential and Scale Mixtures and Equilibrium Distributions," *Scandinavian Actuarial Journal*, in press.

KAAS, R., VAN HEERWAARDEN, A.E., AND GOOVAERTS, M.J. 1994. *Ordering of Actuarial Risks*, Education Series 1. Brussels: CAIRE.

LAWLESS, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

LOWE, S.P., AND STANARD, J.N. 1996. "An Integrated Dynamic Financial Analysis and Decision Support System for a Property Catastrophe Reinsurer," *Transactions of the XXVII ASTIN Colloquium*, Copenhagen, Denmark.

MEYERS, G.G. 1991. "The Competitive Market Equilibrium Risk Load Formula for Increased Limits Ratemaking," *Proceedings of the Casualty Actuarial Society* LXXVIII:163–200.

MICCOLIS, R.S. 1977. "On the Theory of Increased Limits and Excess-of-Loss Pricing," *Proceedings of the Casualty Actuarial Society* LXIV:27–59.

PANJER, H.H., AND WILLMOT, G.E. 1992. *Insurance Risk Models*. Schaumburg, Ill.: Society of Actuaries.

PHILBRICK, S.W. 1994. "Accounting for Risk Margins," *Casualty Actuarial Society Forum* 1(Spring):4–87.

RACHEV, S.T. 1991. *Probability Metrics and the Stability of Stochastic Models*. Chichester, England: Wiley.

RAMSAY, C.M. 1993. "Loading Gross Premiums for Risk Without Using Utility Theory," *Transactions of the Society of Actuaries* XLV:305–49.

ROTHSCHILD, M., AND STIGLITZ, J.E. 1970. "Increasing Risk I: a Definition," *Journal of Economic Theory* 2:225–43.

SCHMEIDLER, D. 1986. "Integral Representation without Additivity," *Proceedings of the American Mathematical Society* 97:255–61.

VAN HEERWAARDEN, A.E., AND KAAS, R. 1992. "The Dutch Premium Principle," *Insurance: Mathematics and Economics* 11:129–133.

VENTER, G.G. 1991. "Premium Calculation Implications of Reinsurance without Arbitrage," *ASTIN Bulletin* 21:223–30.

WANG, S. 1995. "Insurance Pricing and Increased Limits Ratemaking by Proportional Hazards Transforms," *Insurance: Mathematics and Economics* 17:43–54.

WANG, S. 1996a. "Premium Calculation by Transforming the Layer Premium Density," *ASTIN Bulletin* 26:71–92.

WANG, S. 1996b. "Ambiguity-Aversion and the Economics of Insurance," *Proceedings of Risk Theory Seminar*, University of Wisconsin–Madison.

WANG, S. 1996c. "Ordering of Risks under PH Transforms," *Insurance: Mathematics and Economics* 18, no. 2:109–114.

WANG, S. 1997. "Implementation of PH-Transforms in Ratemaking," *Proceedings of the Casualty Actuarial Society*, in press.

WANG, S., AND DHAENE, J. 1997. "Comonotonicity, Dependency, and Stop-Loss Premiums," working paper.

WANG, S., AND YOUNG, V.R. 1998. "Ordering of Risks: Utility Theory Versus Yaari's Dual Theory of Risk," *Insurance: Mathematics and Economics*, in press.

WANG, S., YOUNG, V.R., AND PANJER, H.H. 1997. "Axiomatic Characterization of Insurance Prices," *Insurance: Mathematics and Economics* 21:173–83.

WILLMOT, G. 1987. "The Poisson-Inverse Gaussian as an Alternative to the Negative Binomial," *Scandinavian Actuarial Journal*, 113–127.

YAARI, M.E. 1987. "The Dual Theory of Choice under Risk," Econometrica 55:95–115.

ZOLOTAREV, V.M. 1979. "Ideal Metrics in the Problem in Topological Space and Probability Theory," *The Australian Journal of Statistics* 21:193–208.

APPENDIX

EXTENSION TO REAL-VALUED RANDOM VARIABLES

For a real-valued random variable X, we have

$$E[X] = \int_{-\infty}^{0} [S_X(t) - 1]\,dt + \int_{0}^{\infty} S_X(t)\,dt.$$

In terms of Mallows metric, the right-tail deviation can be extended to real-valued random variables as:

$$D[X] = M(S_X, \sqrt{S_X}) = \int_{-\infty}^{\infty} [\sqrt{S_X(t)} - S_X(t)]\,dt.$$

In the context of insurance assets, the main concern is the downside risk, that is, poorer than expected performance in the investment portfolios or failure of promised payments by the borrowers. If one looks at the distribution for investment prospects, the left-tail constitutes the major investment risk. As a measure of the left-tail risk, we define the following quantity:

$$D^*[X] = D[-X] = \int_{-\infty}^{\infty} \left[\sqrt{F_X(t)} - F_X(t)\right] dt$$

Now we also define the *two-sided deviation* as

$$\begin{aligned}\Delta[X] &= \frac{D[X]+D^*[X]}{2} \\ &= \frac{1}{2}\left\{\int_{-\infty}^{\infty} \left[\sqrt{S_X(t)} + \sqrt{1 - S_X(t)} - 1\right] dt\right\},\end{aligned}$$

and the *two-sided tail index* as

$$\delta(X) = \frac{\Delta[X]}{E[X]}.$$

Note that the two-sided deviation can be readily expressed in terms of Mallows metric

$$\Delta[X] = \frac{1}{2} M\left(\sqrt{S_X}, 1 - \sqrt{1 - S_X}\right).$$

Proposition 12.1

- $\Delta[X + b] = \Delta[X]$, for $-\infty < b < \infty$.
- $\Delta[aX] = a\Delta[X]$, for $-\infty < a < \infty$.
- $\Delta[X + Y] \le \Delta[X] + \Delta[Y]$.
- If X and Y are comonotonic, then $\Delta[X + Y] = \Delta[X] + \Delta[Y]$.

Remarks

The term $\Delta[X]$ may be difficult to evaluate numerically for many parametric distributions, but there is no such problem for empirical sample distributions.

