ROE, Utility, and the Pricing of Risk

Leigh J. Halliwell

Abstract

Actuaries, economists, underwriters, and regulators are rightly convinced that the task of pricing insurance should depend on the uncertainty of the amount and the timing of the insured losses, as well as on the correlation of those losses with those already insured. The most common approach to pricing is to allocate equity to the insurance transaction and to achieve a certain expected return on that equity. This paper will argue that this ROE approach is illogical and inconsistent, and that the proper approach is to accept no price that reduces the expected utility of the insurer. The utility-theoretic approach will enable logical and consistent pricing, and will involve no unusual practical difficulties. The paper will argue also that capital should play a role in assessing the strength of an insurer, but capital allocation should play no role in an insurer's pricing.

1) Return on Allocated Capital

A typical finance problem is the following: A company is considering entry into the widget business, which entails the purchase of a machine to produce widgets. The company estimates that the machine will last five years, and that the profits from the sale of its widgets over those five years will be \$100,000, \$125,000, \$125,000, \$100,000, and \$75,000. If the beta of the widget industry is 1.20, what is the maximum price that the company should pay for the machine?

The profits are *expected to be* those five amounts. But the profit of the first year, almost surely, will not be \$100,000. If the estimate is unbiased, the profit will be \$100,000 $\pm\ \sigma$. But the variance of the profit is ignored. The rationale for ignoring it is that the "market" accounts for it in the beta of the widget industry. If the risk-free rate of return is five percent per year, and the expected rate of return on the universe of risky assets is ten, then the Capital Asset Pricing Model (CAPM) prescribes the expected rate of return for the widget industry to be $r_f + \beta(E[R_m] - r_f) = 0.05 + 1.2(0.10 - 0.05) = 0.11$, or eleven percent per year. Therefore, the company is justified in purchasing the machine if it costs no more than $\frac{\$100{,}000}{(1.11)^1} + \frac{\$125{,}000}{(1.11)^2} + \frac{\$125{,}000}{(1.11)^3} + \frac{\$100{,}000}{(1.11)^4} + \frac{\$75{,}000}{(1.11)^5} = \$403{,}000$. It might then be said that the company is allocating \$403,000 of capital to its widget business and is expecting a return on that capital of eleven percent per year.

The author distrusts this ROE approach to finance because it makes the CAPM (or some other rate-of-return model) the panacea for variance. Planners can justify a deterministic projection of the cash flow by claiming that some risk-adjusted rate of return on the expected cash flow takes into account its variance, and that the CAPM or some other model yields the proper rate. This approach also makes time to be of the essence of pricing risk. So the proper number is not a return (e.g., eleven percent), but rather a *rate* of return (e.g., eleven percent *per year*). That this is illogical and makes for inconsistent pricing will be demonstrated in the next two sections and in Appendices A and E. But for now one need only ponder how to price instantaneous cash flows, such as the flipping of a coin and the receipt of one dollar if the coin lands heads.

2) Return on Capital Allocated to the Insurance Business

If the ROE approach to pricing risk is problematic enough for the widget business, it is even more problematic for the insurance business. For the widget business behaves as a typical investment in that money first goes out with the desirable expectation that later it will come in. But the insurance business, at least superficially, behaves in reverse: money (in the form of premiums) first comes in with the undesirable expectation that later it will go out (in the form of losses). So if a company were considering to insure a risk whose losses were expected to be $100,000, $125,000, $125,000, $100,000, and $75,000 over five years, and the beta for the insurance industry were 1.20, it would not accept a pure premium of $403,000. In fact, it would not even accept a pure premium of these amounts discounted at the risk-free rate of return:

$$\frac{\$100{,}000}{(1.05)^1}+\frac{\$125{,}000}{(1.05)^2}+\frac{\$125{,}000}{(1.05)^3}+\frac{\$100{,}000}{(1.05)^4}+\frac{\$75{,}000}{(1.05)^5}=\$457{,}631$$

This has led Robert Butsic [4] and others to modify the CAPM formula to something like:

$$r_f-\beta\left(E[r_m]-r_f\right)=0.05-1.2(0.10-0.05)=-0.01$$

This would make for a risk-adjusted pure premium of:

$$\frac{\$100{,}000}{(0.99)^1}+\frac{\$125{,}000}{(0.99)^2}+\frac{\$125{,}000}{(0.99)^3}+\frac{\$100{,}000}{(0.99)^4}+\frac{\$75{,}000}{(0.99)^5}=\$540{,}342$$

Such modifications of the rate of return move the pure premium in the right direction, viz., to being greater than the present value of the expected losses. However, it seems unbecoming of a theory that it should have a plus sign when returns are desirable and a minus sign when they are undesirable. Moreover, time is still of the essence of this pricing. What would the risk-adjusted pure premium be to insure a one-dollar loss that is incurred if a coin to be flipped in the next moment lands heads?

The sensing of these problems has led many to a more sophisticated concept of the insurance business. The insurance *cash flow* is opposite to that of a typical investment; but the order of the *capital* or *equity flow* is typical, i.e., an outflow followed by an inflow. Although the insurer first receives the premium, it must set up reserves and pay expenses. Usually this entails that the insurer's net worth or equity decreases immediately after writing the policy, any eventual increase depending on how much loss is incurred and how much investment income is earned. But this is not the whole story. The insurer must convince prospective clients, as well as regulators and rating agencies, that it has sufficient capital on hand to cover worse-than-expected losses. Thus, every insured risk places some burden on the insurer's capital. Hence, many have come to conceive of the financing of an insurance transaction as the allocation of some amount of capital on which the profits of the transaction are returns.

The author believes that this more sophisticated ROE approach is no less problematic. For no suggested method for allocating capital has been widely accepted, nor is it likely that one will be (cf. [1] and [13]; also Appendix D). And even if some method were devised, there would remain the problem of determining what expected rate of return the allocated capital should earn. Also, the CAPM affords financial planners the convenience of a deterministic framework; but the allocation of capital to insurance risks requires the taking into account of variances and covariances.

Moreover, the expected rate of return on allocated capital seems to depend on the amount of allocated capital. For suppose that companies A and B bid for the same risk, that both believe the expected losses to be $140, and that their expense loads are similar. But A proposes to allocate $100 of its capital to the risk for one year, at a cost of ten percent per year. Thus A's premium will include a ten-dollar risk charge. Company B, which boasts of its disaster-proof security, proposes to allocate $200 for one year, at the same cost of ten percent. So B's premium will include a twenty-dollar risk charge. Suppose that A quotes a premium of $200 and B quotes $210. Company A argues that its capital allocation corresponds to a 97.5% confidence level, whereas company B lays claim to a 99.0% confidence level. It is questionable whether the prospective client will consider the additional 1.5% confidence level to be worth the additional ten dollars, especially if a guaranty fund backs up the companies. Company B will be hard pressed either to allocate less capital or to charge less than ten percent per year for that capital.

And finally, time continues to be of the essence of this more-sophisticated ROE approach. If capital c is allocated to an instantaneous risk, such as the coin flip, at a cost of capital of κ per time period, then the risk charge will be $c\kappa\Delta t \to 0$. Risk loads evanesce in the realm of the instantaneous, which is inconsistent with the fact that a lot of money can be lost in an instant. The next section will reinforce this critique of the ROE approach with a simple mathematical example.

3) An Example of the Inconsistency of the ROE Pricing Approach

We will use the ROE approach to price the following risk: a policy will cover the time interval $[0, T]$. Claim generation is a Poisson process with a constant frequency of λ claims per time unit. The severity of every claim is s dollars. The risk-free yield curve is flat at a force of interest ρ per time unit. Somehow it is determined that c dollars of capital should be allocated to this risk and that this capital should earn a risk charge continuously at a rate of κ per time unit. It is clear that the example is simple, particularly in that the uniform Poisson process with constant severity justifies the constancy of the allocated capital. But if the ROE approach proves to be illogical and inconsistent on a simple level, how much more illogical and inconsistent must it be on a complex level?

The expected loss in the interval $[t, t+dt]$ is $s\lambda dt$. And the present value of the loss in that interval is $e^{-\rho t} s\lambda dt$. Therefore, the present value of the expected loss is:

$$
\begin{aligned}
PV[E[L]] &= \int_{t=0}^{T} e^{-\rho t} s\lambda dt \\
&= s\lambda \frac{1}{\rho} \int_{t=0}^{T} e^{-\rho t} \rho dt \\
&= s\lambda \frac{1}{\rho}\left(e^{-\rho t}\Big|_{T}^{0} \right) \\
&= s\lambda \frac{1-e^{-\rho T}}{\rho} \\
&= s\lambda T \frac{1-e^{-\rho T}}{\rho T}
\end{aligned}
$$

L'Hospital's rule can be used to extend this equation to the interesting case of $\rho T = 0$:

$$
PV[E[L]] = \begin{cases} s\lambda \dfrac{1-e^{-\rho T}}{\rho} & \rho T \neq 0 \\ \\ s\lambda T & \rho T = 0 \end{cases}
$$

The allocated capital c will earn in the interval $[t, t+dt]$ a risk charge of $c\kappa dt$. The present value of the entire risk charge is:

$$
\begin{aligned}
RC &= \int_{t=0}^{T} e^{-\rho t} c\kappa dt \\
&= c\kappa \frac{1}{\rho} \int_{t=0}^{T} e^{-\rho t} \rho dt \\
&= c\kappa \frac{1-e^{-\rho T}}{\rho} \\
&= c\kappa T \frac{1-e^{-\rho T}}{\rho T}
\end{aligned}
$$

Again, L'Hospital's rule allows for the extension to the case of $\rho T = 0$:

$$RC = \begin{cases} c\kappa \dfrac{1 - e^{-\rho T}}{\rho} & \rho T \neq 0 \\ c\kappa T & \rho T = 0 \end{cases}$$

Summing these two, we have the formula for the risk-adjusted pure premium:

$$\begin{aligned} RAPP &= PV\left[E[L]\right] + RC \\ &= (s\lambda + c\kappa) \cdot \begin{cases} \dfrac{1 - e^{-\rho T}}{\rho} & \rho T \neq 0 \\ T & \rho T = 0 \end{cases} \end{aligned}$$

The formula possesses a very attractive property, viz., that the premium for a policy covering $[0, T_2]$ equals the sum of the premium for a policy covering $[0, T_1]$ and the present value of the premium for a policy covering $[T_1, T_2]$, where $0 \leq T_1 \leq T_2$:

$$RAPP[0,T_2]=(s\lambda+c\kappa)\cdot\begin{cases}\dfrac{1-e^{-\rho T_2}}{\rho} & \rho T_2\neq 0\\ T_2 & \rho T_2=0\end{cases}$$

$$=(s\lambda+c\kappa)\cdot\begin{cases}\dfrac{1-e^{-\rho T_1}+e^{-\rho T_1}-e^{-\rho T_2}}{\rho} & \rho T_2\neq 0\\ T_1+(T_2-T_1) & \rho T_2=0\end{cases}$$

$$=(s\lambda+c\kappa)\cdot\begin{cases}\dfrac{1-e^{-\rho T_1}}{\rho}+e^{-\rho T_1}\dfrac{\left(1-e^{-\rho(T_2-T_1)}\right)}{\rho} & \rho T_2\neq 0\\ T_1+e^{-0}(T_2-T_1) & \rho T_2=0\end{cases}$$

$$=(s\lambda+c\kappa)\cdot\begin{cases}\dfrac{1-e^{-\rho T_1}}{\rho} & \rho T_1\neq 0\\ T_1 & \rho T_1=0\end{cases}+e^{-\rho T_1}(\lambda s+c\kappa)\cdot\begin{cases}\dfrac{1-e^{-\rho(T_2-T_1)}}{\rho} & \rho(T_2-T_1)\neq 0\\ T_2-T_1 & \rho(T_2-T_1)=0\end{cases}$$

$$=RAPP[0,T_1]+e^{-\rho T_1}RAPP[T_1,T_2]$$

Now we will modify the example. Let the expected loss $E[L]=s\lambda T$ be constant regardless of T. This is easily accomplished by making the frequency inversely proportional to T: $\lambda(T)=\dfrac{\lambda_0}{T}=\dfrac{E[L]}{sT}$. If the allocated capital c remains constant, the formula for the risk-adjusted pure premium becomes:

$$RAPP = \left(s\lambda(T) + c\kappa\right) \cdot \begin{cases} \dfrac{1 - e^{-\rho T}}{\rho} & \rho T \neq 0 \\ T & \rho T = 0 \end{cases}$$

$$= \left(\frac{E[L]}{T} + c\kappa\frac{T}{T}\right) \cdot \begin{cases} \dfrac{1 - e^{-\rho T}}{\rho} & \rho T \neq 0 \\ T & \rho T = 0 \end{cases}$$

$$= \left(E[L] + c\kappa T\right) \cdot \begin{cases} \dfrac{1 - e^{-\rho T}}{\rho T} & \rho T \neq 0 \\ 1 & \rho T = 0 \end{cases}$$

But again, divide the coverage term into $[0, T_1]$ and $[T_1, T_2]$, and pro-rate the expected loss: $E[L_1] = \frac{T_1}{T_2} E[L] = w_1 E[L]$ and $E[L_2] = \frac{(T_2 - T_1)}{T_2} E[L] = w_2 E[L]$, where $w_1 + w_2 = 1$.

