A Critique of Risk-Adjusted Discounting

Leigh J. Halliwell

Abstract

The adjustment of the discount function for risk has led actuaries and other financial decision-makers into a labyrinth, some of whose branches are risk-adjusted returns, capital allocation, and piecemeal risk loads. This paper will define a stochastic cash flow, and will prove that to free the pricing of such a cash flow from arbitrage one must adjust its probability measure, not the discount function. In a less theoretical vein, the paper will proceed to show from simple examples the inconsistency of risk-adjusting the discount function. Then it will draw out implications for actuarial practice, one of the most important being that there are risk elements for which premiums neither can nor should be loaded. Five appendices delve into these and related matters, and make clear that the ideas herein are mere beginnings.

1) Introduction

More than a dozen years ago Hans Bühlmann wrote of a three-stage evolution of actuaries: "The Actuary of the Second Kind, contrary to his colleague of the First Kind in life assurance, whose methods were essentially deterministic, had to master the skills of probabilistic thinking." [4:137] He predicted the evolution of an actuary of the "Third Kind," a financial actuary – a prediction that is proving to be accurate. However, today even the non-life actuaries of the "Second Kind" routinely perform the financial task of loading pure premiums for risk. But the most common method of risk loading, viz., risk-adjusted discounting, is a vestige of the deterministic thinking of the "First Kind," a vestige surviving not only in actuaries, but in financial decision-makers of all species. This paper shows that risk-adjusted discounting is inconsistent, and recommends that actuaries of the Second Kind catch up with the probabilistic thinking of risk-adjusted probability measures.

2) The Present Value of a Cash Flow

A *cash flow* is a function $C:\Re\rightarrow\Re$, where $\Re$ is the set of real numbers. The domain of the function represents time t; and $C(t)$ represents the cumulative amount of cash, or money, received before or at time t. Normally the present time is $t = 0$, and $C(t)$ is irrelevant for $t < 0$. The receipt of the amount x_1 of money at time t_1 is represented by the function:

$$C(t)=\begin{cases} k & t<t_1 \\ k+x_1 & t\ge t_1 \end{cases}$$

The constant k, in effect, the money received at time $-\infty$, is arbitrary. Moreover, since normally we are not concerned with the past, $\lim_{t\to 0^-} C(t)$ is irrelevant. What matters is not level of the function, but the change of the function, i.e., not $C(t)$, but $dC(t)$.

The definition of a cash flow as a function from $\Re$ into $\Re$ allows for continuous cash flows. For example, $C(t) = \rho t$ represents the constant reception of ρ units of money per unit of time. (On the consistency of the dimensions of an equation see Appendix A.) But since financial reality is discrete, and since Calculus can lead us from the discrete to the continuous, we can confine our attention here to discrete cash flows. A function that represents the reception of amounts x_i of money at times t_i is $C(t)=\sum_{t_i\le t} x_i$.

The cash flow of receiving amount x_1 at time t_1 and returning it in the next moment is:

$$C(t)=\begin{cases} 0 & t<t_1 \\ x_1 & t=t_1 \\ 0 & t>t_1 \end{cases}$$

This and even more pathological cash flows can be ruled out by assuming C to be continuous from the right, i.e., $\lim_{s\to t^+} C(s)=C(t)$. We will always assume this.

Cash flows can be combined. A linear combination of cash flows $C=\sum_i a_i C_i$ is the function from $\Re$ into $\Re$ such that $C(t)=\sum_i a_i C_i(t)$. Let U_s be the step function at s:

$$U_s(t) = \begin{cases} 0 & t < s \\ 1 & t \geq s \end{cases}$$

Confining our attention to discrete time, we may write the cash flow $C(t) = \sum_{t_i \leq t} x_i$ as the linear combination $\sum_i x_i U_{t_i}(t)$.