Discussions on this paper can be submitted until October 1, 1998. The author reserves the right to reply to any discussion. See the Submission Guidelines for Authors for detailed instructions on the submission of discussions.

new scheme would probably not have gotten off the ground. A defined-contribution scheme can generate adequate pensions only for those who enter it when young. For older entrants, under the civilized principle of social cohesiveness and "solidarity," some other means has to be found, like the Chilean government guarantee. Also, without the underpinning of this guarantee, it would probably have been politically impossible to start the new system. Thus the government guarantee is an essential, initial, ongoing component of the new system. So Chile is in fact a *hybrid* system.

2. Chile gives the employee the best of both worlds: the pension generated by his own defined contributions or the guaranteed minimum pension if higher. There is also a financial exposure for government if an AFP goes belly-up. These propositions will be too expensive for Pakistan. They will create a large current and future burden on government, which it cannot afford. I regret that those who advocated the Chilean scheme for Pakistan failed to mention this aspect.
3. According to literature available here, it seems Chile does not yet have well-developed employer-sponsored retirement benefit systems. In countries with such systems, it would not be a good idea to mandate a defined-contribution wage-related system like that in Chile.
4. According to its official summary, the World Bank's October 1994 study *Averting the Old Age Crisis* recommended a basic state-provided defined-benefit flat-rate subsistence pension, indexed to the cost of living, plus a mandatory second-tier funded system, which could be either defined-benefit or defined contribution. The defined contribution or AFP component of the Chilean system is analogous to the second tier recommended by the World Bank; but the government-guaranteed defined-benefit minimum pension is analogous to the World Bank's first-tier, means-tested against the AFP pension. A first-tier statutory scheme has to be defined benefit to meet its basic objective of alleviating poverty for elderly persons. Defined-contribution schemes simply cannot respond to this need. For the basic state-provided defined-benefit flat-rate subsistence pension, the World Bank's publication advocated paygo. I have already indicated my views on this.

For these reasons, plus the cost, compliance, and other issues mentioned in Dr. Brown's paper, the Chilean model is not necessarily suitable for all countries. My own country, Pakistan, has a number of employer-sponsored funded retirement schemes, both defined contribution and defined benefit. Instead of the Chilean system, we should modify the present first-tier statutory scheme to make it a flat-rate defined-benefit subsistence pension for all working Pakistanis, financed by employers and employees, with no government contribution, as I advocated in January 1993. There should be no automatic increases or indexation, because automatics can create severe problems that we cannot afford. But the flat rate should be reviewed periodically in the light of actuarial recommendations, which would take into account the long-range soundness of the scheme, as well as the cost of living.

And the government should do what it can to strengthen and build upon the existing second-tier system of gratuity funds, provident funds, and pension funds run by employers, usually with some employee contributions. However, I do not think this second tier should become mandatory.

"An Actuarial Index of the Right-Tail Risk," Shaun Wang, April 1998

James G. Berberian*

The right-tail index introduced by Dr. Wang is a useful representation of the risk associated with large-loss events. It is an interesting approach that seems to outperform the common measures in identifying and ordering right-tail risks.

The "pessimist's premium" concepts defined below provide another framework for analyzing such risks. This methodology can be generalized to accommodate utility preferences, including the right-tail index.

Definitions

Assume $X \sim F(x)$ is a continuous, non-negative distribution and $S(x) = 1 - F(x)$. Then define the tail function

$$T(t) = \int_t^\infty S(x)dx.$$

*James G. Berberian, A.S.A., is a Senior Manager at KPMG Peat Marwick LLP, 303 Peachtree Street NE, Suite 2000, Atlanta, GA 30308, e-mail, jberberian@kpmg.com.

Generate the means:

$$\mu_1 = \int_0^\infty S(x)dx = \mathrm{E}(X) = T(0) = T(\mu_0)$$

$$\{\text{where } \mu_0 \equiv 0 \text{ by definition}\},$$

$$\mu_2 = \mathrm{E}[X|X > \mu_1] = \mu_1 + \frac{T(\mu_1)}{S(\mu_1)},$$

$$\mu_{i+1} = \mathrm{E}[X|X > \mu_i] = \mu_i + \frac{T(\mu_i)}{S(\mu_i)}.$$

Using these, define

$$T = \sum_{i=0}^{\infty} T(\mu_i)$$

and the risk premium,

$$r = \frac{T - \mu}{\mu}.$$

Example 1

Consider an exponential distribution with mean μ. $S(x) = e^{-x/\mu}$ and $T(t) = \mu e^{-t/\mu}$, generating $\mu_i = i\mu$. Then $T(\mu_i) = \mu e^{-i}$ and $T = \mu[e/(e - 1)]$, giving $r = 1/(e - 1)$. The risk premium is about 58%. Note that the right-tail index is 100%.

Example 2

For $S(x) = [1000/(1000 + x)]^2$, it can be shown that $\mu = 1000$, $T = 2000$, and $r = 100\%$. Note that the right-tail index over $(0, \infty)$ does not converge since $[S(x)]^{1/2}$ is of order $1/x$.

Pessimist's Premium

The terms of T can be rearranged to gain insight into the nature of the associated risk premium. From the definitions above, $T(\mu_i) = S(\mu_i)(\mu_{i+1} - \mu_i)$. Then,

$$T = \sum_{i=0}^{\infty} T(\mu_i) = \sum_{i=0}^{\infty} S(\mu_i)(\mu_{i+1} - \mu_i)$$

$$= \sum_{i=0}^{\infty} \mu_{i+1}\,[S(\mu_i) - S(\mu_{i+1})]$$

$$\{\text{by rearrangement}\},$$

$$= \sum_{i=0}^{\infty} \mu_{i+1}\,P[\mu_i < X \le \mu_{i+1}].$$

This expresses T as the expected value of a discrete distribution of $\{\mu_i, p_i\}$. From the pessimist's point of view, the mean $\mu_1 = \mu$ is a reasonable estimator for X as long as X does not exceed μ_1. Hence, the pessimist accords μ_1 only the weight $P[\mu_0 < X \le \mu_1]$ (remember $\mu_0 \equiv 0$ by definition). If X exceeds μ_1, then perhaps μ_2 is an acceptable estimator for $X|X > \mu_1$ (weighted $P[\mu_1 < X \le \mu_2]$ by the pessimist), unless, of course, X exceeds μ_2 . . . . Continuing, we rationalize the expected value T. The resulting risk premium r might then be called the pessimist's premium.