Then:

$$RAPP = (E[L] + c\kappa T_2) \cdot \begin{cases} \dfrac{1-e^{-\rho T_2}}{\rho T_2} & \rho T_2 \neq 0 \\ 1 & \rho T_2 = 0 \end{cases}$$

$$= (E[L] + c\kappa T_2) \cdot \begin{cases} \dfrac{1-e^{-\rho T_1}}{\rho T_2} + e^{-\rho T_1} \dfrac{1-e^{-\rho(T_2-T_1)}}{\rho T_2} & \rho T_2 \neq 0 \\ w_1 + w_2 & \rho T_2 = 0 \end{cases}$$

$$= (E[L] + c\kappa T_2) \cdot \begin{cases} w_1 \dfrac{1-e^{-\rho T_1}}{\rho T_1} + e^{-\rho T_1} w_2 \dfrac{1-e^{-\rho(T_2-T_1)}}{\rho(T_2-T_1)} & \rho T_2 \neq 0 \\ w_1 + e^{-0} w_2 & \rho T_2 = 0 \end{cases}$$

$$= w_1 (E[L] + c\kappa T_2) \cdot \begin{cases} \dfrac{1-e^{-\rho T_1}}{\rho T_1} & \rho T_1 \neq 0 \\ 1 & \rho T_1 = 0 \end{cases} + e^{-\rho T_1} w_2 (E[L] + c\kappa T_2) \cdot \begin{cases} \dfrac{1-e^{-\rho(T_2-T_1)}}{\rho(T_2-T_1)} & \rho(T_2-T_1) \neq 0 \\ 1 & \rho(T_2-T_1) = 0 \end{cases}$$

$$= (E[L_1] + c\kappa T_1) \cdot \begin{cases} \dfrac{1-e^{-\rho T_1}}{\rho T_1} & \rho T_1 \neq 0 \\ 1 & \rho T_1 = 0 \end{cases} + e^{-\rho T_1} (E[L_2] + c\kappa (T_2 - T_1)) \cdot \begin{cases} \dfrac{1-e^{-\rho(T_2-T_1)}}{\rho(T_2-T_1)} & \rho(T_2-T_1) \neq 0 \\ 1 & \rho(T_2-T_1) = 0 \end{cases}$$

$$= RAPP[0, T_1] + e^{-\rho T_1} RAPP[T_1, T_2]$$

So far, the ROE approach is behaving as desired.

If $\rho > 0$ and $T \to \infty$, then $RAPP = (E[L] + c\kappa T)\dfrac{1-e^{-\rho T}}{\rho T} \to \dfrac{c\kappa}{\rho}$. In other words, the present value of the expected loss becomes insignificant and the premium is the present value of the perpetuity of the risk charge. But as $\rho \to 0^+$, the value of this perpetuity becomes infinite. This is not a desirable behavior, since a risk-adjusted pure premium could be greater than a policy limit or maximum loss. Even if there were no maximum loss to the policy, a premium could be much greater than any probable loss. And the results are nonsensical if $\rho < 0$. Although a negative force of interest is unrealistic, a mathematical theory of pricing risk that suffers a discontinuity and excepts a domain is suspect.

On the other hand, as $T \to 0^+$, $RAPP = (E[L] + c\kappa T)\dfrac{1 - e^{-\rho T}}{\rho T} \to E[L]$. This is the problem of the evanescence of risk loads in the realm of the instantaneous, as mentioned at the end of the previous section. This too is not a desirable behavior. The author has tried to put a flattering construction on the ROE approach to pricing risk; nevertheless, this approach underprices very short-term risks and may well overprice very long-term risks. In fact, it is arguable that $RAPP[0, T]$ should monotonically decrease for $T \geq 0$, if $E[L]$ remains constant. For the maximum risk load should pertain to $RAPP[0, 0^+]$, since if the policy covering $[0, 0^+]$ loses money, it does so without the consolation prize of investment income.

The author does not expect that this critique of ROE pricing will convince everyone, or even most. Many will go back to their drawing boards and seek to rehabilitate this approach or to vindicate it against this critique. But the author will next turn to the utility-theoretic approach to the pricing of risk, hoping that the building up of a new approach will be more effective than the tearing down of an old. It is proverbial that one can catch more flies with honey than with vinegar. Those who wish for further argument against the ROE approach should consult Appendix E.

4) The Utility-Theoretic Approach to Pricing Risk

Perhaps Daniel Bernoulli was the father of utility theory; he employed it in his solution to the St. Petersburg Paradox (Appendix B). It is the foundation of much of modern economics and decision-making ([7], [15], and [26]). We will introduce utility theory with the example of the coin flip. A coin is about to flip, and if it lands heads it will cost a certain person one dollar. Otherwise, it will cost that person nothing. The person has come to us to purchase one dollar of insurance against the event of heads. What premium should we quote?

If we were *risk-neutral*, we would quote a pure premium equal to the expected loss, or fifty cents. If our quote were accepted and we received the fifty cents, then in the next moment with equal probability we would be either fifty cents richer (tails) or fifty cents poorer (heads). But if we were *risk-averse*, we would dread the loss of fifty cents more than we would welcome the gain of fifty cents. Then our quote would be greater than fifty cents, perhaps sixty cents. If we received sixty cents, after the flip we would be either sixty cents richer (tails) or forty cents poorer (heads). And if we were *risk-inclined*, we would welcome the gain of fifty cents more than we would dread the loss of fifty cents. In this case our quote would be less than fifty cents, perhaps forty cents. Then, after the flip, we would be either forty cents richer (tails) or sixty cents poorer (heads). If we were extremely risk-averse we would quote $1-\varepsilon$ dollars, and if we were extremely risk-inclined we would quote ε dollars, where ε is a small positive number.

But whatever our risk-tolerance might be, we would quote a pure premium greater than zero and less than one. This is a realistic description of how individuals and companies assess risk, and it makes no use of allocated capital and returns thereon.

The pricing of risk is really the weighing of alternatives, or utilities. Individuals and companies might not be aware that they have utility functions; they might operate on a visceral level. But, as Socrates said, "The unexamined life is not worth living." What happens on a visceral level can and should be brought before the intellect. So let us take for granted that there is some function of wealth $u(w)$ by which we weigh our alternatives. For now, we need to know only that u must be continuous and strictly increasing. Let w be our present wealth, and let p be the pure premium. After the coin flip our wealth will be either $w + p$ (tails) or $w + p - 1$ (heads). Since heads and tails are equally likely, this insurance transaction will change our expected utility from $u(w)$ to $0.5u(w+p-1)+0.5u(w+p)=f(p)$.

Since u is continuous and strictly increasing, f too must be continuous and strictly increasing. However, $f(0)=0.5u(w-1)+0.5u(w)<0.5u(w)+0.5u(w)=u(w)$. And $f(1)=0.5u(w)+0.5u(w+1)>0.5u(w)+0.5u(w)=u(w)$. So, $f(0)<u(w)<f(1)$. Since f is continuous, there must be some $0 < p^* < 1$ such that $f(p^*)=u(w)$. Entering into this transaction for any premium less than p^* decreases our expected utility, and entering for any premium greater than p^* increases it. p^* is the equilibrating premium. Our rule is that we should enter into no transaction that decreases our expected utility.

In general, if there are n possible outcomes for a loss L to happen in the next moment (l_1, … , l_n) with respective probabilities π_1, … , π_n, where $\sum_{i=1}^{n} \pi_i = 1$, then an acceptable premium satisfies the equation $\sum_{i=1}^{n} \pi_i u(w + p - l_i) > \sum_{i=1}^{n} \pi_i u(w) = 1 \cdot u(w) = u(w)$. If the outcomes are uncountable, then an integral form is appropriate:

$$\int u(w + p - l) f(l) dl > \int u(w) f(l) dl = 1 \cdot u(w) = u(w)$$

The current state may not be one of fixed wealth w, but of a distribution of wealths w_i and probabilities π_i. One can decide to change the distribution to wealths w_i* and probabilities π_i*. The decision is good if it increases the expected utility:

$$\sum_i \pi_i^* u(w_i^*) > \sum_i \pi_i u(w_i)$$

This formulation, which shows the effect of a decision on one's stochastic wealth, allows for the consideration of covariance among risks.

So far, the decisions have taken place instantaneously. Earlier the ROE approach was criticized for making time to be of the essence of the pricing of risk. This is not the case with utility theory. Consider a variation of the game show *The Price is Right*. There are three doors, behind which are \$100, \$200, and \$300 cash prizes. You may choose one door, but you do not know which prize is behind which door. However, you must pay to play this game. If your current wealth is fixed at w, your bid p is ruled by the equation:

$$\frac{1}{3} u(w - p + 100) + \frac{1}{3} u(w - p + 200) + \frac{1}{3} u(w - p + 300) > u(w)$$

Obviously, $\$100 < p < \$300$. But now we introduce the time element. The prizes are \$100 now, \$200 one year from now, and \$300 two years from now. Suppose that the risk-free yield curve is flat at six percent per year. Then the present values of the prizes are \$100, \$200/1.06, and $\$300/1.06^2$, at which values the prizes can be cashed in. This modifies the equation:

$$\frac{1}{3}u\left(w-p+\frac{100}{1.06^0}\right)+\frac{1}{3}u\left(w-p+\frac{200}{1.06^1}\right)+\frac{1}{3}u\left(w-p+\frac{300}{1.06^2}\right)>u(w)$$

The equilibrating p, which must be greater than $\$100/1.06^0$ and less than $\$300/1.06^2$, has nothing to do with some such ROE equation as:

$$p=\frac{1}{3}\frac{100}{(1+ROE)^0}+\frac{1}{3}\frac{200}{(1+ROE)^1}+\frac{1}{3}\frac{300}{(1+ROE)^2}$$

Time is not of the essence of utility-theoretic pricing. That this is a desirable feature is argued in Appendix A.

5) Covariance and Order Dependence

The price of insuring a new risk may depend on the risks that the insurer currently insures. The reason for this is that the new risk may covary with the current risks. This means that if A and B are two risks, what the insurer will quote to insure them may depend on the order in which they are submitted. Both ROE and utility-theoretic approaches to pricing consider the covariance phenomenon. But it is instructive to see how cleanly it is handled by utility theory.

Returning to the coin flip example, suppose that we decided to insure heads for the equilibrating premium p^*. Prior to this we had a non-stochastic wealth w. We have dropped the assumption that the coin is fair for the general probabilities π_H and π_T. Therefore, $u(w)=(\pi_H+\pi_T)u(w)=\pi_H u(w+p^*-1)+\pi_T u(w+p^*)$. Now suppose that someone wishes for us to quote for a dollar of insurance against tails. We calculate the equilibrating premium q^*, *acknowledging that we have already insured heads*, as:

$$\begin{aligned}
u(w) &= \pi_H u(w+p^*-1)+\pi_T u(w+p^*) \\
&= \pi_H u(w+p^*-1+q^*)+\pi_T u(w+p^*+q^*-1) \\
&= \pi_H u(w+(p^*+q^*)-1)+\pi_T u(w+(p^*+q^*)-1) \\
&= (\pi_H+\pi_T)u(w+(p^*+q^*)-1) \\
&= u(w+(p^*+q^*)-1)
\end{aligned}$$

But since u is a strictly increasing function:

$$\begin{aligned}
u(w) &= u(w+(p^*+q^*)-1) \\
w &= w+(p^*+q^*)-1 \\
1 &= p^*+q^*
\end{aligned}$$

The equilibrating premium to insure tails has been affected by the prior decision to insure heads. If the tails policy were issued first, p^* and q^* would be different, but their sum would still be one.

The covariance phenomenon sometimes annoys actuaries, because it implies that risk charges increase as positively correlated risks are added to a portfolio. So, for example, the second home insured in Miami, FL should be charged more than the first, and so forth. However, in practice this is not a problem, for an insurer usually decides to underwrite a block of correlated risks, in which case each risk is charged an average amount. The premium for the whole block may be affected by the presence of other blocks of insurance, but within the block order dependence is ignored.