Define a present, or instantaneous, cash flow, as a cash flow $C(t)$ for which there exists an x such that $C(t) = xU_0(t)$. So every present cash flow is of the form:

$$C(t) = \begin{cases} 0 & t < 0 \\ x & t \geq 0 \end{cases}$$

Present cash flows, now identified with a dot, are easily ordered (cf. [14:18-24]):

Order	Symbol	Meaning
Indifference	$\dot{C}_1 \sim \dot{C}_2$	$\dot{C}_1(0) = \dot{C}_2(0)$
Preference	$\dot{C}_1 \succ \dot{C}_2$	$\dot{C}_1(0) > \dot{C}_2(0)$

Present value is an operator on cash flows, mapping cash flows to present cash flows. $PV[C(t)] = \dot{C}(t)$ indicates that the one using this operator is indifferent to the general cash flow C and the present cash flow $\dot{C}$, in symbols, $C \sim \dot{C}$. Present value allows for the ordering of general cash flows, the preference of C_1 to C_2 depending on the preference of $PV[C_1]$ to $PV[C_2]$, that in turn depending on the relation of $\dot{C}_1(0)$ to $\dot{C}_2(0)$.

The present value of a cash flow is not directly the price that one would pay for that cash flow; rather it is the present cash flow that makes for indifference. Of course, the price of $\dot{C}(t)$ must be $\dot{C}(0)$; otherwise, either the buyer or the seller would get something for nothing. But it is this reasonable demand for the conservation of value that links the price of cash flow C with $\dot{C}(0)$.

Just as reasonable is the demand that every present-value operator be linear, i.e., that for all cash flows $C_i(t)$ and factors a_i, $PV\left[\sum_i a_i C_i(t)\right] = \sum_i a_i PV[C_i(t)]$. But in discrete time, cash flow $C(t) = \sum_i x_i U_{t_i}(t)$. Thus:

$$\begin{aligned} PV[C(t)] &= PV\left[\sum_i x_i U_{t_i}(t)\right] \\ &= \sum_i x_i PV\left[U_{t_i}(t)\right] \end{aligned}$$

So linearity implies that a present-value operator is uniquely specified by how it operates on $U_s(t)$. But $PV[U_s(t)]$ is a present, or instantaneous, cash flow; thus for every s there exists some real number, $v(s)$, such that $PV[U_s(t)] = v(s)U_0(t)$. There is a one-to-one correspondence between present-value operators and functions $v:\Re\rightarrow\Re$. Hence:

$$\begin{aligned} PV[C(t)] &= \sum_i x_i PV\left[U_{t_i}(t)\right] \\ &= \sum_i x_i v(t_i) U_0(t) \end{aligned}$$

And Calculus will lead us into the realm of the continuous, where summation is replaced by integration (Stieltjes integration, if necessary; cf. [7:21] and [10:1]):

$$
\begin{aligned}
PV[C(s)] &= \lim_{n\to\infty}\sum_{i=1}^{n}\Delta C(t_i)v(t_i)U_0(s)\\
&= \int_{t=-\infty}^{\infty} v(t)dC(t)U_0(s)\\
&= \begin{cases} 0 & s<0\\ \int_{t=-\infty}^{\infty} v(t)dC(t) & s\ge 0\end{cases}
\end{aligned}
$$

Of course, the coupling of this result with the demand for the conservation of value implies that the price of cash flow $C(t)$ for one using this operator is $\int_{t=-\infty}^{\infty} v(t)dC(t)$. And in the normal case that $C(t)$ is constant for $t < 0$, i.e., $dC(t) = 0$: $\int_{t=-\infty}^{\infty} v(t)dC(t) = \int_{t=0}^{\infty} v(t)dC(t)$.

The linearity of present value leads of necessity to a discount function $v(t)$. Dimensional analysis (Appendix A) requires that $v(t)$ be unitless. Conservation of value demands that $v(0) = 1$. The universal preference to receive money sooner than later and to pay it later than sooner demands that $v(t)$ decrease monotonically. And since the reception of an amount of money at time $t+\Delta t$ is equivalent to the reception of it at time t as Δt approaches zero, $v(t)$ must be continuous. The reception of amount x at time t is equivalent to the reception of $xv(t)$ units of money now, at time zero. The reception of one unit of money at time t is preferable to the reception of zero units; therefore, $1v(t) > 0v(t) = 0$. So a discount function is everywhere positive, as well as continuous and decreasing.