Discrete or Sample Distributions

The risk premium defined by $(T - \mu)/\mu$ can be easily extended to non-negative discrete or empirical distributions. Without loss of generality, $x_0 = 0$:

$$\mu_1 = \sum_{i=0}^{\infty} S(x_i)(x_{i+1} - x_i)$$

and

$$\mu_{k+1} = x_k + [S(x_k)]^{-1} \sum_{i=k}^{\infty} S(x_i)(x_{i+1} - x_i),$$

where x_k is the largest x_i less than or equal to μ_k. (The process can be terminated if $x_z = x_{z-1}$, or if finite data are exhausted). Then,

$$T = \sum_{k=0}^{\infty} T(x_k).$$

Example 3

Consider an annuity that pays \$1 at the beginning of each year while an individual is alive. Assume a fixed interest rate of 8% applies during the period of payments. Let X be the (random variable) present value of such a policy. Using a common mortality table, the actuarial present value of such a policy issued to a person age 65 is \$9.11 (= expected value = μ_1). Calculating T yields a risk premium $r = 17.88\%$, implying that a pessimistic insurer might charge \$10.74 for such a policy. Note that the right-tail index for this scenario is 14.18%.

Utility Preferences

The pessimist's premium can be generalized to reflect other utility preferences. Define

$$T^* = \sum_{i=0}^{\infty} w_i\, T(\mu_i),$$

and the resulting

$$r^* = \frac{T^* - \mu}{\mu},$$

for which the w_i can be chosen to reflect the individual's aversion to large-loss events. For example, the "risk-neutral" individual who focuses only on the expected loss can be represented with $w_0 = 1$ and $w_i = 0$ for all other i. The pessimist's premium results from setting $w_i = 1$ for all i. Probabilities too remote for real-world consideration can be eliminated by setting $w_i = 0$ for all i greater than or equal to a given cutoff (remember Bernoulli's paradox). Risk aversion exceeding the pessimist's premium can be reflected with appropriate w_i, as can risk-seeker profiles where $T^* < \mu$ holds.

The right-tail index can be generated using specific w_i. If the integral converges, we can write

$$\int_0^\infty [S(x)]^{1/2}\,dx = \int_{\mu_0(=0)}^{\mu_1} [S(x)]^{1/2}\,dx + \int_{\mu_1}^{\mu_2} [S(x)]^{1/2}\,dx + \int_{\mu_2}^{\mu_3} [S(x)]^{1/2}\,dx + \ldots$$

Then, choosing w_i such that

$$w_i = \frac{\int_{\mu_i}^{\mu_{i+1}} [S(x)]^{1/2}\,dx}{S(\mu_i)(\mu_{i+1} - \mu_i)}$$

will give r^* equal to the right-tail index. For a function $g[S(x)]$ whose integral converges, we can similarly generate the corresponding r^* with weights

$$w_i = \frac{\int_{\mu_i}^{\mu_{i+1}} g[S(x)]dx}{S(\mu_i)(\mu_{i+1} - \mu_i)}.$$

Multiple Policies

If a risk premium r' has been determined for a group of n policies, the expected profit when $(1 + r')\,\mu$ is charged for each of the n policies is $\mu n r'$. If the underlying claim distribution has a finite variance σ^2, the sum of the policy profits should be approximately normally distributed with mean $\mu n r'$ and standard error $n^{1/2}\sigma$. Probabilities related to profits can be derived from the normal values.

Alternatively, we can use the normal approximation to create r_z that will generate profit with probability $p_z = P[N(0, 1) > -z]$. (For example, if $z = 2.326342$, the normal table shows $P[N(0, 1) > -2.326342] = 99\%$.) Set $r_z = z\sigma/(\mu n^{1/2})$. Then $\mathrm{E}[Profit] = \mu n r_z = z\sigma\, n^{1/2}$, and the standard error $\sigma[Profit] = \sigma n^{1/2}$. From the normal approximation,

$$\begin{aligned} &P[Profit > 0] \\ &\quad = P[N(z\sigma\, n^{1/2}, \sigma n^{1/2}) > 0] \\ &\quad = P[N(0, 1) > -z] = p_z. \end{aligned}$$

Note that this r_z can be created through T^* and r^* by setting $w_0 = 1$,

$$w_1 = \frac{r_z\mu}{T(\mu)} = \frac{z\sigma}{[n^{1/2}T(\mu)]},$$

and $w_i = 0$ for all i greater than 1. Then $T^* = \mu + r_z\mu$ and $r^* = r_z$. This corresponds to a utility preference that evaluates only the probability of profit under a normal approximation.

A pessimist, however, might not be comfortable with $w_i = 0$ for all i greater than 1. If instead $w_i = r_z\mu/T(\mu)$ for all i greater than 0 ($w_0 = 1$), then it follows that $r^* = r_z\{r\mu/T(\mu)\}$. Here, r is the pessimist's premium derived previously. Hedging the risk premium r_z upwards is equivalent to loading σ by $\{r\mu/T(\mu)\}$ in the development of r_z. Accordingly, $\sigma^* = \sigma\{r\mu/T(\mu)\}$ can be considered a pessimist's standard error.

Example 4

In the life annuity Example 3, $\mu = \$9.11$ and $\sigma = \$2.80$. The pessimist's premium is $r = 0.1788$ and $T(\mu) = 1.2423$. Consider 100 identical, independent policies. The risk premium r_z for $z = 2.326342$ is $(2.326342)(\$2.80)/\{\$9.11(10)\} = 0.071501$. Expected profit is $(2.326342)(\$2.80)(10) = \$65.14$. Alternatively, the risk premium implies a price of $\$9.11(1.071501) = \$9.76$ and expected profit $(\$9.76 - \$9.11)100 = \$65$ (difference due to rounding). The total price for 100 policies sold at risk premium r_z is \$976, with expected profit \$65 (6.7% of total). Under the normal approximation, the probability of making a profit should be about 99%.