6) A Versatile Utility Function

Until now $u(w)$ has been unspecified, only that it had to be continuous and strictly increasing. The fundamental theorem here is that a decision based on a utility function is unaffected by a positive linear transformation of that function. Let the linear transformation be $u^*(w) = au(w) + b$, where a is positive. Then u^*, like u, will be continuous and strictly increasing. Now suppose that the wealth distribution W is preferable to distribution W^* according to utility function u. Hence, $\sum_i \pi_i u(w_i) > \sum_i \pi_i^* u(w_i^*)$. Then, since $a > 0$, the order will not be affected by the following algebraic manipulations:

$$\sum_i \pi_i u(w_i) > \sum_i \pi_i^* u(w_i^*)$$
$$a\sum_i \pi_i u(w_i) > a\sum_i \pi_i^* u(w_i^*)$$
$$\left(a\sum_i \pi_i u(w_i)\right) + b > \left(a\sum_i \pi_i^* u(w_i^*)\right) + b$$
$$\left(a\sum_i \pi_i u(w_i)\right) + \sum_i \pi_i b > \left(a\sum_i \pi_i^* u(w_i^*)\right) + \sum_i \pi_i^* b$$
$$\sum_i \pi_i \left(au(w_i) + b\right) > \sum_i \pi_i^* \left(au(w_i^*) + b\right)$$
$$\sum_i \pi_i u^*(w_i) > \sum_i \pi_i^* u^*(w_i^*)$$

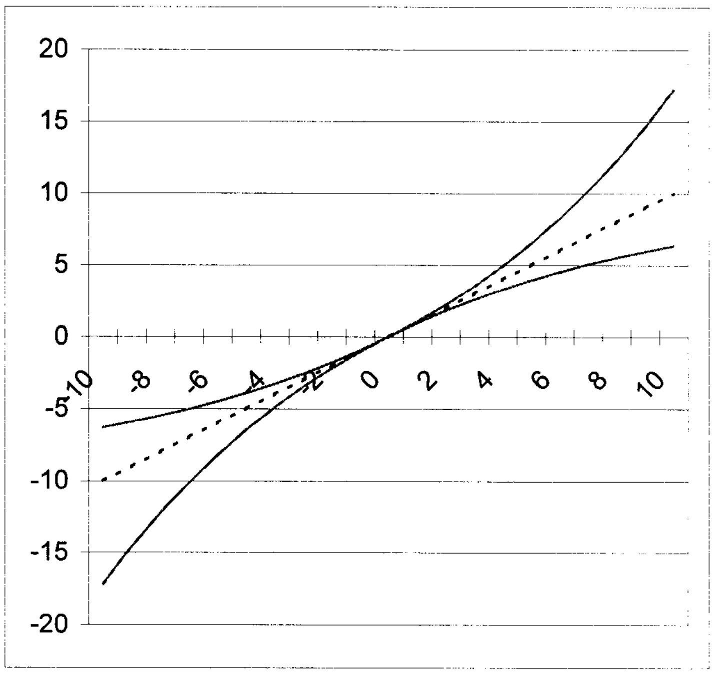
The proof holds whether the relation between W and W^* is '>', '<', '=', '≥', or '≤'. Therefore, positive linear transformations of utility functions are equivalent.

We desire for a utility function to be defined over all the real numbers. This excludes quadratic utility functions, because a quadratic function must somewhere be decreasing. This excludes also the logarithmic utility function, of which Daniel Bernoulli was so fond (Appendix B). For the logarithm of a negative number is not real. We will want to evaluate utilities such as $u(w - L)$, were $w - L$ represents an insurer's wealth after a loss. But the loss could be large enough to make $w - L$ negative. Some might reply that this would bankrupt the insurer, which is equivalent to a wealth of zero, or $\max(0, w - L)$. We disagree; utility-theoretic decisions should allow for negative wealth, since this will build into the decision not just the *fact* of the bankruptcy, but also its *magnitude*. Responsible decision-making should ignore limited corporate liability and guarantee funds; someone will have clean up the mess of a bankruptcy. And logarithmic utility fails with the formula $\max(0, w - L)$, if $\text{Prob}[w - L \le 0] > 0$, since $\ln(0)$ is undefined.

Probably we are not giving up much if we demand that u be differentiable, or even analytic. Not only should a realistic utility increase a little with a little increase of wealth; probably it should increase smoothly. Therefore, $u'(0)$ will exist and be positive. Then, the positive linear transformation $u^*(w) = \frac{1}{u'(0)}u(w) - \frac{u(0)}{u'(0)}$ is equivalent to u. But u^* has standard properties of passing through the origin with a slope of one.

A very versatile family of utility functions satisfies the differential equation $u'(w) = \alpha u(w) + 1$, with the initial condition that $u(0) = 0$. This implies that $u'(0) = 1$. And $u''(0) = \alpha u'(0) = \alpha$. If $\alpha = 0$, $u(w) = w$. This represents a risk-neutral utility. And if

$\alpha \neq 0$, the solution of the differential equation is $u(w;\alpha) = \dfrac{e^{\alpha w}-1}{\alpha}$. It can be proven by l'Hospital's rule that $\lim_{\alpha\to 0} u(w;\alpha) = \lim_{\alpha\to 0} \dfrac{e^{\alpha w}-1}{\alpha} = w = u(w;0)$. A graph of $u(w;\alpha)$ for three values of α (-0.1, 0, and 0.1) follows (w on the x-axis, $u(w;\alpha)$ on the y-axis):



The top curve ($\alpha = 0.1$) represents risk inclination. The chance of gaining is more useful than the chance of loosing. The middle curve ($\alpha = 0$ with dotted line) represents risk neutrality. And the bottom curve ($\alpha = -0.1$) represents risk aversion. The chance of gaining is less useful that that of loosing. There is a certain antisymmetry to u in that $u(-w;-\alpha) = -u(w;\alpha)$. So a loss of w dollars under ($-\alpha$) risk-aversion is as unpleasant as a gain of w dollars under (α) risk-inclination is pleasant.

Let X represent a stochastic (present) value with density function f. What would be its cash value to an investor whose risk tolerance is α and whose current (non-stochastic) wealth is w? Denote this value by $\xi_\alpha(X; w)$:

$$
\begin{aligned}
u(w) &= \int_{x=-\infty}^{\infty} u(w-\xi+x) f_X(x)dx \\
\frac{e^{\alpha w}-1}{\alpha} &= \int_{x=-\infty}^{\infty} \frac{e^{\alpha(w-\xi+x)}-1}{\alpha} f_X(x)dx \\
e^{\alpha w} &= \int_{x=-\infty}^{\infty} e^{\alpha(w-\xi+x)} f_X(x)dx \\
&= e^{\alpha(w-\xi)} \int_{x=-\infty}^{\infty} e^{\alpha x} f_X(x)dx \\
e^{\alpha\xi} &= \int_{x=-\infty}^{\infty} e^{\alpha x} f_X(x)dx \\
\xi_\alpha(X;w) &= \frac{\ln\left(\int_{x=-\infty}^{\infty} e^{\alpha x} f_X(x)dx\right)}{\alpha} \\
&= \frac{\ln(M_X(\alpha))}{\alpha}
\end{aligned}
$$

M_X denotes the moment generating function. With this particular family of utility functions (in effect, the exponential family) the initial wealth of the investor is irrelevant. Some will disagree, but the author believes this to be a desirable property.

Moreover, making α to change with the wealth of the company is a natural way of making the company's risk tolerance depend on its size (cf. [17:90]). Economists know that large firms are less risk-averse than small firms. It is reasonable that if company A is twice the size of company B, then company A can bear twice the loss at the same level of pain. Suppose that A is k times the size of B, and the cash value of X to company B is $\xi_\alpha(X)$. If A's risk tolerance is α/k, then the cash value of kX to A is:

$$\xi_{\frac{\alpha}{k}}(kX)=\frac{\ln\left(M_{kX}\left(\frac{\alpha}{k}\right)\right)}{\left(\frac{\alpha}{k}\right)}=\frac{k}{\alpha}\ln\left(M_X\left(\frac{\alpha}{k}k\right)\right)=k\frac{\ln(M_X(\alpha))}{\alpha}=k\xi_\alpha(X)$$

Similarly, if tomorrow the government were to declare the Q to be the new currency and one Q to equal k old dollars, a decision-making entity would go on its merry way just by changing its risk tolerance from α to αk. (The unit of α is currency^{-1}.)

If stochastic (present-valued) cash flows are *independent*, their cash values will be additive:

$$\begin{aligned}\xi_\alpha(X_1+X_2)&=\frac{\ln(M_{X_1+X_2}(\alpha))}{\alpha}\\&=\frac{\ln(M_{X_1}(\alpha)M_{X_2}(\alpha))}{\alpha}\\&=\frac{\ln(M_{X_1}(\alpha))+\ln(M_{X_2}(\alpha))}{\alpha}\\&=\xi_\alpha(X_1)+\xi_\alpha(X_2)\end{aligned}$$

This seems to be another desirable property.

If X is a normal random variable with mean μ and variance σ^2, $M_X(\alpha)=e^{\alpha\mu+\alpha^2\sigma^2/2}$. In this case $\xi_\alpha(X)=\frac{\ln\left(e^{\alpha\mu+\alpha^2\sigma^2/2}\right)}{\alpha}=\frac{\alpha\mu+\alpha^2\sigma^2/2}{\alpha}=\mu+\left(\frac{\alpha}{2}\right)\sigma^2$. A risk-inclined investor will pay more than expected value, a risk-averse one will pay less. In this case the risk load is proportional to the variance of the cash flow. Risk loads proportional to variances are widely thought to be reasonable.

If the company's stochastic present value is X_1, and it considers a project that will (immediately) change that value to X_2, and the values can be treated as normal random variables (or equivalently, if only the first two moments are considered), the project would be undertaken if:

$$\mu_1 + \left(\frac{\alpha}{2}\right)\sigma_1^2 < \mu_2 + \left(\frac{\alpha}{2}\right)\sigma_2^2$$

(Appendix D elucidates the properties of this relation.) The company would only need to know its risk tolerance α.

7) Estimating One's Risk Tolerance

If one were impressed with the versatility of the exponential family of utility functions, he would have to estimate his tolerance toward risk, the parameter α. This is not a formidable undertaking; in fact, it may be easier to estimate α than to estimate the risk-adjusted rate-of-return of the ROE approach.

As an example of how one might estimate α, let company A have a net worth of one billion dollars. Let the management of company A ponder the additional utility of the company's gaining another \$0.5B of net worth. Then ask it what loss of net worth would diminish the utility by the same amount. If the company were risk-neutral, a \$0.5B loss would be as unpleasant as a \$0.5B gain would be pleasant. Suppose that the management says that a \$0.3B loss countervails a half-billion dollar gain. This yields the equation:

$$\frac{e^{\alpha 1.0\mathrm{B}}-1}{\alpha}-\frac{e^{\alpha 0.7\mathrm{B}}-1}{\alpha}=\frac{e^{\alpha 1.5\mathrm{B}}-1}{\alpha}-\frac{e^{\alpha 1.0\mathrm{B}}-1}{\alpha}$$

$$2\frac{e^{\alpha 1.0\mathrm{B}}-1}{\alpha}=\frac{e^{\alpha 0.7\mathrm{B}}-1}{\alpha}+\frac{e^{\alpha 1.5\mathrm{B}}-1}{\alpha}$$

The solution is $\alpha = -1.305 \times 10^{-9}$. This method is simple; but it may rely too much on one judgment. A more robust method is to let the management judge the relations among the utilities of several amounts and to estimate the α that best fits these judgments.

8) A Utility-Theoretic Approach to the Poisson Example

Section 3 treated an example that revealed the inconsistency of the ROE approach. It is only fair to see whether utility theory fares any better. The problem was to price a policy covering the time interval $[0, T]$. Claim generation was a Poisson process with a constant frequency of λ claims per time unit, the severity of every claim was s dollars, and the force of interest was ρ.

The problem is that the distribution of the aggregate loss is not analytic. However, we can derive expressions for the mean, the variance, and the higher moments of the aggregate loss. For convenience we will use only the mean and the variance; and our utility will be exponential. Appendix C shows how the exponential family of utility functions works with the higher moments.

Let $L[t, t+dt]$ be the random variable of the loss incurred in the interval $[t, t+dt]$. The probability of a loss is λdt. Hence, the claim count is Bernoulli-distributed with parameter λdt, or Bernoulli[λdt]. So $L[t, t+dt]$ is distributed as $s \cdot$ Bernoulli[λdt], whose mean is $s\lambda dt$, and whose variance is $s^2\lambda dt(1 - \lambda dt)$. But higher-order terms in dt can be ignored: $Var[L[t,t+dt]] = s^2\lambda dt - s^2\lambda^2(dt)^2 \approx s^2\lambda dt$. The present value of $L[t, t+dt]$, $PV[L[t, t+dt]]$, equals $e^{-\rho t}L[t,t+dt]$; therefore, $PV[L[t, t+dt]]$ has mean $e^{-\rho t}s\lambda dt$ and variance $e^{-2\rho t}s^2\lambda dt$.