There are functions that are continuous everywhere and differentiable nowhere; but they increase and decrease in every interval. It seems that a decreasing function such as $v(t)$

must be differentiable almost everywhere, i.e., that the set of *t* at which *v*(*t*) is not differentiable must be countable. So $v'(t)$ almost always exists, and when it exists it is negative. Its unit is 1/[time]. The forward rate, frequently a useful function, $\phi(t) = -\frac{v'(t)}{v(t)}$, exists if and only if $v'(t)$ exists. A division by zero is impossible, since *v*(*t*) is everywhere positive. Its unit also is 1/[time].

From a few reasonable axioms, we have developed the theory of the present value of a cash flow. Now we will extend it to the realm of uncertainty.

3) The Certainty-Equivalent Value of a Stochastic Cash Flow

A *stochastic cash flow* ***C*** is a function from a sample space Ω into the set of cash-flow functions. So each state ω of the sample space has a cash flow $C_\omega(t)$. If $C_\omega(t)$ is the same for all ω, $\boldsymbol{C}(t)$ is trivial; it behaves like a general cash flow, and certainty-equivalent value reduces to present value.

Moreover, let $\boldsymbol{U}_{s,\omega}$ be the step function at *s* in state ω of the sample space. Representing the reception of amount $x_{i,\omega}$ of money at time t_i in state ω as the function $x_{i,\omega}\boldsymbol{U}_{t_i,\omega}(t)$, we can write a discrete stochastic cash flow as $\boldsymbol{C}(t) = \sum_{\omega\in\Omega}\sum_{i} x_{i,\omega}\boldsymbol{U}_{t_i,\omega}(t)$.

Certainty-equivalent value is an operator on stochastic cash flows, mapping stochastic cash flows to present cash flows. $CEV[\boldsymbol{C}(t)] = \dot{C}(t)$ indicates that the one using this operator is indifferent to the stochastic cash flow $\boldsymbol{C}$ and the present cash flow $\dot{C}$, in symbols, $\boldsymbol{C} \sim \dot{C}$.

Certainty-equivalent value, like present value, should be a linear operator. Therefore, in the discrete case:

$$\begin{aligned} CEV[\boldsymbol{C}(t)] &= CEV\left[\sum_{\omega\in\Omega}\sum_{i} x_{i,\omega}\boldsymbol{U}_{t_i,\omega}(t)\right] \\ &= \sum_{\omega\in\Omega}\sum_{i} x_{i,\omega} CEV\left[\boldsymbol{U}_{t_i,\omega}(t)\right] \\ &= \sum_{\omega\in\Omega}\sum_{i} x_{i,\omega} v(t_i,\omega) U_0(t) \end{aligned}$$

So there is a one-to-one correspondence between certainty-equivalent operators and functions $v:(\Re\times\Omega)\rightarrow\Re$. The function $v(t, \omega)$ is a stochastic discount function. The one using this function is indifferent to the reception of one unit of money at time t in state ω and the reception of $v(t, \omega)$ units of money at time zero regardless of the state. As was the case with $v(t)$ and present value, so too $v(t, \omega)$ must be positive, decreasing, and continuous with respect to t.

The stochastic cash flow that represents the reception of one unit of money at time t regardless of the state is equivalent to a general cash flow. Its certainty-equivalent value must be the present value of the general cash flow. Therefore, for all t, $\sum_{\omega\in\Omega} v(t,\omega) = v(t)$. This suggests the usefulness of the function $\psi(t,\omega) = v(t,\omega)/v(t)$. So, for all t, $\sum_{\omega\in\Omega}\psi(t,\omega) = 1$ and $0 < \psi(t,\omega) \leq 1$. ψ is continuous also.