The pessimist might load the standard error, preferring

$$\sigma^* = (2.80)\{(0.1788)(9.11)/(1.2423)\} = 3.671279.$$

This generates $r^* = 0.093750$ and price \$9.96. Expected profit is $(2.326342)(\$3.671279)(10) = \$85.41$, or alternatively $(\$9.96 - \$9.11)100 = \$85$. This represents about 8.5% of the \$996 total price.

Summary

The pessimist's premium framework provides a flexible measure of right-tail risk that can be generalized to reflect diverse utility preferences. The right-tail index is a special case in which the weights of T^* are based on integrals of $[S(x)]^{1/2}$.

One final link between the two methodologies can be drawn from the parameter risk section of Mr. Wang's paper. The normal approximation presented above for multiple policies leads to a risk premium formula $\mathrm{E}[X] + r_z \mathrm{E}[X]$, which can be written

$$\mathrm{E}[X] + r_z \mathrm{E}[X] = \mathrm{E}[X] + \frac{z}{n^{1/2}} \frac{\sigma}{\mu} \mathrm{E}[X].$$

Approximating (σ/μ) by $d[X] = D[X]/\mathrm{E}[X]$ yields

$$\mathrm{E}[X] + \frac{z}{n^{1/2}} \frac{D[X]}{\mathrm{E}[X]} \mathrm{E}[X] = \mathrm{E}[X] + \beta D[X].$$

This indicates that the right-tail index may be especially useful in situations in which normal approximations are reasonable, as is often the case with a large number of finite-variance policies. Environments in which the number of policies is small or tail probabilities are excessive might be more appropriately addressed through explicit utility functions or the flexible forms r^*.

"BONUS-MALUS SYSTEMS: THE EUROPEAN AND ASIAN APPROACH TO MERIT RATING," JEAN LEMAIRE, JANUARY 1998

LIVIANA PICECH,* PATRIZIA GIGANTE,† AND LUCIANO SIGALOTTI‡

This discussion on the paper by Jean Lemaire aims to point out some comments on merit-rating systems in automobile liability insurance, inspired by recent Italian experience. After the deregulation originating from the application of the Third European Economic Community Directive, Italian insurance companies began to offer a wide supply of different tariff structures based on increasingly personalized rating systems. In particular, the insurers are now using more *a priori* classification variables and various bonus-malus systems (BMSs) are emerging. Moreover, some merit-rating systems recently appearing in the market combine bonus-malus and deductibles: the policyholders are spread into classes with different premium coefficients and different deductible levels. In this way a further element of personalization is introduced.

Some of the systems recently proposed by the Italian companies have been evaluated and compared by Gigante, Picech, and Sigalotti (1996) by analyzing: the distribution of the policyholders among the bonus-malus classes, the average premium level, and the equilibrium premium and their time development, which allows us to study the dynamics of the system along the periods preceding the steady-state.

We want to give here a brief illustration of the comparison between a specific BMS with deductible (which we call system A) and the BMS in use in Italy since before the deregulation (system B). To model the number of losses of an individual policyholder, we consider both a Poisson-gamma process and a Markov chain process. For the evaluation of the above-mentioned quantities, we use recursive procedures developed in Sigalotti (1994) and Gigante (1997). This approach also allows us to deal with open insurance portfolios; in this way we can study the impact of new policyholders on the financial stability of the system.

System A. The BMS with deductible is organized as follows:

Rules: • starting class 15

$$\bullet\ j = \begin{cases} max(h - 1, 1) & r = 0 \\ min(21, h + 3r - 1) & r = 1, 2, 3 \end{cases}$$

where h indicates the initial class, j the final class, and r the number of claims incurred in one year exceeding the deductible

• Table 1 reports the deductible levels (expressed as a percentage of each class premium) and the premium scale.

System B. The BMS in use in Italy since before the deregulation is organized as follows:

Rules: • starting class 14

$$\bullet\ j = \begin{cases} max(h - 1, 1) & n = 0 \\ min(18, h + 3n - 1) & n = 1, 2, 3, 4 \end{cases}$$

where h indicates the initial class, j the final class, and n the number of claims in one year

• Table 2 reports the premium scale.

Tables 3, 4, 5, and 6 show the distributions of the policyholders among the classes of the two systems for

*Liviana Picech is a Researcher in the Dipartimento di Matematica Applicata alle Scienze Economiche Statistiche ed Attuariali: "Bruno de Finetti," Università di Trieste, Piazzale Europa 10-34127 Trieste, Italy, e-mail, livianap@econ.univ.trieste.it.

†Patrizia Gigante is a Researcher in the Dipartimento di Matematica Applicata alle Scienze Economiche Statistiche ed Attuariali: "Bruno de Finetti," Università di Trieste, Piazzale Europa 10-34127 Trieste, Italy, e-mail, patriziag@econ.univ.trieste.it.

‡Luciano Sigalotti is a Professor in the Dipartimento di Finanza dell'Impresa e dei Mercati Finanziari, Università di Udine, Via Tomadini 30/A-033100 Udine, Italy, e-mail, luciano.sigalotti@dfimf.uniud.it.

"An Actuarial Index of the Right-Tail Risk," Shaun Wang, April 1998

BENJAMIN W. WURZBURGER*

In his paper Dr. Wang defines H[X] as the "certainty equivalent to risk X, or the price for transferring the risk X to other parties," then proposes five axioms for this H function and uses this framework (supplemented by the assumption that a particular power function is a square root function) to generate a measure of the right-tail thickness. He then ranks several continuous probability distributions according to this measure, and reports that this metric, unlike the Gini index for example, is in accord with the commonly perceived notion of tail thickness. Dr. Wang also deals nicely with several other important topics; his paper is quite informative.

It is important to note that Dr. Wang's notion of certainty equivalence is very different from the standard notion of certainty equivalence as applied by, say, Longley-Cook (1998) or Gerber and Pafumi (1998). The canonical formulation relies on the von Neumann-Morgenstern (1953) axiomatization (henceforth vN-M) about preferences; Dr. Wang's formulation relies however on the Yaari (1987) axiomatization, which explicitly deviates from the vN-M. Accordingly, the article's results—despite its title—should not be adopted by actuaries who would like to apply the canonical vN-M expected-utility risk theory to the important problem of evaluating the right-tail risk of low-frequency/large-loss events. For those of us conventional thinkers who prefer to operate under the vN-M paradigm, the big unanswered question is whether Dr. Wang's interesting results, such as on the ranking of various probability distributions, can be extended to the vN-M framework.