The expectation of the sum equals the sum of the expectations. Hence:

$$\begin{aligned}
E[PV[L[0,T]]] &= \int_{t=0}^{T} E[PV[L[t,t+dt]]] \\
&= \int_{t=0}^{T} e^{-\rho t} s\lambda dt \\
&= s\lambda \frac{1}{\rho} \int_{t=0}^{T} e^{-\rho t} \rho dt \\
&= s\lambda \frac{1-e^{-\rho t}}{\rho} \\
&= s\lambda T \frac{1-e^{-\rho T}}{\rho T}
\end{aligned}$$

And, since the increments of a Poisson process are independent, the variance of the sum equals the sum of the variances:

$$\begin{aligned}
Var[PV[L[0,T]]] &= \int_{t=0}^{T} Var[PV[L[t,t+dt]]] \\
&= \int_{t=0}^{T} e^{-2\rho t} s^2\lambda dt \\
&= s^2\lambda \frac{1}{2\rho} \int_{t=0}^{T} e^{-2\rho t} 2\rho dt \\
&= s^2\lambda \frac{1-e^{-2\rho t}}{2\rho} \\
&= s^2\lambda T \frac{1-e^{-2\rho T}}{2\rho T}
\end{aligned}$$

As in Section 3, let the expected loss be constant for all T by making $\lambda(T) = \frac{\lambda_0}{T} = \frac{E[L]}{sT}$:

$$E[PV[L[0,T]]] = s\lambda(T)T\frac{1-e^{-\rho T}}{\rho T} = E[L]\frac{1-e^{-\rho T}}{\rho T}$$

$$Var[PV[L[0,T]]] = s^2\lambda(T)T\frac{1-e^{-2\rho T}}{2\rho T} = sE[L]\frac{1-e^{-2\rho T}}{2\rho T}$$

In Section 6 we proved the cash value of wealth inflow X to an investor whose utility is exponential with risk tolerance is α to be $\xi_\alpha(X) = \frac{\ln(M_X(\alpha))}{\alpha}$. (This assumes that X does not covary with the investor's current distribution of wealth.) And if X is normal with mean μ and variance σ^2, then $\xi_\alpha(X) = \frac{\ln\left(e^{\alpha\mu+\alpha^2\sigma^2/2}\right)}{\alpha} = \frac{\alpha\mu+\alpha^2\sigma^2/2}{\alpha} = \mu + \left(\frac{\alpha}{2}\right)\sigma^2$. A loss is a wealth outflow, or a negative inflow. And a premium is money received, or negative money spent. Therefore, the equilibrating premium of loss L to an insurer is:

$$p = -\xi_\alpha(-L) = -\frac{\ln(M_{-L}(\alpha))}{\alpha} = \frac{\ln\left(E\left[e^{-\alpha L}\right]\right)}{-\alpha} = \frac{\ln(M_L(-\alpha))}{-\alpha}$$

It is worthwhile to corroborate this equation by deriving it in full:

$$u(w)=\int_{l=-\infty}^{\infty}u(w+p-l)f_L(l)dl$$

$$\frac{e^{\alpha w}-1}{\alpha}=\int_{l=-\infty}^{\infty}\frac{e^{\alpha(w+p-l)}-1}{\alpha}f_L(l)dl$$

$$e^{\alpha w}=\int_{l=-\infty}^{\infty}e^{\alpha(w+p-l)}f_L(l)dl$$

$$=e^{\alpha(w+p)}\int_{l=-\infty}^{\infty}e^{-\alpha l}f_L(l)dl$$

$$e^{-\alpha p}=\int_{l=-\infty}^{\infty}e^{-\alpha l}f_L(l)dl$$

$$p_\alpha(L;w)=\frac{\ln\left(\int_{l=-\infty}^{\infty}e^{-\alpha l}f_L(l)dl\right)}{-\alpha}$$

$$=\frac{\ln(M_L(-\alpha))}{-\alpha}$$

So, if loss L is normal with mean μ and variance σ^2, or if loss L is effectively normal by the consideration of just its mean and variance, then:

$$p_\alpha(L)=\frac{\ln\left(e^{-\alpha\mu+\alpha^2\sigma^2/2}\right)}{-\alpha}=\frac{-\alpha\mu+\alpha^2\sigma^2/2}{-\alpha}=\mu-\left(\frac{\alpha}{2}\right)\sigma^2=\mu+\left(\frac{-\alpha}{2}\right)\sigma^2.$$

If the insurer is risk-averse, then its risk tolerance α is negative and it will charge more that the pure premium μ.

We can now formulate the utility-theoretic answer to the Poisson example:

$$RAPP[0,T]=p_\alpha(L)$$

$$=E[PV[L[0,T]]]+\left(\frac{-\alpha}{2}\right)Var[PV[L[0,T]]]$$

$$=E[L]\frac{1-e^{-\rho T}}{\rho T}+\left(\frac{-\alpha}{2}\right)sE[L]\frac{1-e^{-2\rho T}}{2\rho T}$$

$$=E[L]\varphi(\rho T)+\left(\frac{-\alpha}{2}\right)sE[L]\varphi(2\rho T)$$

At the end of Section 3 we wondered whether *RAPP* should be a decreasing function of T. Now we have an answer. $\varphi(x)=\begin{cases}\dfrac{1-e^{-x}}{x} & x\neq 0\\ 1 & x=0\end{cases}$ is a decreasing function of x.

Therefore, if $\rho > 0$, *RAPP* decreases with T, a confirmation of our intuition. The other conditions of ρ, though unrealistic, also confirm our intuition: If $\rho = 0$, *RAPP* is constant with T; and if $\rho < 0$, *RAPP* increases with T. Utility theory fares well in this example; no inconsistency has arisen. And one can show that the utility-theoretic answer here has the same desirable characteristic as the ROE answer in Section 3 had, viz., that $RAPP[0,T_2]=RAPP[0,T_1]+e^{-\rho T_1}RAPP[0,T_2-T_1]$. So utility theory has the advantage without the disadvantage.

The author thinks that an insurer at any time should have a fairly good idea of the present value of its stochastic cash flows, whether they stem from assets or from liabilities. It should have an idea about the flows of a contemplated project and how they covary with its own flows. Then with a utility function it can decide whether the project is worthwhile. In the estimation of the cash flows lies a multitude of details; but the details plague every approach. Utility theory has no more of them than do the others, perhaps less.

9) Capital and the Strength of an Insurer

Section 6 argued that responsible decision-making should not countenance such protections to the risk-taker as limited corporate liability and guaranty funds. Many managers of insurance companies do not think beyond the probability of ruin, if they think even that far. But both the utility-theoretic and the ROE approaches encourage considering the magnitude of insolvency as well as the probability thereof.

Undoubtedly, the focus of the utility-theoretic approach to pricing is not solvency. But on the other hand, neither is solvency the focus of the ROE approach, although some seem to believe that it derives naturally, or almost invariably, from a "solvency first" mentality. But with either approach there will be times when an insurer will have to decline a good deal simply because it is loaded to capacity, as James Stone [25] has demonstrated. Conversely, at times an insurer will accept a bad deal simply because it is *not* loaded to capacity. It may try to convince itself that the deal really is not too bad, or that it has intangible benefits; but the truth may be that the insurer, like an impatient child, just can't stand to be idle. And as long as the insurer is not in a make-believe world, its low-balling injures no one. Indeed, would anyone purchase a policy from a company that couldn't afford to give to him for free? Of course, if the company gave free insurance to all comers, eventually it would become insolvent. But on the margin, for the next few policies, or every now and then, a healthy company could afford to give away insurance. All this indicates how loose the link is between pricing and solvency, especially in the short term.

Perhaps a prudent utility function will make insolvency tolerably unlikely. Perhaps a prudent ROE approach will also make it unlikely. More likely, prudence itself will make insolvency tolerably unlikely. Whatever the approach to pricing, it is prudent for an insurer's financial strength to be assessed by regulators, rating agencies, and the insurer itself. Financial ratios, stress tests, risk-based capital, and dynamic financial analysis should prove to be helpful for this assessment, in addition to non-quantitative means.

But assessing how well an insurer is capitalized does not necessitate the allocation of its capital to individual risks. However the company might internally allocate its capital, every dollar's worth of it is on call to meet the loss of any risk. A strong man who contracts the flu is still strong; but a weak man is weak even on his best days. Likewise, a strong company may have ailing divisions; but the insurance issued by these divisions is just as secure as that issued by the healthy divisions. Conversely, the insurance issued by a profitable division of a weak company is not secure. The only caveat is that the capital must truly be one; the company can't have firewalls between parts of its capital in the form of legally distinct entities. But this goes without saying. So *capital* is the sum and substance of an insurer's financial strength. *Capital allocation* is nonsensical for the assessment of financial strength; and it yields illogical and inconsistent results as a pricing technique.

References

[1] Bass, Irene K, and Khury, C. K., "Surplus Allocation: An Oxymoron," *Insurer Financial Solvency*, 1992 CAS *Discussion Paper Program*, 553-584.

[2] Bernstein, Peter L., *Against the Gods: The Remarkable Story of Risk*, New York, John Wiley & Sons, 1998.

[3] Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbitt, C. J., *Actuarial Mathematics*, Society of Actuaries,1986.

[4] Butsic, Robert, "Determining the Proper Interest Rate for Loss Reserve Discounting: An Economic Approach," *Evaluating Insurance Company Liabilities*, 1988 CAS *Discussion Paper Program*, 147-188.

[5] D'Arcy, Stephen P., and Dyer, Michael A., "Ratemaking: A Financial Economics Approach," *PCAS* LXXXIV (1997), 301-390.

[6] Daykin, C. D., Pentikäinen, T., and Pesonen, M., *Practical Risk Theory for Actuaries*, London, Chapman & Hall, 1994.

[7] Debreu, Gerard, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New Haven, Yale, 1987.

[8] Dixit, Avinash K., *Optimization in Economic Theory*, Second Edition, Oxford, 1990.

[9] Duffie, Darrell, *Dynamic Asset Pricing Theory*, Second Edition, Princeton, 1996.

[10] Feldblum, Sholom, "Risk Loads for Insurers," *PCAS* LXXVII (1990), 160-195.

[11] Gogol, Daniel F., "Discussion of Paper Published in Volume LXXVII: 'Reinsurer Risk Loads from Marginal Surplus Requirements,' by Rodney E. Kreps," *PCAS* LXXIX (1992), 362-366.

[12] Halliwell, Leigh J., "Conjoint Prediction of Paid and Incurred Losses," *1997 CAS Loss Reserving Discussion Papers*, 241-379.

[13] Kneuer, Paul J., "Allocation of Surplus for a Multi-Line Insurer," *Financial Analysis of Insurance Companies*, 1987 CAS *Discussion Paper Program*, 191-228.

[14] Kneuer, Paul J., "Discussion of Paper Published in Volume LXXVII: 'Reinsurer Risk Loads from Marginal Surplus Requirements' by Rodney E. Kreps," *PCAS* LXXXV (1998 [forthcoming]).

[15] Kreps, David M., *A Course in Microeconomic Theory*, Princeton, 1990.

[16] Kreps, Rodney E., "Reinsurer Risk Loads from Marginal Surplus Requirements," *PCAS* LXXVII (1990), 196-203.

[17] Longley-Cook, Alastair G., "Risk-Adjusted Economic Value Analysis," *North American Actuarial Journal*, Volume 2 (Jan 1998), 87-100.

[18] Lowe, Stephen P., and Stanard, James N., "An Integrated Dynamic Financial Analysis and Decision Support System for a Property Catastrophe Reinsurer," *CAS Forum* (Spring 1996), 89-118.

[19] Mango, Donald F., "An Application of Game Theory: Property Catastrophe Risk Load," *CAS Forum* (Spring 1997), 31-51.

[20] Meyers, Glenn, "Managing the Catastrophe Risk," *Incorporating Risk Factors in Dynamic Financial Analysis*, 1995 CAS *Discussion Paper Program*, 111-150.

[21] Neftci, Salih N., *An Introduction to the Mathematics of Financial Derivatives*, San Diego, Academic Press, 1996.

[22] Panjer, Harry H., editor, *Financial Economics: with Applications to Investments, Insurance and Pensions*, Schaumburg, IL, Actuarial Foundation, 1998.

[23] Philbrick, Stephen, "Accounting for Risk Margins," *CAS Forum* (Spring 1994), 1-87.

[24] Searle, S. R., *Linear Models*, New York, John Wiley & Sons, 1971.

[25] Stone, James M., "The Theory of Capacity and the Insurance of Catastrophe Risks," *The Journal of Risk and Insurance*, Volume XL, 1973, 339-355.