Now consider the certainty-equivalent value of $\boldsymbol{C}(t) = x_1\boldsymbol{U}_{t_1,\omega^*}(t) + x_2\boldsymbol{U}_{t_2,\omega^*}(t)$, a stochastic cash flow that is null except in one particular state ω^*, wherein amounts x_1 and x_2 are received at times t_1 and t_2:

$$\begin{aligned} CEV[\boldsymbol{C}(t)] &= CEV\left[x_1\boldsymbol{U}_{t_1,\omega^*}(t) + x_2\boldsymbol{U}_{t_2,\omega^*}(t)\right] \\ &= x_1 CEV\left[\boldsymbol{U}_{t_1,\omega^*}(t)\right] + x_2 CEV\left[\boldsymbol{U}_{t_2,\omega^*}(t)\right] \\ &= x_1 v(t_1,\omega^*)U_0(t) + x_2 v(t_2,\omega^*)U_0(t) \\ &= \left(x_1 v(t_1,\omega^*) + x_2 v(t_2,\omega^*)\right)\cdot U_0(t) \end{aligned}$$

If $x_1 = -1/v(t_1,\omega^*)$ and $x_2 = 1/v(t_2,\omega^*)$, then $CEV[\boldsymbol{C}(t)] = 0\cdot U_0(t) = 0$. In this case, the present values of all the state-dependent cash flows are zero, except for that of state ω^*:

$$\begin{aligned} PV\left[C_{\omega^*}(t)\right] &= \left(x_1 v(t_1) + x_2 v(t_2)\right)U_0(t) \\ &= \left(\frac{-1}{v(t_1,\omega)}v(t_1) + \frac{1}{v(t_2,\omega)}v(t_2)\right)U_0(t) \\ &= \left(\frac{-1}{\left(\dfrac{v(t_1,\omega)}{v(t_1)}\right)} + \frac{1}{\left(\dfrac{v(t_2,\omega)}{v(t_2)}\right)}\right)U_0(t) \\ &= \left(\frac{-1}{\psi(t_1,\omega)} + \frac{1}{\psi(t_2,\omega)}\right)U_0(t) \end{aligned}$$

But the one using this certainty-equivalent operator would be indifferent to $\boldsymbol{C}(t)$ and a null cash flow, the present value of whose cash flow in state ω^* is $0\cdot U_0(t) = 0$. If $\dfrac{-1}{\psi(t_1,\omega)} + \dfrac{1}{\psi(t_2,\omega)}$ were not zero, one of $\boldsymbol{C}(t)$ and the null cash flow would be preferable to the other, which would contradict the design of certainty equivalence. Therefore, for all times t_1 and t_2, $\dfrac{-1}{\psi(t_1,\omega)} + \dfrac{1}{\psi(t_2,\omega)} = 0$, i.e., $\psi(t_1,\omega) = \psi(t_2,\omega) = \psi(\omega)$. ψ is independent of time, and $v(t,\omega) = \psi(\omega)v(t)$.

We reach the important conclusion:

$$\begin{aligned} CEV[\boldsymbol{C}(t)] &= \sum_{\omega\in\Omega}\sum_{i} x_{i,\omega} v(t_i,\omega) U_0(t) \\ &= \sum_{\omega\in\Omega}\sum_{i} x_{i,\omega} \psi(\omega) v(t_i) U_0(t) \\ &= \sum_{\omega\in\Omega} \psi(\omega)\left(\sum_{i} x_{i,\omega} v(t_i) U_0(t)\right) \\ &= \sum_{\omega\in\Omega} \psi(\omega) PV[C_\omega(t)] \end{aligned}$$

Since $\sum_{\omega\in\Omega}\psi(\omega)=1$, the certainty-equivalent value of a stochastic cash flow is a weighted average of the present values of its state-dependent cash flows. $\psi(\omega)$ is called the "state price," in effect, what the one using the corresponding certainty-equivalent operator would pay at time zero for the immediate reception of one unit of money in state ω. The consistency of valuation implies the existence of a state-price vector ψ; and conversely, the existence of a state-price vector implies the consistency of valuation.