Dr. Wang's article does not mention this Wang/vN-M distinction nor does it mention vN-M. In my discussion, I'll be citing two articles by Wang and coauthors (Wang et al. 1997 and Wang and Young 1998), which note that they are extending Yaari and deviating from vN-M. I'll often characterize that latter system as W-Y. These latter two articles are quite abstract and mathematically sophisticated.

It is Wang's Axiom 3 (page 92) that represents the key departure of the W-Y system from the vN-M. In Section 1, I review this axiom and provide an example of its implications. Section 2 presents the competing vN-M axiom and discusses its relation to Wang Axiom 3 and to risk neutrality. Section 3 comments on some of the relative merits of W-Y versus vN-M.

1. Wang-Yaari Axiom 3 and an Example of Its Implications

According to Wang Axiom 3, "If X and Y are comonotonic, then $H[X + Y] = H[X] + H[Y]$." In his notation, X and Y refer to contingent liability risks which have non-negative payouts. As mentioned earlier, H is the certainty equivalent, namely how much an insurance company would be willing to pay in order to eliminate the exposure. As far as comonotonicity, "two risks are defined to be comonotonic if there exists a random variable Z and nondecreasing real functions u and v such that $X = u(Z)$ and $Y = v(Z)$ with probability one." Accordingly, a sufficient but not necessary condition for X and Y to be comonotonic is that X and Y have a correlation of 1. In the forthcoming example, we will consider a special case of comonotonicity, the case where X and Y are identical.

An Example of the Implications of Axiom 3:

Suppose an insurance company faces a contingent liability to pay \$1 billion in case of a (specified-strength) hurricane in Florida in 1999, an event that has a 1% probability. Because of risk aversion, the insurance company would presumably be willing to pay more than \$10 million to unload this risk; denote this amount as $10 + m$ (in units of \$million).

Now, suppose the same insurance company faced a contingent liability to pay out \$2 billion in case of the aforementioned hurricane. Per Wang's Axiom 3, applied to the case where X and Y are identical, the certainty equivalent of this contingent liability would be $20 + 2m$. Such a property is however implausible; because of risk aversion, the insurer should be willing to pay more than $20 + 2m$ in order to unload this risk. (It is because of that risk aversion that insurers often engage in risk sharing and partial reinsurance. I'll further discuss this issue in Section 3, paragraph 2.)

*Benjamin W. Wurzburger, Ph.D., is Associate Investment Policy Officer at John Hancock Mutual Life Insurance Company, John Hancock Place, P. O. Box 111, Boston, Massachusetts 02117, e-mail, bwurzburger@jhancock.com.

2. The Competing von Neumann-Morgenstern Independence Axiom and its Relationship to Wang-Yaari Axiom 3 and to Risk Neutrality

Let P denote the space of probabilities of possible outcomes, and let $>$ denote a preference relation on P (see Huang and Litzenberger 1998, page 7.) then the vN-M independence axiom, commonly called the substitution axiom, runs as follows:

Independence Axiom: For all p, q, $r \in p$ and $\lambda \in (0, 1]$, $p > q$ implies $\lambda p + (1 - \lambda)r > \lambda q + (1 - \lambda)r$.

Following the Huang and Litzenberger suggestion in the consumption context, think of p, q, r as lotteries and $\lambda p + (1 - \lambda)r$ as a compound lottery; according to the independence axiom, "satisfaction of consumption in a given event does not depend on what the consumption would have been if another event had occurred." On the face of it, this assumption appears to be logically compelling and innocuous, as the counterfactual "would have been" is irrelevant for ordering preferences on what actually occurs.

Under the independence axiom, in conjunction with some other very innocuous axioms, textbooks demonstrate that risk-averse preferences can be represented in terms of an expected utility function, $\Sigma\, u(x)p(x)$, where u is a concave utility function, and x is the payout variable with a p.d.f. of $p(x)$.

Comparison with Risk Neutrality and with Wang-Yaari

According to the risk neutrality assumption, the certainty equivalence (or expected utility) is equal to $\Sigma x\, p(x)$, that is, it is linear in both the dollar payouts and the probabilities. According to vN-M, the expected utility is, as indicated above, linear in the probabilities but concave in the dollars x. According to W-Y, the certainty equivalence H function is linear in x (see Wang's article, page 92), but it is not linear in the probabilities. (Wang uses a concave square root power function; more generally, as in Wang and Young 1997 [p. 178], "market prices are just expectations with respect to a new measure; however this measure is not necessarily additive.") It is because of this interchange of the roles of dollars and probabilities that Yaari describes his framework as a dual theory relative to the standard theory.

Numerical Illustration in Terms of Our Hurricane Example

As discussed in Section 1, according to W-Y the certainty equivalent of the 1% probability of paying out \$2 billion will be twice the certainty equivalent of a 1% probability of paying out \$1 billion. According to vN-M, the certainty equivalent of a 2% probability of paying out \$1 billion will be approximately twice the certainty equivalent of a 1% probability of paying out \$1 billion. Under risk neutrality, both of the above statements will be valid, as the certainty equivalent will be the product of the payout and the probability.

Please note that in the above paragraph the qualifier "approximately" arises because a finite probability of a large loss generates a finite "wealth" or "income" effect, which would impact the absolute risk aversion. The vN-M axiomatization implies that, in the limit, as the probability P approaches zero, the ratio of certainty equivalents—the certainty equivalent of a $2P$ probability event paying out \$1 billion divided by the certainty equivalent of a P probability event paying out \$1 billion—will converge onto 2.

3. Wang-Yaari Versus von Neumann-Morgenstern: A Comparison of Some of Their Respective Merits

Are the Axioms Self-Evident?