[26] Von Neumann, John, and Morgenstern, Oskar, *The Theory of Games and Economic Behavior*, Princeton, 1972.

Appendix A: The Fundamental Theorem of Asset Pricing

Imagine that you walk into a casino and see a table of payoffs and prices. The casino offers n games, the outcome of each game depending on which one of s possible states is realized. So the table shows an $(n \times s)$ matrix D, whose ij^{th} element is the payoff from one unit of game i in state j. (The notation is Duffie's [9:3f.], for whom 'D' stands for 'dividend'.) The table also shows an $(n \times 1)$ vector q, whose i^{th} element is the cost per unit of game i. So the table looks like the partitioned matrix $[\mathrm{D} \mid \mathrm{q}]$.

You may choose any amount of any game. These amounts are the elements of an $(n \times 1)$ vector θ, which is your portfolio. You may be long or short in any game; negative elements of θ are perfectly legitimate. The payoff of your portfolio is the $(s \times 1)$ vector $\mathrm{D}'\theta$, for which you must pay the cost $\mathrm{q}'\theta$. Let J_s be the $(s \times 1)$ vector of ones. Then the profit of your portfolio is the $(s \times 1)$ vector $\mathrm{D}'\theta - \mathrm{J}_s\mathrm{q}'\theta = (\mathrm{D}' - \mathrm{J}_s\mathrm{q}')\theta$. The time value of money is not an issue here, because immediately after you position yourself with a portfolio θ, a random variable $S \in \{1, 2, \ldots, s\}$ is realized and your profit is determined.

Being a cunning person, you would like to take advantage of the casino by finding some portfolio θ that produces a positive profit in every state. A little less ambitious is to produce a non-negative profit in every state. Let $\Re^s_+$ (the non-negative orthant [7:26]) be the set of vectors in s-space all whose elements are non-negative. Similarly, let $\Re^s_{++}$ (the positive orthant) be the set of vectors in s-space all whose elements are positive. You would like to find a portfolio whose profit is in the positive orthant, or at least in the non-negative orthant. Of course, $\theta = 0$, the null portfolio, will produce a profit of 0, which belongs to the non-negative orthant. So you'd be happy with a profit anywhere in the non-negative orthant except for 0. This means that you have some possibility of a positive profit and no possibility of a negative profit — upside potential without downside potential.

Define L as the span of $\mathrm{D}' - \mathrm{J}_s \mathrm{q}'$, i.e., the set of all attainable profit vectors. For every vector $\mathrm{x} \in L$, there is a portfolio θ such that $\mathrm{x} = (\mathrm{D}' - \mathrm{J}_s \mathrm{q}')\theta$. L is a linear subspace of $\Re^s$, since $0 \in L$ and if $\mathrm{x}, \mathrm{y} \in L$, then for any scalars a and b $a\mathrm{x} + b\mathrm{y} \in L$. L intersects $\Re^s_+$ at least at 0. But the sure bets would be at non-zero vectors belonging to $L \cap \Re^s_+$. Thus we can define arbitrage. The table $[\mathrm{D} \,\vdots\, \mathrm{q}]$ is *arbitrage-free* if and only if $L \cap \Re^s_+ = \{0\}$.

We make one more useful definition. The $(s\times1)$ vector ψ is a *state-price vector* if and only if $\psi \in \Re^s_{++}$ and $\mathrm{J}'_s \psi = 1$. If $\mathrm{x} \in \Re^s_{++}$, then $\mathrm{J}'_s \mathrm{x} > 0$, so $(\mathrm{J}'_s \mathrm{x})^{-1}$ exists and $\mathrm{x}(\mathrm{J}'_s \mathrm{x})^{-1}$ qualifies as a state-price vector. A state-price vector can be interpreted as a set of non-zero probabilities and as a weighted-average operator. If x is an $(s\times1)$ vector, then $\psi'\mathrm{x}$ is a weighted average of x, each element receiving some weight. $\mathrm{x} - \mathrm{J}_s \psi'\mathrm{x}$ is the vector of deviations from the weighted average. It is important to note that $\mathrm{x} - \mathrm{J}_s \psi'\mathrm{x} \in \Re^s_+$ if and only of x is constant. Otherwise at least one deviation will be negative. But if x is constant, then $x - J_s\psi'x = 0$. Therefore, if ψ is a state-price vector and M is the span of $I_s - J_s\psi'$, then $M \cap \Re_+^s = \{0\}$.

The fundamental theorem of asset pricing states that $[D \mid q]$ is arbitrage-free if and only if there exists a state-price vector ψ such that $q = D\psi$. In our imaginary casino, if such a vector exists, there is no sure bet; upside potential is inseparable from downside potential.

The 'if' part of the proof (the '$\Leftarrow$' direction) is easy. Suppose that there exists a state-price vector ψ such that $q = D\psi$. Then the profit of portfolio θ is:

$$\begin{aligned}(D' - J_s q')\theta &= \left(D' - J_s (D\psi)'\right)\theta \\ &= (D' - J_s\psi'D')\theta \\ &= (I_s - J_s\psi')D'\theta\end{aligned}$$

But, since ψ is a state-price vector, the span of $I_s - J_s\psi'$ intersects $\Re_+^s$ only at 0. The span of $(I_s - J_s\psi')D'$ is a subset of the span of $I_s - J_s\psi'$; therefore, its intersection with $\Re_+^s$ must be a subset of $\{0\}$. But 0 does belong to the span. Therefore, the span of $(I_s - J_s\psi')D'$, which is the span of $D' - J_s q'$, intersects $\Re_+^s$ only at 0. So if there exists a state-price vector ψ such that $q = D\psi$, then $[D \mid q]$ is arbitrage-free.

The 'only if' part of the proof (the '$\Rightarrow$' direction) requires the Separating Hyperplane Theorem [22:525]. According to this theorem, if L is a linear subspace of $\Re^s$ and $L \cap \Re_+^s = \{0\}$, then there exists a vector $\varphi \in \Re_{++}^s$ such that φ is orthogonal (or perpendicular) to every vector of L. In other words, if $x \in L$, then $\varphi'x = 0$. Geometrical intuition in two or three dimensions suggests the truth of this theorem; but to prove it, as Debreu does [7:24], is arduous.

Now suppose that $[D \mid q]$ is arbitrage-free, and let L be the span of $D' - J_s q'$. Then, by the definition of arbitrage, $L \cap \Re^s_+ = \{0\}$. L is a linear subspace of $\Re^s$, so by the Separating Hyperplane Theorem there exists a vector $\varphi \in \Re^s_{++}$ such that φ is orthogonal to every vector of L. But every vector of L is of the form $(D' - J_s q')\theta$. Therefore, $\forall_{\theta \in \Re^n} \varphi'(D' - J_s q')\theta = 0$. That this equation holds true for all θ requires that $\varphi'(D' - J_s q') = 0$. Simplifying, we have:

$$
\begin{aligned}
0 &= \varphi'(D' - J_s q') \\
\varphi' J_s q' &= \varphi' D' \\
(\varphi' J_s q')' &= (\varphi' D')' \\
q J_s' \varphi &= D\varphi \\
q J_s' \varphi (J_s' \varphi)^{-1} &= D\varphi (J_s' \varphi)^{-1} \\
q &= D\varphi (J_s' \varphi)^{-1} \\
&= D\psi
\end{aligned}
$$

As mentioned above, since $\varphi \in \Re^s_{++}$, $\psi = \varphi(J_s' \varphi)^{-1}$ exists and qualifies as a state-price vector. Therefore, if $[D \mid q]$ is arbitrage-free, then there exists a state-price vector ψ such that $q = D\psi$.

The span of D', a linear subspace of $\Re^s$, contains every payoff attainable by some portfolio θ. If r is the rank of matrix D, then the span of D' will be r-dimensional. This means that the span of D' will not fill $\Re^s$ unless $r = s$, i.e., unless D is of full column rank. A market is *complete* [9:8 and 22:192] if and only if every payoff of $\Re^s$ is attainable. In our casino the completeness of payoffs would require that there be at least s games, or $n \geq s$. But more precisely, it would require that there be s linearly independent games.

If casino $[\mathrm{D} \mid \mathrm{q}]$ is arbitrage-free, then there exists a state-price vector ψ such that $\mathrm{q} = \mathrm{D}\psi$. The payoff of the portfolio of games θ is $\mathrm{D}'\theta$, and the cost to play that portfolio is:

$$\begin{aligned} \mathrm{q}'\theta &= (\mathrm{D}\psi)'\theta \\ &= \psi'\mathrm{D}'\theta \\ &= \psi'(\mathrm{D}'\theta) \\ &= E_\psi[\mathrm{D}'\theta] \end{aligned}$$

Therefore, arbitrage-free prices are weighted averages of the outcomes, i.e., expectations over the probability vector ψ. This vector may be unrelated to the probabilities of the s states. The only caveat is that the probability of each of the states must be positive (in this finite setting). But the casino would be deceiving its clients if it listed columns in the D matrix that couldn't be realized. An honest casino might disclose the probabilities of the states as a row heading above the D matrix. We would ignore any states whose probability is zero, and prefer that the casino not list them.

Let e_i be the $(s\times 1)$ vector whose i^{th} element is one and whose other elements are zeroes. Whether or not e_i is an attainable payoff in the arbitrage-free casino $[\mathrm{D} \mid \mathrm{q}]$, the cost of playing for payoff e_i is $\psi'\mathrm{e}_i = \sum_{k=1}^{s} \psi_k [\mathrm{e}_i]_k = \psi_i$. Thus, ψ_i is price for a game which pays one unit of money if state i is realized and zero units otherwise. This explains why ψ is called a "state-price" vector [9:3 and 22:184]. It makes sense that $0 < \psi_i \leq 1$, since there is the possibility of a positive payoff and no possibility of a negative payoff. Also, the cost of playing for payoff J_s is $\psi' J_s = \sum_{k=1}^{s} \psi_k 1 = \sum_{k=1}^{s} \psi_k = 1$. It too makes sense that a constant payoff of one unit should cost one unit.

As mentioned above, the probabilities of the state-price vector may be unrelated to the real probabilities of the states. This shows that arbitrage is not the whole story. The missing elements are linear optimization and equilibrium, which [7], [8], [9], and [22] discuss. But arbitrage is foundational; and in a complete arbitrage-free market the price of any newcomer (whether an asset or a game) is determined by the prices of what is already in the market.

Though arbitrage is not the whole story, it can take us a long way. We will illustrate this with two examples. The first example is the pricing of an option to call a stock when the stock will realize only two prices. This simple model of the call option is sometimes called the binomial model or the one-period model. The casino offers two games (n = 2), the stock game and the call-option game. There are two possible states (s = 2), the up-market state and the down-market state. If the market goes up, the stock will sell for S_u and the call will sell for C_u. If the market goes down, the stock will sell for S_d and the call will sell for C_d. The cost to play one unit of the stock game is S, and that of the call-option game is C. So the casino displays the table:

$$[D \mid q] = \left[\begin{array}{cc|c} S_u & S_d & S \\ C_u & C_d & C \end{array}\right]$$

We will assume that D is non-singular, i.e., that its determinant is non-zero. For the call option has an excise price E, and its payoff is $C = \max(0, S - E)$. So C_u and C_d should be linearly independent of S_u and S_d. Of course, $S_d < S_u$, which implies that $C_d \leq C_u$,. Equality is possible only if the excise price is greater than S_u, which ensures that the call has a value of zero. But this is not an interesting possibility; moreover, the determinant of D would be zero. So we will assume that $C_d < C_u$.

If the table $[\mathrm{D} \mid \mathrm{q}]$ is arbitrage-free, then there exists a state-price vector ψ such that $\mathrm{q} = \mathrm{D}\psi$. Since D is non-singular, D^{-1} exists and ψ must equal $\mathrm{D}^{-1}\mathrm{q}$. Hence:

$$\begin{aligned}
\psi &= \mathrm{D}^{-1}\mathrm{q} \\
&= \begin{bmatrix} S_u & S_d \\ C_u & C_d \end{bmatrix}^{-1} \begin{bmatrix} S \\ C \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} C_d & -S_d \\ -C_u & S_u \end{bmatrix} \begin{bmatrix} S \\ C \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} SC_d - S_d C \\ S_u C - SC_u \end{bmatrix}
\end{aligned}$$

But ψ must also be a state-price vector, which means that its elements must sum to one. This will yield an equation that relates the prices of the stock and the call:

$$\begin{aligned}
\frac{(SC_d - S_d C) + (S_u C - SC_u)}{S_u C_d - S_d C_u} &= 1 \\
(SC_d - S_d C) + (S_u C - SC_u) &= S_u C_d - S_d C_u \\
-S(C_u - C_d) + C(S_u - S_d) &= S_u C_d - S_d C_u
\end{aligned}$$

Since the price of the call derives from the price of the stock:

$$\begin{aligned}
C(S_u - S_d) &= S_u C_d - S_d C_u + S(C_u - C_d) \\
C &= C_u \frac{S - S_d}{S_u - S_d} + C_d \frac{S_u - S}{S_u - S_d}
\end{aligned}$$

Therefore, the fundamental theorem of asset pricing determines the price of the call option as a weighted average of its terminal values. The weights $w_u = \dfrac{S - S_d}{S_u - S_d}$ and $w_d = \dfrac{S_u - S}{S_u - S_d}$, being unrelated to the probabilities of the up and down markets, are artificial probabilities.