The function $\psi:\Omega\rightarrow\Re$ is a probability measure ([9:271] and [16:599]). It need not be the same as $\pi:\Omega\rightarrow\Re$, the true probability measure of Ω; in fact, usually it is not. The only qualification is that ψ must be "equivalent" to π ([9:272] and [16:600]), meaning that $\{\omega\in\Omega:\psi(\omega)=0\}=\{\omega\in\Omega:\pi(\omega)=0\}$. Above, we tacitly assumed that every state was possible, i.e., that the probability of every state was positive.

The interpretation of ψ as a probability measure allows us to express the certainty-equivalent value of a stochastic cash flow as the expectation with respect ψ of a random function:

$$CEV[\boldsymbol{C}(t)] = \sum_{\omega \in \Omega} \psi(\omega) PV[C_{\omega}(t)] = E_{\psi}[PV[C_{\omega}(t)]]$$

This underscores the linearity of certainty-equivalent value. Moreover, the certainty-equivalent value respects only the *present values* of the state-dependent cash flows, not the state-dependent cash flows themselves. Changing the state-dependent cash flows without changing their present values has no effect on the certainty-equivalent value.

The linearity of certainty-equivalent value bears on the collective risk model, $\boldsymbol{S} = \sum_{i=1}^{\boldsymbol{N}} \boldsymbol{X}_i$, where ***N***, the number of claims is a random variable. For $CEV[\boldsymbol{S}] = \sum_{i=1}^{\boldsymbol{N}} CEV[\boldsymbol{X}_i]$. Current actuarial practice is to obtain the aggregate loss distribution from nominal claim severity, and then somehow to discount it. The theory here shows that what matters is the certainty-equivalent value of the claims. And the distribution on which this depends is that of the present value of the claims, not that of their nominal value. When pricing risk one ought to deal with the economic realities from the beginning, rather than to ignore them throughout and to try to back into them at the end.

Finally, we note two properties of certainty-equivalent value by which we will test risk-adjusted discounting. First, adjusting the probability measure from π to ψ, or "tilting" the probabilities, allows one to reward some states and to punish others. But the certainty-equivalent value must still fall within the envelope bounded by the minimum and the maximum of $\{PV[C_{\omega}(t)]\}$. In particular, if all the state-dependent cash flows have the same present value, the certainty-equivalent value must be that value. Second and consequently, if stochastic cash flow $\boldsymbol{D}$ is defined such that $D_\omega(t) = C_\omega(t) + F(t)$, i.e., $\boldsymbol{D}$ is the combination of $\boldsymbol{C}$ and a non-stochastic cash flow, then $CEV[\boldsymbol{D}] = CEV[\boldsymbol{C}] + PV[F]$.

4) The Ætiology of Risk-Adjusted Discounting

We begin by citing three of the many assertions of risk-adjusted discounting.

> "Value is defined as the present value of a future cash-flow stream. The stream may have one or more cash flows. The present value is found by discounting the future flow or flows at a discount rate which reflects the riskiness of the stream." [2:82]

> "The standard criticisms of the NPV approach are that cash flows are uncertain, there may be different views as to the proper discount rate and projects are assumed to be independent. The first two criticisms are assumed to be resolved by the market process. Because cash flows are uncertain, they are discounted at a rate that reflects this uncertainty rather than at the risk-free rate." [6:21]

Of particular relevance to actuaries is the following excerpt from a Standard of Practice:

> "The key element in an actuarial appraisal is the projection of the future stream of earnings attributable to the evaluated business. … The projected earnings are then discounted at appropriate risk-adjusted rate(s) of return to derive value." [1:§3]

This Standard of Practice ends, as most do, with a provision for actuaries to differ from it, provided that they give "appropriate and explicit" statements of their differences. However, actuaries are reluctant to criticize their profession's standards, much less to differ from them. Nevertheless, standards are not dogmas; and the more recent *Principles Underlying Actuarial Science* [5], by its avoidance of the concept of risk-adjusted discounting, may evidence the uneasiness of some actuaries with the concept.