The von Neumann independence axiom is an "intuitively obvious" restriction on the structure of preferences, that the realized satisfaction or utility depends solely on what has occurred and not on what might have happened but didn't. The Wang Axioms (Wang Axiom 3 and especially the related Wang Axiom 5) impose nonintuitive restrictions on a more complicated construct, namely certainty equivalence/insurance prices. Axiom 5, an axiom that relates to "the reduction of compound Bernoulli risks," is especially elusive. That axiom states: "Let $Y = BX$ be a compound Bernoulli risk, where the Bernoulli frequency random variable B is independent of the loss severity random variable $X = Y \mid Y > 0$. Then the market prices for risks $Y = BX$ and $BH[X]$ are equal."

While Wang and Young (1997) motivate Wang's Axioms 3 and 5 via *no arbitrage* arguments, their system does not in fact preclude arbitrage (see Wang and Young, page 182, where they note that it is only for the case of comonotonic risks that they preclude arbitrage). The vN-M axiomatization, on the other hand, does not deal with prices; nevertheless, it readily lends itself to be embedded in a system which fully satisfies the no-riskless-arbitrage condition. For this embedding, see any good textbook. Along these lines, Gerber and Pafumi (1998, Section 10) present a competitive-equilibrium, no-arbitrage framework and cite both Borch and Bühlmann.

How Do People and Institutions Actually Behave?

The vN-M axioms are generally regarded as self-evident, and the expected utility representation, which is implied by these axioms, has become standard. Nevertheless, it has long been recognized that actual behaviour is often inconsistent with the vN-M axioms. Of these axioms, it is the independence axiom—a VN-M axiom which is absent from the W-Y system—that is most often violated in empirical experiments. The best known of these violations is the Allais Paradox (see Allais 1953). Because of its commitment to the expected-utility paradigm, the literature has tended to describe such violations as irrational.

The Hurricane Example

Suppose we were to ask senior insurance executives how much they would be willing to pay to avoid a 1% probability of a $1 billion loss and another set of executives how much they would be willing to pay in order to avoid a 2% probability of a $1 billion loss. I suspect that the answers provided in the second case would run well below twice the value of the answers in the first case, a seeming violation of the vN-M axioms. (I suspect that, in practice, senior executives may not really believe a reassurance that a certain event has only a 1% probability. Also, many insurance executives may be psychologically uncomfortable facing material ambiguity and uncertainty, and even a 1% probability may be sufficient to trigger this discomfort. A 10%/5% experiment would also be quite interesting, as the 1% probability may conceivably lie below a critical threshold value for concern.) This example further illustrates that the vN-M axioms are often more *prescriptive* of what is "rational" behaviour than *descriptive* of actual behaviour.

The Overriding Consideration, Concluding Remarks

The overriding consideration is that the various sciences—actuarial risk theory, economics, and finance—have all adopted the vN-M paradigm. Wang and Young (1998) suggest that it is on account of its relative newness that Yaari's theory has not been widely applied in insurance economics. However, it's already been a decade since Yaari's article, and I conjecture that a decade from now vN-M will still dominate.

By developing and extending a competing theory, Yaari (1987) and subsequently Dr. Wang (along with his coauthors) have certainly helped advance our understanding of the textbook vN-M. And, who knows, maybe my conjecture will prove wrong—and the Wang-Yaari axiomatization will someday supplant the von Neumann-Morgenstern.

ACKNOWLEDGMENT

The discussant thanks Associate Editor Robert R. Reitano for his valuable suggestions.

REFERENCES

ALLAIS, M. 1953. "Le comportement de l'homme rationnel devant le risque, critique des postulates et axioms de l'ecole Americaine," *Econometrica* 21:503–46.

GERBER, H.U., AND PAFUMI, G. 1998. "Utility Functions: From Risk Theory to Finance," *North American Actuarial Journal* 2(3):87–100.

HUANG, C., AND LITZENBERGER, R.H. 1988. *Foundations for Financial Economics.* New York: North-Holland.

LONGLEY-COOK, A.G. 1998. "Risk-Adjusted Economic Value Analysis," *North American Actuarial Journal* 2(1):74–100.

VON NEUMANN, J., AND MORGENSTERN, O. 1953. *Theory of Games and Economic Behavior.* Princeton, N.J.: Princeton University Press.

WANG, S., YOUNG, V.R., AND PANJER, H.H. 1997. "Axiomatic Characterization of Insurance Prices," *Insurance: Mathematics and Economics* 21:173–83.

WANG, S., AND YOUNG, V.R. 1998. "Ordering of Risks: Utility Theory Versus Yaari's Dual Theory of Risk," *Insurance: Mathematics and Economics* 22:145–61.

YAARI, M.E. 1987. "The Dual Theory of Choice under Risk," *Econometrica* 55:95–115.

AUTHOR'S REPLY

SHAUN WANG*

I would like to thank Mr. Berberian [*NAAJ* Vol. 2, no. 4 (October 1998:140)] and Dr. Wurzburger for their discussions of my April 1998 NAAJ paper "An Actuarial Index of the Right-Tail Risk." Mr. Berberian proposed an interesting way of constructing a summary

*Shaun Wang, A.S.A., Ph.D., is an Associate Actuary at the SCOR Reinsurance Company, One Pierce Place, Itasca, Illinois 60143-4049, e-mail, swang@scor.com.

statistics of the right-tail risk. Dr. Wurzburger's discussion is at a more theoretical level of economic theory; his discussion is also broader in commenting on some related writings by me and coauthors Virginia Young and Harry Panjer. In response to their discussions, I would like to comment on some properties of the PH-transform and underlying axioms, from which the right-tail index was derived.

An Important Fact About the PH-Transform

Consider a real-valued random variable X with decumulative distribution function: $\mathrm{S}(u) = 1 - \mathrm{F}(u) = \Pr\{\mathrm{X} > u\}$, for $-\infty < u < \infty$. The PH-mean of X with (positive) index r is defined as the mean value after applying a proportional hazards transform:

$$\mathrm{H}_r[X] = \int_{-\infty}^{0} [S(u)^r - 1]\, du + \int_{0}^{\infty} S(u)^r\, du.$$

In general, we have: (i) $\mathrm{E}[X] \leq \mathrm{H}_r[X] \leq \max[X]$ when $0 < r < 1$; (ii) $\mathrm{H}_r[X] = \mathrm{E}[X]$ when $r = 1$; and (iii) $\mathrm{H}_r[aX + b] = a\mathrm{H}_r[X] + b$ for all positive numbers a and real numbers b.