The state price vector must be:

$$
\begin{aligned}
\psi &= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} SC_d - S_d C \\ S_u C - SC_u \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} SC_d - S_d\left(w_u C_u + w_d C_d\right) \\ S_u\left(w_u C_u + w_d C_d\right) - SC_u \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} \left(S - w_d S_d\right)C_d - S_d w_u C_u \\ S_u w_d C_d + \left(S_u w_u - S\right)C_u \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} \left(S - \left(\dfrac{S_u - S}{S_u - S_d}\right)S_d\right)C_d - S_d\left(\dfrac{S - S_d}{S_u - S_d}\right)C_u \\ S_u\left(\dfrac{S_u - S}{S_u - S_d}\right)C_d + \left(S_u\left(\dfrac{S - S_d}{S_u - S_d}\right) - S\right)C_u \end{bmatrix} \\
&= \frac{1}{S_u C_d - S_d C_u} \begin{bmatrix} S_u\left(\dfrac{S - S_d}{S_u - S_d}\right)C_d - S_d\left(\dfrac{S - S_d}{S_u - S_d}\right)C_u \\ S_u\left(\dfrac{S_u - S}{S_u - S_d}\right)C_d - S_d\left(\dfrac{S_u - S}{S_u - S_d}\right)C_u \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{S - S_d}{S_u - S_d} \\ \dfrac{S_u - S}{S_u - S_d} \end{bmatrix} \\
&= \begin{bmatrix} w_u \\ w_d \end{bmatrix}
\end{aligned}
$$

This makes sense since the state-price vector must contain the prices of one unit of payoff for each state. So w_u is the price for one unit of payoff if the market goes up, and w_d is the price for one unit of payoff if the market goes down. As a check:

$$
\begin{aligned}
\mathrm{D}\psi &= \begin{bmatrix} S_u & S_d \\ C_u & C_d \end{bmatrix} \begin{bmatrix} w_u \\ w_d \end{bmatrix} \\
&= \begin{bmatrix} S_u w_u + S_d w_d \\ C_u w_u + C_d w_d \end{bmatrix} \\
&= \begin{bmatrix} S_u \left(\dfrac{S - S_d}{S_u - S_d} \right) + S_d \left(\dfrac{S_u - S}{S_u - S_d} \right) \\ C_u w_u + C_d w_d \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{S_u S - S_d S}{S_u - S_d} \\ C_u w_u + C_d w_d \end{bmatrix} \\
&= \begin{bmatrix} S \\ C \end{bmatrix} \\
&= \mathrm{q}
\end{aligned}
$$

The hedge ratio is the number of call options that will immunize one share of stock:

$$
\begin{aligned}
S_u - hC_u &= S_d - hC_d \\
h(C_u - C_d) &= S_u - S_d \\
h &= \frac{S_u - S_d}{C_u - C_d}
\end{aligned}
$$

All these formulas agree with those of the binomial, or one-period, option model; but we derived them explicitly from the fundamental theorem of asset pricing.

The second example shows how to handle the present value of a payoff. So far, the casino game has been instantaneous; the payoff comes immediately after one positions himself (i.e., places his bets). But suppose that the casino offered a new game: to pay d_i at time t_i, if state i is realized. Until now, all the t_is have been 0^+.

We take for granted the existence of a present-value function $v(t)$, the price that the "market" will pay at time 0 for the certain reception of one unit of value at time t. No one would contract to pay you a unit at time t for less than $v(t)$, since he could get $v(t)$ from the market buyers. And why would you agree to pay more than $v(t)$, knowing that the market sellers were willing to contract for $v(t)$?

The price for this casino game must be the same as the price for the game that will pay $v(t_i)d_i$ at time 0^+, if state i is realized. The reason for this is that one can cash in the certain reception of d_i at time t_i to the market for the reception of $v(t_i)d_i$ now. This suggests that when pricing a stochastic cash flow one should present-value the flows by multiplying them by $v(t)$, and then price the stochastic payoff at time 0^+. This would imply that time is not of the essence of pricing stochastic cash flows and that discounting them by so-called risk-adjusted rates of return is faulty [12:349-356].

Appendix B: The St. Petersburg Paradox

In the 1730s the St. Petersburg Paradox gave Daniel Bernoulli the inspiration for utility theory. A coin is flipped until heads appears. If heads appears on the n^{th} flip, the player of the game will receive 2^n units of money. How much should one pay to play this game?

Let N be the number of the first heads flip. N is a geometric random variable and $\Pr[N = n] = 2^{-n}$. The payoff is equal to 2^N. The expected value of the payoff is:

$$E[2^N] = \lim_{n\to\infty}\sum_{i=1}^{n} 2^i \Pr[N = i] = \lim_{n\to\infty}\sum_{i=1}^{n} 2^i 2^{-i} = \lim_{n\to\infty}\sum_{i=1}^{n} 1 = \lim_{n\to\infty} n = \infty$$

As we would say today, if one were risk-neutral, then no price would be too large for one to play this game. Bernoulli knew that real persons would not pay an unlimited price for the game, and he reasoned that a payoff of 2^N was not really worth that much to the players. He argued that "the utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods previously possessed," [2:105] or:

$$\begin{aligned} du(x) &\propto \frac{1}{x}dx \\ &= \frac{a}{x}dx \\ u(x) &= u(1) + a\ln(x) \end{aligned}$$

Since $u(1)$ can be standardized to zero, $u(x) = a\ln(x)$. He also supposed multi-commodity utility to be additive. In other words, if x, y, z, … are amounts of different commodities, there are positive constants a, b, c, … such that:

$$
\begin{aligned}
u(x, y, z, \ldots) &= au(x) + bu(y) + cu(z) + \ldots \\
&= a\ln(x) + b\ln(y) + c\ln(z) + \ldots \\
&= \ln(x^a) + \ln(y^b) + \ln(z^c) + \ldots \\
&= \ln(x^a y^b z^c \ldots)
\end{aligned}
$$

An indifference curve, on which $u(x, y, z, \ldots)$ is constant, or on which $x^a y^b z^c \ldots$ is constant, is an hyperboliod. Modern Economics textbooks follow Bernoulli when they show indifference, or demand, curves as hyperbolas in the positive xy quadrant. When Bernoulli spoke of diminishing marginal utility, he thought of the logarithmic function, with its positive first derivative and its negative second derivative.

A general form of the St. Petersburg Paradox for a person whose initial wealth is w and will pay p to play the game is $u(w) = \sum_{i=1}^{\infty} u(w - p + 2^i)\Pr[N = i] = \sum_{i=1}^{\infty} u(w - p + 2^i)2^{-i}$.

However, many utility functions, one of which is the logarithmic, do not allow for an analytic solution for p. So Bernoulli made the simplification that $w = p$:

$$u(p) = \sum_{i=1}^{\infty} u(p - p + 2^i)2^{-i} = \sum_{i=1}^{\infty} u(2^i)2^{-i}$$

If the utility function is $u(x) = a\ln(x)$, we can ignore the positive constant a and formulate the equation: $\ln(p) = \sum_{i=1}^{\infty} \ln(2^i)2^{-i} = \ln(2)\sum_{i=1}^{\infty} \frac{i}{2^i}$. We can solve for p:

$$
\begin{aligned}
\frac{\ln(p)}{2} &= \ln(2)\sum_{i=1}^{\infty}\frac{i}{2^{i+1}} \\
&= \ln(2)\sum_{i=2}^{\infty}\frac{i-1}{2^{i}} \\
&= \ln(2)\left(\sum_{i=2}^{\infty}\frac{i}{2^{i}} - \sum_{i=2}^{\infty}\frac{1}{2^{i}}\right) \\
&= \ln(2)\left(\sum_{i=2}^{\infty}\frac{i}{2^{i}} - \frac{1}{2}\right) \\
&= \ln(2)\left(-\frac{1}{2} + \sum_{i=1}^{\infty}\frac{i}{2^{i}} - \frac{1}{2}\right) \\
&= -\ln(2) + \ln(2)\sum_{i=1}^{\infty}\frac{i}{2^{i}} \\
&= -\ln(2) + \ln(p) \\
\ln(p) &= 2\ln(2) = \ln(4) \\
p &= 4
\end{aligned}
$$

Perhaps one reason for Bernoulli's attraction to logarithmic utility is the invariance of its answers to changes in scale. Suppose that we have solved for p in the utility equation:

$$
\begin{aligned}
u(p) &= \sum_{i}\pi_i u(x_i) \\
\ln(p) &= \sum_{i}\pi_i \ln(x_i)
\end{aligned}
$$

If we multiply each of the payoffs by a constant, making the transformation $x_i \to cx_i$, then, because the π_is sum to one, p will similarly be transformed to cp:

$$
\begin{aligned}
\ln(cp) &= \ln(c) + \ln(p) \\
&= \sum_{i}\pi_i \ln(c) + \sum_{i}\pi_i \ln(x_i) \\
&= \sum_{i}\pi_i (\ln(c) + \ln(x_i)) \\
&= \sum_{i}\pi_i \ln(cx_i)
\end{aligned}
$$

So the answer to the St. Petersburg Paradox is independent of the unit of currency.

Appendix C: Exponential Utility and Cumulants

In Section 6 we proved the cash value of the stochastic inflow X to an investor whose utility is exponential to be $\xi_\alpha(X) = \frac{\ln(M_X(\alpha))}{\alpha}$. And if X is normal with mean μ and variance σ^2, then $\xi_\alpha(X) = \frac{\ln\left(e^{\alpha\mu + \alpha^2\sigma^2/2}\right)}{\alpha} = \frac{\alpha\mu + \alpha^2\sigma^2/2}{\alpha} = \mu + \left(\frac{\alpha}{2}\right)\sigma^2$. But there is a generalization of the ξ formula to non-normal random variables.

The function $\psi_X(\alpha) = \ln(M_X(\alpha))$ is called the cumulant generating function of X [6:23]. The i^{th} derivative of this function evaluated at $\alpha = 0$ is called the i^{th} cumulant of X:

$$\kappa_i = \left.\frac{\partial^i \psi(\alpha)}{\partial \alpha^i}\right|_{\alpha=0} = \psi^{[i]}(0)$$

Cumulants have a pleasing property that most moments lack, viz., that the cumulant of a sum of independent random variables equals the sum of the cumulants of those random variables. For if $X = \sum_i X_i$ and the X_is are independent:

$$\begin{aligned}
\psi_X(\alpha) &= \ln(M_X(\alpha)) \\
&= \ln\left(M_{\sum_i X_i}(\alpha)\right) \\
&= \ln\left(\prod_i M_{X_i}(\alpha)\right) \\
&= \sum_i \ln(M_{X_i}(\alpha)) \\
&= \sum_i \psi_{X_i}(\alpha)
\end{aligned}$$

And so, $\kappa_j(X) = \psi_X^{[j]}(0) = \sum_i \psi_{X_i}^{[j]}(0) = \sum_i \kappa_j(X_i)$.

For convenience we define $\kappa_0 = \psi_X^0(0) = \psi_X(0) = \ln(M_X(0)) = \ln(1) = 0$. Then we can express ψ as a Maclaurin series:

$$\begin{aligned}\psi_X(\alpha) &= \sum_{i=0}^{\infty}\frac{1}{i!}\psi_X^{[i]}(0)\alpha^i \\ &= \sum_{i=0}^{\infty}\frac{1}{i!}\kappa_i\alpha^i \\ &= \sum_{i=1}^{\infty}\frac{1}{i!}\kappa_i\alpha^i\end{aligned}$$

Hence:

$$\begin{aligned}\xi_\alpha(X) &= \frac{\ln(M_X(\alpha))}{\alpha} \\ &= \frac{\psi_X(\alpha)}{\alpha} \\ &= \frac{\sum_{i=1}^{\infty}\frac{1}{i!}\kappa_i\alpha^i}{\alpha} \\ &= \sum_{i=1}^{\infty}\frac{1}{i!}\kappa_i\alpha^{i-1} \\ &= \kappa_1 + \frac{1}{2}\kappa_2\alpha + \frac{1}{6}\kappa_3\alpha^2 + \frac{1}{24}\kappa_4\alpha^3 + \ldots\end{aligned}$$

If μ_i represents the i^{th} central moment of X, it can be shown [6:24]:

$$\begin{aligned}\kappa_1 &= \mu \\ \kappa_2 &= \sigma^2 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\sigma^4\end{aligned}$$

Therefore:

$$\begin{aligned}\xi_\alpha(X) &= \kappa_1 + \frac{1}{2}\kappa_2\alpha + \frac{1}{6}\kappa_3\alpha^2 + \frac{1}{24}\kappa_4\alpha^3 + \ldots \\ &= \mu + \frac{1}{2}\sigma^2\alpha + \frac{1}{6}\mu_3\alpha^2 + \frac{1}{24}\left(\mu_4 - 3\sigma^4\right)\alpha^3 + \ldots \\ &\approx \mu + \frac{1}{2}\sigma^2\alpha\end{aligned}$$

We conclude, then, that the ξ function works for non-normal variables, that using just the mean and the variance yields a first-order approximation in α, and that using higher moments or cumulants will yield higher-order approximations.