The normal distributions form a location-scale family. For any $X \sim \mathrm{N}(\mu, \sigma^2)$, we have $X = \mu + \sigma \cdot Z$ with $Z \sim \mathrm{N}(0, 1)$. Because of the aforementioned third property of PH-mean, for any $X \sim \mathrm{N}(\mu, \sigma^2)$ we have $\mathrm{H}_r[X] = \mu + \beta \cdot \sigma = \mathrm{E}[X] + \beta \cdot \mathrm{SD}[X]$. The constant β depends on r:

$$\beta = \mathrm{H}_r[Z] = \int_{-\infty}^{\infty} [\Phi^c(u)^r - 1]du + \int_{-\infty}^{\infty} [\Phi^c(u)]^r\, du,$$

where $\Phi^c = 1 - \Phi$ is the standard normal decumulative distribution. In other words, *for normal distributions, the PH-transform loading is equivalent to the standard deviation loading:* $\mathrm{H}_r[X] = \mathrm{E}[X] + \beta \cdot \mathrm{SD}[X]$. The value of β can be calculated numerically for any given r (see the table below).

Index r	β
0.50	0.7041
0.55	0.5971
0.60	0.5025
0.65	0.4179
0.70	0.3417
0.75	0.2724
0.80	0.209
0.85	0.1507
0.90	0.0968
0.95	0.0467

This interesting property of the PH-transform was mentioned by Todd Bault in his presentation at the 1998 CAS Reinsurance Seminar on Capital Allocations. Independently, Stavros Christofides made this observation in his prize-winning paper "Pricing for Risk in Financial Transactions," presented at the 1998 ASTIN meeting. In that paper, he investigated the risk/reward trade-off using the PH-transform of empirical loss distributions in various financial transactions. He also demonstrated by numerical example that the PH-loading is equivalent to the standard deviation principle for extreme value distributions with cdf: $F(u) = \exp\{-\exp[-(u - a)/b]\}$, for $-\infty < u < \infty$. This should become apparent by noting that the extreme value distributions form a location-scale family.

For normal distributions, the standard deviation method is a well-accepted loading principle. The PH-transform loading preserves the tradition of standard deviation loading for normal distributions, while it also reflects the tail risk for skewed loss distributions. In contrast, the standard deviation loading does not account for the skewness of most distributions encountered in actuarial science.

In my "actuarial index" paper, I defined the right-tail deviation as $\mathrm{RD}[X] = \mathrm{H}_{0.5}[X] - E[X]$. For normal distributions, the right-tail deviation is simply a multiple of the standard deviation: $\mathrm{RD}[X] = 0.7041 \cdot SD[X]$.

Some Comments on Our Axiom 3 (Wang, Young, and Panjer 1997)

In his discussion, Dr. Wurzburger used the example of an insurance company facing a contingent liability to pay $1 billion in case of a (specified-strength) hurricane in Florida in 1999, an event that has a 1% probability. I would like to offer my viewpoint using this same example. While the expected-utility theory of von Neumann and Morgenstern explains an individual's behavior related to risk aversion, our Axiom 3 offers a different perspective of the same risk-sharing process. Assuming that a $2 billion loss can be shared proportionally among twice as many reinsurers, the "market price" for this potential $2 billion-dollar loss should be twice as much as that for a $1 billion-dollar potential loss. Our Axiom 3 offers a different perspective on the same risk-sharing process. The expected-utility theory motivates how individuals would want risk sharing, while our Axiom 3 tries to capture the market prices after risk-sharing schemes are already in place.

The right-tail deviation, like the standard deviation, is a tool for measuring risk. The final result depends upon not only the tool being employed, but also *how* the tool was applied. The subadditivity of the right-tail deviation can be used to explain the risk reduction by pooling independent risks, provided we apply this risk measure to the aggregate risk portfolio. From a risk portfolio's perspective, the PH-mean can be used to explain the risk-averse behavior of an individual entity. Doubling the loss dollar amount is equivalent to covering two perfectly correlated risks, which would result in a higher aggregate PH-mean than that for covering two independent and identically distributed risks.

A Justification for Our Axiom 5 (Wang, Young, and Panjer 1997)

Dr. Wurzburger commented that our Axiom 5 regarding compound Bernoulli risks is not quite obvious to many readers. To assist in understanding Axiom 5, I offer the following practical situation in increased-limits ratemaking. In liability insurance, manual rates are available for basic-limit policies and increased limits can be rated using the ISO published increased-limits factors (ILFs). For example, assume that the basic \$100,000 limit policy has a pure premium (expected loss plus risk charge, excluding expenses) of \$2,000. If the increased limits factor for a \$500,000 limit policy is 1.4, then the pure premium for a \$500,000 policy can be calculated as \$2,800 = \$2,000*1.4.

Assume that the claim frequency can be modeled by a Bernoulli(q) variable B. Let X_1 represent the basic limit severity and X_2 represent the increased limit severity. Then both the basic limit risk and an increased limit risk are compound Bernoulli risks, $B \cdot X_1$ and $B \cdot X_2$ respectively. Our Axiom 3 would imply that the ILF should not depend on the frequency q:

$$\mathrm{ILF} = \frac{\mathrm{H}[B \cdot X_2]}{\mathrm{H}[B \cdot X_2]} = \frac{\mathrm{H}[B] \cdot \mathrm{H}[X_2]}{\mathrm{H}[B] \cdot \mathrm{H}[X_1]} = \frac{\mathrm{H}[X_2]}{\mathrm{H}[X_1]}.$$

I believe that our Axiom 3 is consistent with ISO practice in that only the severity curve was used in developing the ILFs. This is intuitive because an ILF indicates only the relativity between the basic-limit premium and the increased-limit premium. The ground-up frequency should not affect this relativity.