Appendix D: Exponential Utility and Covariance

Section 6 concluded that if an entity's stochastic present value is X_1, and it considers a project that will immediately change that value to X_2, and the values can be treated as normal random variables, the project should be undertaken if and only if:

$$\mu_1 + \left(\frac{\alpha}{2}\right)\sigma_1^2 < \mu_2 + \left(\frac{\alpha}{2}\right)\sigma_2^2$$

This appendix will elucidate this conclusion, especially as to how utility-theoretic pricing accounts for the correlation of risks. The assumption of normal random variables is nearly unavoidable, since few multivariate density functions other than the normal specify correlation. Nevertheless, Appendix C demonstrated that normality, or consideration of just the mean and the variance, is a first-order approximation to the truth.

Let X_1 be normal with mean μ_1 and variance σ_1^2, and X_2 be normal with mean μ_2 and variance σ_2^2. Let the correlation coefficient between the two be ρ. Then the covariance between the two is $\rho\sigma_1\sigma_2$. If X_1 represents an entity's current wealth, or stochastic present value, the payment of p for asset X_2 will change its wealth to $X_1 - p + X_2$, which is normal with mean $\mu_1 - p + \mu_2$ and variance $\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2$. The entity's utility is exponential with risk tolerance α; so a comparison of its utilities before and after acquiring asset X_2 is:

$$\mu_1 + \left(\frac{\alpha}{2}\right)\sigma_1^2 \sim \mu_1 - p + \mu_2 + \left(\frac{\alpha}{2}\right)\left(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2\right),$$

where the tilde stands for some comparison relation. The following algebraic manipulations will preserve the comparison:

$$\mu_1 + \left(\frac{\alpha}{2}\right)\sigma_1^2 \sim \mu_1 - p + \mu_2 + \left(\frac{\alpha}{2}\right)\left(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)$$

$$0 \sim -p + \mu_2 + \left(\frac{\alpha}{2}\right)\left(2\rho\sigma_1\sigma_2 + \sigma_2^2\right)$$

$$p \sim \left(\mu_2 + \frac{\alpha}{2}\sigma_2^2\right) + \alpha\rho\sigma_1\sigma_2$$

The acquisition increases expected utility if $p < \left(\mu_2 + \frac{\alpha}{2}\sigma_2^2\right) + \alpha\rho\sigma_1\sigma_2$.

The expression $\mu_2 + \frac{\alpha}{2}\sigma_2^2$ represents the equilibrating price of an independent asset, denoted in Section 6 as $\xi_\alpha(X_2)$. A risk-averse entity, for whom $\alpha < 0$, will pay less than expected value (μ_2) for the asset. But correlation adds the term $\alpha\rho\sigma_1\sigma_2$ to the price. A risk-averse entity will pay less than $\mu_2 + \frac{\alpha}{2}\sigma_2^2$ for an asset that is positively correlated with its current wealth, but will pay more that that for a negatively correlated asset.

It will be instructive to derive the utility comparison more generally and more rigorously than was done in Section 6. Let **x** be a multivariate random variable with mean μ and variance Σ. If **x** is n–dimensional, then μ is an $n\times 1$ vector and Σ is an $n\times n$ matrix. Σ must be symmetric, and $\sigma_{ij} = Cov[X_i, X_j] = \rho_{ij}\sigma_i\sigma_j$. Σ must also be non-negative definite, which implies that if any diagonal, or variance, element (σ_{ii}) is zero, then all the elements of the row and the column intersecting that element must be zero (i.e., all the σ_{ij} and σ_{ji}

elements). The moment generating function of **x** is defined as $M_{\mathbf{x}}(\mathrm{t}) = E\left[e^{\mathrm{t'x}}\right]$, where t is an $n\times 1$ vector. This is analogous to the definition of the moment generating function of a scalar, and the moments of **x** are the partial derivatives of $M_{\mathbf{x}}(\mathrm{t})$ with respect to t, when evaluated at t = 0. Searle [24:44] proves that if **x** is multivariate normal with mean μ and variance Σ, then $M_{\mathbf{x}}(\mathrm{t}) = e^{\mathrm{t'}\mu + \frac{1}{2}\mathrm{t'}\Sigma\mathrm{t}}$. The proof is analogous to that for a scalar normal variable, involving nothing more than an n–dimensional completion of a square.

Now let **x** be a bivariate normal variable (n = 2), consisting of elements X_1 and X_2. Therefore, $E[\mathbf{x}] = \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, and $Var[\mathbf{x}] = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. And

$$M_{\mathbf{x}}(\mathrm{t}) = M_{\mathbf{x}}\left(\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\right) = e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}\left(t_1^2\sigma_1^2 + 2t_1t_2\sigma_{12} + t_2^2\sigma_2^2\right)}.$$

The utility comparison becomes:

$$E[u(X_1)] \sim E[u(X_1 - p + X_2)]$$

$$\left.\begin{matrix} E[X_1] & \alpha = 0 \\ E\left[\dfrac{e^{\alpha X_1} - 1}{\alpha}\right] & \alpha \neq 0 \end{matrix}\right\} \sim \left\{\begin{matrix} E[X_1 - p + X_2] & \alpha = 0 \\ E\left[\dfrac{e^{\alpha(X_1 - p + X_2)} - 1}{\alpha}\right] & \alpha \neq 0 \end{matrix}\right.$$

If $\alpha = 0$, the comparison reduces to $\mu_1 \sim \mu_1 - p + \mu_2$, or $p \sim \mu_2$. What is interesting is the comparison when $\alpha \neq 0$, which is continuous with the comparison when $\alpha = 0$:

$$E\left[\frac{e^{\alpha X_1}-1}{\alpha}\right]\sim E\left[\frac{e^{\alpha(X_1-p+X_2)}-1}{\alpha}\right]$$

$$\frac{1}{\alpha}E\left[e^{\alpha X_1}\right]-\frac{1}{\alpha}\sim\frac{1}{\alpha}E\left[e^{\alpha(X_1-p+X_2)}\right]-\frac{1}{\alpha}$$

$$\frac{1}{\alpha}E\left[e^{\alpha X_1}\right]\sim\frac{1}{\alpha}E\left[e^{\alpha(X_1-p+X_2)}\right]$$

$$\frac{1}{\alpha}E\left[e^{\alpha X_1}\right]\sim\frac{e^{-\alpha p}}{\alpha}E\left[e^{\alpha X_1+\alpha X_2}\right]$$

$$\frac{e^{\alpha p}}{\alpha}E\left[e^{\alpha X_1+0X_2}\right]\sim\frac{1}{\alpha}E\left[e^{\alpha X_1+\alpha X_2}\right]$$

$$\frac{e^{\alpha p}}{\alpha}M_{\mathbf{x}}\left(\begin{bmatrix}\alpha\\0\end{bmatrix}\right)\sim\frac{1}{\alpha}M_{\mathbf{x}}\left(\begin{bmatrix}\alpha\\\alpha\end{bmatrix}\right)$$

The comparison has not been affected by these manipulations, since none involved multiplication by a non-positive number. But the comparison cannot be multiplied by α, because α can be negative. We can continue:

$$\frac{e^{\alpha p}}{\alpha}e^{\alpha\mu_1+\frac{\alpha^2}{2}\sigma_1^2}\sim\frac{1}{\alpha}e^{\alpha\mu_1+\alpha\mu_2+\frac{\alpha^2}{2}\left(\sigma_1^2+2\sigma_{12}+\sigma_2^2\right)}$$

$$\frac{e^{\alpha p}}{\alpha}\sim\frac{1}{\alpha}e^{\alpha\mu_2+\frac{\alpha^2}{2}\left(2\sigma_{12}+\sigma_2^2\right)}$$

If α is positive, $e^{\alpha p}\sim e^{\alpha\mu_2+\frac{\alpha^2}{2}\left(2\sigma_{12}+\sigma_2^2\right)}$, $\alpha p\sim\alpha\mu_2+\frac{\alpha^2}{2}\left(2\sigma_{12}+\sigma_2^2\right)$, and finally, $p\sim\mu_2+\frac{\alpha}{2}\left(2\sigma_{12}+\sigma_2^2\right)$. And if α is negative, the relation commutes with each multiplication and division: $e^{\alpha\mu_2+\frac{\alpha^2}{2}\left(2\sigma_{12}+\sigma_2^2\right)}\sim e^{\alpha p}$, $\alpha\mu_2+\frac{\alpha^2}{2}\left(2\sigma_{12}+\sigma_2^2\right)\sim\alpha p$, and

$p \sim \mu_2 + \frac{\alpha}{2}(2\sigma_{12} + \sigma_2^2)$. Whether positive or negative, the outcome is the same. And this is continuous with the comparison when $\alpha = 0$: $p \sim \mu_2$.

So the entity's expected utility increases if and only if $p < \mu_2 + \frac{\alpha}{2}(2\sigma_{12} + \sigma_2^2)$. In Section 8 we modified the relation for the pricing of a liability. Liability X_2 is a negative asset, which makes **x** to have a mean of $\begin{bmatrix} \mu_1 \\ -\mu_2 \end{bmatrix}$ and a variance of $\begin{bmatrix} \sigma_{11} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{22} \end{bmatrix}$. And the entity charges a premium to assume X_2, which is a negative payment. Hence the comparison becomes:

$$-p \sim -\mu_2 + \frac{\alpha}{2}(-2\sigma_{12} + \sigma_2^2)$$

$$\mu_2 + \frac{1}{2}(-\alpha)(\sigma_2^2 - 2\sigma_{12}) \sim p$$

The expected utility increases if and only if $\mu_2 + \frac{1}{2}(-\alpha)(\sigma_2^2 - 2\sigma_{12}) < p$.

Recall that a negative α signals risk aversion. Thus, a risk-averse entity will charge more than the expected liability on the basis of positive amounts $(-\alpha)$ and σ_2^2. However, a positive covariance of the liability with the entity's current worth (σ_{12}) will serve to lower that premium. This is reasonable, for if the liability should turn out greater than expected, a risk-averse entity would prefer that it be coupled with greater-than-expected current worth X_1. Conversely, a negative covariance will raise the premium. For greater-than-expected liability tends to be coupled with less-than-expected current worth, which to a risk-averse entity is not offset by the coupling of less-than-expected liability with greater-than-expected current worth.

Let us return to a world of assets, understanding that liabilities can be fitted into this world as negative assets. The equilibrating price for new asset X_2 is

$$\begin{aligned} p_2 &= \mu_2 + \frac{\alpha}{2}\left(2\sigma_{12} + \sigma_2^2\right) \\ &= \mu_2 + \frac{\alpha}{2}\left(\sigma_{12} + \sigma_{21} + \sigma_{22}\right) \end{aligned}$$

Suppose that the entity has just passed from a non-stochastic balance sheet (c dollars in cash) by the acquisition of asset X_1. Since cash has zero variance, X_1 does not covary with it, and the equilibrating price for X_1 is $p_1 = \mu_1 + \frac{\alpha}{2}\sigma_{11}$. So if the entity acquires X_2, its stochastic balance sheet **x** is:

$[\mathbf{x}]$	$E[\mathbf{x}]$	$Var[\mathbf{x}]$		
$c-(p_1+p_2)$	$c-(p_1+p_2)$	0	0	0
X_1	μ_1	0	σ_{11}	σ_{12}
X_2	μ_2	0	σ_{21}	σ_{22}

If the assets were packaged, the equilibrating price would be:

$$\begin{aligned} p_{1\cup 2} &= \mu_1 + \mu_2 + \frac{\alpha}{2}\left(\sigma_{11} + \sigma_{12} + \sigma_{21} + \sigma_{22}\right) \\ &= \mu_1 + \frac{\alpha}{2}\sigma_{11} + \mu_2 + \frac{\alpha}{2}\left(\sigma_{12} + \sigma_{21} + \sigma_{22}\right) \\ &= p_1 + p_2 \end{aligned}$$

And the balance sheet would be:

$[\mathbf{x}]$	$E[\mathbf{x}]$	$Var[\mathbf{x}]$	
$c-(p_1+p_2)$	$c-(p_1+p_2)$	0	0
$X_1 \cup X_2$	$\mu_1+\mu_2$	0	$\sigma_{11}+\sigma_{12}+\sigma_{21}+\sigma_{22}$

Perhaps this seems equivalent to the first balance sheet; however, when the assets are acquired separately, the price of X_1 is adjusted only for σ_{11}, whereas the price of X_2 is adjusted for σ_{12}, σ_{21}, and σ_{22}. The price of X_2 has to take into account the fact that X_1 "got there first." X_2 has to pick up an "$\lrcorner$" shape of σ_{ij}s, as Meyers has shown [20:120]. Whether this benefits X_1 or X_2 depends on one's perspective; but order does affect the pricing of risk.