Let me construct another man-made practical situation (insurance is man-made practice anyway): Suppose that there is an aggregate CAT (catastrophe) loss index for a well defined risk-portfolio and there is an active market of exchange based on this CAT index. Suppose the current index is at \$8,000 million as of 7/1/1999 and the CAT index follows a Poisson-Pareto (frequency-severity) process. Consider the following three contracts:

1. Contract E covers the excess of \$8,500 million as of 1/1/2001.

2. As of 7/1/1999, there is a zero-cost futures Contract F: one party will settle the loss in excess of \$8,500 million (which is random) in exchange for a fixed premium of \$30 million. However, this contingent cash flow exchange will take place on 1/1/2001 only if the CAT index exceeds \$8,500 million on 1/1/2001.

3. Contract G promises to pay a predetermined \$30 million in the event that the CAT index exceeds \$8,500 million as of 1/1/2001.

Using the notations of our Axiom 5, Contract E defines a compound Bernoulli risk $Y = B \cdot X$, where B has Bernoulli(q) frequency (q is the probability that the CAT index will exceed \$8,500 on 1/1/2001). Contract F implies that, as of 7/1/1999, the (conditional) excess loss severity X has a market value (conditional certainty-equivalent) of \$30 million, that is, $\mathrm{H}[X] = 30$. Contract G defines a Bernoulli risk $B \cdot \mathrm{H}[X]$. Then our Axiom 5 says that the market premium for Contract E should be the same for Contract G, that is, $\mathrm{H}[B \cdot X] = \mathrm{H}[B \cdot \mathrm{H}[X]] = \mathrm{H}[B] \cdot \mathrm{H}[X]$.

I hope the above two examples may assist in understanding our Axiom 5 regarding valuation of compound Bernoulli risks.

Any measure of risk is somewhat subjective and reflects the individual's perception of the risk. On the other hand, it would be useful to have some simple and specific summary statistics regarding the right-tail risk. In this regard, the square-root function seems to be a good candidate. By fixing the index $r = 0.5$ we get a numeric metric which gives a complete ordering of loss distributions. More importantly, this complete ordering seems to agree with our intuitive perception regarding the ordering of tail thickness among commonly used distributions.

Motivated by the general use of Gini index in measuring income distributions, I chose the square root function to create a right-tail index. However, by no means am I advocating the use of $r = 0.5$ in deciding risk-loads. In fact, some actuaries might feel that the square-root function seems to produce excessive risk loads for high layers; instead, a less severe index (between 0.7 and 0.9) might produce more reasonable risk loading at extreme tails.

In conclusion, the PH-transform methodology parallels the well established paradigm of expected utility. It offers a different perspective of the observed risk-sharing process. I hope that this new perspective, in conjunction with the expected utility theory, will help to draw a better picture of the theory of risk.

"ECONOMIC VALUATION MODELS FOR INSURERS," DAVID F. BABBEL AND CRAIG MERRILL, JULY 1998

IRWIN T. VANDERHOOF*

Babbel and Merrill have made so many contributions to the actuarial literature that I wish we could award them honorary fellowships. This is another fine example of their work.

The paper separates nicely into sections on insurance liabilities, criteria for such models, and a taxonomy of these models with a detailing of the choices available in modeling such liabilities. I believe that this is a special benefit for those who may be new to the wild world of interest rate generators and stochastic parameters.

I do not have criticisms of the material in the paper, but I do have some extensions of the ideas presented. They have to do with the objective of valuations and some of the details of the methodology.

The paper is set around methods of calculating the present value, or market value, or fair value of insurance liabilities. This would follow the pattern on Wall Street of focusing on market values. Mean variance analysis in finance extends beyond the first moment of value to the second moment. While this provides useful information on the riskiness of the total portfolio, its usefulness depends upon the actual probability distribution of results.

I believe that while present value is the only right place to start, it may be an inadequate place for the actuary to end. It would seem to me that while the accountant may be explicitly interested only in expected value of a stream of payments, the actuary must focus on the additional factor of the risk of ruin of the enterprise. Since this depends upon the balance of payments of cash flows into and out of the organization, we would need probability distributions of these cash flows of both assets and liabilities to provide information on safety and solvency. These are the particular responsibilities of the actuary as opposed to the financial consultant.

Such distributions of all cash flows are required by regulators for both assets and liabilities for the New York Seven and will eventually be required for a larger universe of interest rate scenarios.

This might seem like an impossible task. Unless all the answers were available from closed formulas we could have millions of simulations to do. However, it just so happens that Graham Lord and I have been working with Dan Vesper of the ARM group on experiments with what we call adaptive resampling (AR). Graham and I believe that this technique is original with us.

It works this way. Each of the stochastic variables involved has some probability distribution. We can characterize each as having values of the variable that correspond to each possible value of the cumulative distribution function. Since every CDF has values that run from 0 to 1, we can form a unit hypercube where each edge is the value of one of these CDF's. Because we can have a dimension for every day, week, and so on for each variable, every point within the hypercube represents a complete scenario of values for the variables. We can then do a Monte-Carlo simulation with variables running from 0 to 1 and then translating the CDF's into the corresponding values of the variables. If we can identify values of the final results that are of interest—such as insolvency—we can then resample from only that volume of the hypercube to get a more accurate value on the risk of insolvency or whatever other value level is of particular interest. A great advantage of this technique is that the volume of the probability space being resampled is known, and that volume is the probability of the occurrence of values of the result corresponding to that volume.

We have tried this in cases where the results could be easily checked, such as the sum of a series of normal distributions, with good results. We have also had quite satisfactory results with an actual portfolio of assets and SPDA liabilities. We seem to be getting reasonable approximations of ruin probabilities with 100 to 200 simulations in cases where there are 80 dimensions. Steve Craighead has recently reported similar results in his experiments.

The above results are based on the use of low discrepancy point technology.

One of the criteria set in the paper for a viable model is close calibration with observed market prices. I view this as crucial. I believe that it is nonsense to think that we can form a distribution of possible futures. Fortunately we do not need to. The market for default-free interest rate investments is near

*Irwin T. Vanderhoof, F.S.A., A.C.A.S., A.I.A., Ph.D., is Clinical Professor at the Stern School of Business at New York University. His address is 18 Two Bridges Road, Towaco, New Jersey 07082.