One might ask what the prices should be if the assets are acquired simultaneously, or without logical dependence. This really is equivalent to the allocation of the price of the whole to the parts. The answer is that the price should be averaged over all orders. In this simple example, p_1 and p_2 would each account for one of the covariances:

$$p_1 = \mu_1 + \frac{\alpha}{2}(\sigma_{11}+\sigma_{12})$$
$$p_2 = \mu_2 + \frac{\alpha}{2}(\sigma_{21}+\sigma_{22})$$

In a general problem with n assets and an $n \times n$ variance matrix Σ, each asset's price would account for a row (or, by symmetry, a column) of σ_{ij}s:

$$p_i = \mu_i + \frac{\alpha}{2}\sum_j \sigma_{ij}$$

This solution to the allocation problem was implicit in Gogol [11:363]. Mango derived it from game theory [19:41-46], and Halliwell from the linear statistical model [12:346-348]. This seems to be *the* solution to the allocation problem; but, as mentioned in

Section 2, it is not likely to win general acceptance. But, as stated in Section 5, if a block of risks were priced according to utility theory, the price of the block would be allocated to each risk according to this solution. The pricing of the blocks themselves would remain order dependent; and that is just how life is.

Appendix E: An Examination of One Sophisticated ROE Approach

Sections 2 and 3 criticized the ROE approach for its undesirable (or illogical and inconsistent) behavior. If a risk is allocated capital $c(t)$ at time t, and this capital should be rewarded with an expected rate of return of κ, then the risk-taker should receive $c(t)\kappa dt$ for taking risk during the time interval $[t, t+dt]$. If $v(t)$ represents the present value (at time 0) of one dollar to be received with certainty at time t, then the risk-taker can receive the whole risk charge up front at the amount $RC[0,T] = \int_{t=0}^{T} v(t)c(t)\kappa dt$.

The major defect of this approach is that $\lim_{T \to 0} RC[0,T] = 0$, a phenomenon that the paper calls "the evanescence of risk loads in the realm of the instantaneous." The simplest instantaneous risk is to bet on a coin that will be flipped in a few seconds. But Section 3 showed also that this approach makes the risk charge too large (even infinitely large) as T approaches infinity. The ROE approach does not relate properly with time. Again, Section 3 argued that as the loss is allowed more future in which to stretch out, the risk charge should decrease. For if the loss is going to happen, it is better that it happen farther into the future. Thus not only the present value of the loss but also the present value of the risk charge should decrease as the loss is allowed more future.

As a simple example, a coin will flip at some random time uniformly distributed over [0, T]. If heads results, a dollar will be paid to someone at that time. If this were the only risk in the portfolio, the risk-taker would be justified in setting $c(t)$ equal to one dollar. The ROE approach would have the risk-taker to collect a risk charge of κ each time period. So, present value aside, the expected loss is always half a dollar and the risk charge for a $[0, T]$ interval is κT. For realistic yield curves the risk-adjusted pure premium $RAPP[0, T]$ increases with T. And $RAPP[0, 0^+]$ always equals half a dollar. This revolts against our intuition that the risk-adjusted pure premium should decrease with T; and the utility-theoretic example of Section 8 confirmed our intuition.

One of the most sophisticated ROE approaches is that of Rodney Kreps [16]. The current net worth of an insurer has mean R and standard deviation S. The insurer considers underwriting a risk whose profit has mean r and standard deviation σ. And the correlation of the profit with the current net worth is C. Therefore, immediately after underwriting the risk, the net worth has mean $R' = R + r$ and standard deviation $S'^2 = S^2 + 2CS\sigma + \sigma^2$. According to the second footnote [16:197], "We take all values as present values." This must mean the usual present value (*sans* risk adjustment), since the aim is to calculate a risk load, and one should avoid double counting. The marginal net worth (i.e., change in net worth) has mean $R' - R = r$ and standard deviation:

$$\begin{aligned} S' - S &= (S' - S)\frac{S' + S}{S' + S} \\ &= \frac{S'^2 - S^2}{S' + S} \\ &= \frac{2CS\sigma + \sigma^2}{S' + S} \\ &= \frac{2CS + \sigma}{S' + S}\sigma \end{aligned}$$

Kreps then allocates capital V to the whole book of business as some multiple z of the standard deviation, giving as an example z = 3.1 for a 99.9% confidence level if the business is normally distributed [16:197]. He also says that the allocated capital may be reduced by the expected profit of the business; so his formulas are $V = zS - R$ and $V' = zS' - R'$. Therefore, the marginal capital for writing the new risk is:

$$
\begin{aligned}
V' - V &= (zS' - R') - (zS - R) \\
&= z(S' - S) - (R' - R) \\
&= z\frac{2CS + \sigma}{S' + S}\sigma - r
\end{aligned}
$$

The author disagrees that an insurer should be allowed to treat the expected profit of a risk as part of the capital allocated to the risk; but something greater is a stake here. Nonetheless, the author prefers the formulas $V = zS$, $V' = zS'$, and $V' - V = z\frac{2CS + \sigma}{S' + S}\sigma$.

The last element is y, what management deems to be the required expected return on the marginal capital. The expected marginal profit should equal the required expected return:

$$
\begin{aligned}
r &= y(V' - V) \\
&= y\left(z\frac{2CS+\sigma}{S'+S}\sigma - r\right) \\
r(1+y) &= yz\frac{2CS+\sigma}{S'+S}\sigma \\
r &= \frac{yz}{1+y}\frac{2CS+\sigma}{S'+S}\sigma \\
&= \mathrm{R}\sigma
\end{aligned}
$$

$\mathsf{R} = \frac{yz}{1+y}\frac{2CS + \sigma}{S' + S}$ is "the reluctance to take on risk" [16:198]. If the formulas for V and V' that the author prefers are used, the formula for the reluctance simplifies to $yz\frac{2CS + \sigma}{S' + S}$. The premium is $P = \mu + \mathsf{R}\sigma + E$ [16:199], where μ represents the present value of the expected losses and E the present value of the (non-stochastic) expenses.

This is a sophisticated ROE approach, and has attracted many actuaries (e.g., [11] and [14]). Kreps calculates under some reasonable circumstances a value for R of 0.33 [16:203], and it has become a rule of thumb among some actuaries that the risk load should be about a third of a standard deviation. Also attractive is that the approach considers the covariance of the new risk with the current portfolio.

In fact, it may surprise some that this approach is similar to a utility-theoretic answer, as we will now show. If we use exponential utility and consider just the first two moments of the probability distribution (as Kreps does), the comparison of utility before and after underwriting the new risk is (cf. Section 6 and Appendix D):

$$
\begin{aligned}
R + \frac{\alpha}{2}S^2 &\sim R + P - \mu - E + \frac{\alpha}{2}S'^2 \\
&\sim R + (\mu + RC + E) - \mu - E + \frac{\alpha}{2}S'^2 \\
&\sim R + RC + \frac{\alpha}{2}S'^2 \\
\frac{\alpha}{2}S^2 &\sim RC + \frac{\alpha}{2}S'^2 \\
\frac{1}{2}(-\alpha)(S'^2 - S^2) &\sim RC
\end{aligned}
$$

RC is what Section 3 called the risk charge, which Kreps calls the risk load and sets equal to Rσ. It is convenient to prefix α with a minus sign, since a risk-averse perspective is assumed ($\alpha < 0$), so $-\alpha$ is positive. We simplify:

$$\frac{1}{2}(-\alpha)\left(S'^2 - S^2\right) \sim RC$$
$$\frac{1}{2}(-\alpha)\left(S^2 + 2CS\sigma + \sigma^2 - S^2\right) \sim RC$$
$$\frac{1}{2}(-\alpha)\left(2CS\sigma + \sigma^2\right) \sim RC$$

Compare this with the Krep's equilibrating profit (or risk load):

$$\begin{aligned} r &= \frac{yz}{1+y}\frac{2CS+\sigma}{S'+S}\sigma \\ &= \frac{1}{2}\left(\frac{yz}{1+y}\frac{2}{S'+S}\right)\left(2CS\sigma+\sigma^2\right) \\ &= \frac{1}{2}\left(\frac{yz}{1+y}\frac{1}{\overline{S}}\right)\left(2CS\sigma+\sigma^2\right) \end{aligned}$$

$\overline{S}$ represents the average standard deviation of the portfolio, which will not change dramatically in the short term for a mature portfolio. So we see that $-\alpha$ and $\frac{yz}{1+y}\frac{1}{\overline{S}}$ (or $\frac{yz}{\overline{S}}$, as the author prefers) serve the same purpose.

This settles the debate as to whether risk loads should be proportional to standard deviation or to variance. Frequently the Kreps approach is cited in support of the standard deviation. But the formulation for the risk load Rσ is really quadratic with respect to σ, because $R = \frac{yz}{1+y}\frac{2CS+\sigma}{S'+S}$. Thus this approach really supports the variance. Only "in the very pessimistic case where $C = 1$" [16:199], in which case $S' = S + \sigma$, does Rσ become linear with respect to σ:

$$
\begin{aligned}
\mathrm{R}\sigma &= \frac{yz}{1+y}\frac{2CS+\sigma}{S'+S}\sigma \\
&= \frac{yz}{1+y}\frac{2(1)S+\sigma}{(S+\sigma)+S}\sigma \\
&= \frac{yz}{1+y}\frac{2S+\sigma}{2S+\sigma}\sigma \\
&= \frac{yz}{1+y}\sigma
\end{aligned}
$$

At this point the reader may be wondering why the author has fussed over utility theory when he has demonstrated its similarity with one of the most highly regarded ROE approaches. Just what is the big difference after all?

Recall the critique of the ROE approach that was summarily stated in the beginning of this appendix. It has been instructive to examine the ROE approach of Rodney Kreps. It has merit, and that merit draws it close to the utility-theoretic approach. Perhaps that rapprochement will make it easier for some actuaries to make the transition to utility theory. But the author mentioned above, when digressing on a subtle point, that "something greater is at stake here." That something is the problem of time, to which Sections 2 and 3 have drawn much attention.

Even the Kreps approach is unaware of the problem of time; the second footnote blithely dispensed with it ("We take all values as present values"). But it reasserts itself on considering the dimensions of the variables: R, S, and V are in dollars; C and z are unitless. Therefore, because $r = y(V' - V)$, y too must be unitless. Section 1 showed time to be of the essence of the ROE approach, i.e., that it depends on determining the proper *rate* of return, where the unit of rate is time^{-1}. But this y is a unitless return, rather than a rate of return. How would management determine y? Even if management had, perchance, the beta of insurance liabilities, the CAPM would yield a rate of return, say twelve percent *per year*. But there is no natural unit of time in this formulation.

Many would respond that there is a natural unit of time, viz., the policy term, normally one year. Therefore, if management targeted, for example, a rate of return of twelve percent per year, the y of the Kreps formulation would be 0.120. But all the dollar amounts, especially σ, are present valued, and it makes no difference to the present value of σ whether the risk is covered for one year, for two years, for six months, or for one instant. If the policy term were semiannual, y would be $1.120^{0.5} - 1.000 = 0.058$. If we had a colony on Mars, perhaps the natural Martian policy term would be 687 Earth days. Then y would be $1.120^{\left(\frac{687}{365}\right)} - 1.000 = 0.238$. And if the policy covered the next twenty-four hours, y would be $1.120^{\left(\frac{1}{365}\right)} - 1.000 = 0.000$. Thus the premium would vary for risks whose σ values are the same. Furthermore, the time to runoff makes no difference to σ.

Only Kreps' fifth footnote mentions time [16:198], " ... there are interesting questions with respect to the surplus flow needed to support the expected return of the book and of the contract, and the consequent internal *rate* of return." (author's italics) This means that we are back to square one, as described in the beginning of this appendix, having dealt neither with the evanescence of the risk loads in the realm of the instantaneous nor with the other inconsistencies of the ROE approach. The ROE approach does not relate properly with time; but the utility-theoretic approach does.