Language like that of these citations, language of a rate that "reflects" risk, or that is "appropriate to" or "commensurate with" risk and uncertainty, sounds natural and innocuous. It is the language that most executives, board members, investors, analysts, and regulators expect to hear. Few pause to consider whether a rate of return is suited to the task of reflecting risk. The second citation ("*Because* cash flows are uncertain, they are discounted at a rate … ") contains an unguardedness that the next section will expose. But for the rest of this section we will seek to understand why many find it so natural, so inevitable, to apply risk-adjusted discounting to the valuation of stochastic cash flows.

Using the notation of the previous section, we let π_ω be the probability of state ω, and $x_{i\omega}$ be the amount received at t_i in state ω. And let $v^*(t)$ represent the risk-adjusted discount function, normally but not necessarily an exponential function of a risk-adjusted discount rate: $v^*(t;r) = (1 + r \cdot 1)^{-t/1}$. (The ones assure dimensional consistency; cf. Appendix A.) The authors cited above believe that the certainty-equivalent value of this stochastic cash flow results from the choice of an appropriate discount function v^* (normally the choice of a rate r) in the equation:

$$\begin{aligned} CEV[\boldsymbol{C}(t); \pi, v^*] &= \sum_\omega \sum_i \pi_\omega x_{i,\omega} v^*(t_i) U_0(t) \\ &= \sum_\omega \pi_\omega \left(\sum_i x_{i,\omega} v^*(t_i) U_0(t) \right) \\ &= \sum_\omega \pi_\omega PV[C_\omega(t); v^*] \end{aligned}$$

The formula derived in the last section is $CEV[\boldsymbol{C}(t); \psi, v] = \sum_\omega \psi_\omega PV[C_\omega(t); v]$. The difference is clear: the formulas agree that either the probability measure or the discount function must change; they disagree as to which.

We will understand the appeal of risk-adjusted discounting if we interchange the summations:

$$\begin{aligned}CEV\left[\boldsymbol{C}(t);\pi,v^*\right] &= \sum_{\omega}\sum_{i}\pi_\omega x_{i,\omega}v^*(t_i)U_0(t)\\ &= \sum_{i}\left(\sum_{\omega}\pi_\omega x_{i,\omega}\right)v^*(t_i)U_0(t)\\ &= \sum_{i}E_\pi[\boldsymbol{X}_i]v^*(t_i)U_0(t)\end{aligned}$$

Risk-adjusted discounting requires only the expected cash flow according to the real probability measure. Decision-makers are comfortable with guessing the expected cash flow, even those who are not mathematically sophisticated. And guessing a rate r does not trouble them, especially since comparable investments, the CAPM, and sensitivity testing offer guidance. Risk-adjusted discounting moves with the inertia of the world: accountants and lawyers understand it; consultants advise about it; governments tax and depreciate in accordance with it; everyone seems happy with decisions made on the basis of it. Above all, it is doable. Can that be said about adjusting the probability measure from π to ψ? Nevertheless, as the following section will show, risk-adjusted discounting is riddled with problems. Eventually, decision-makers will abandon it.

5) Inconsistencies of Risk-Adjusted Discounting

The six inconsistencies of this section are roughly in order of ascending seriousness. Others (e.g., [11:352-354] and [18:137]) have noted some of these inconsistencies.

First, the forward rate $\phi(t)$ is not constant. So why should the risk-adjusted forward rate $\phi^*(t)$ be constant, as the form $v^*(t;r)=(1+r\cdot 1)^{-t/1}$ implies? One might suggest that $\phi^*(t)$ should be $\phi(t)$ plus a constant. But why too should this be? Convenience is the only reason. If the discount function is to be adjusted for risk, no one knows the true adjustment and no one knows whether the error of opting for convenience is tolerable or serious. Presumably, the steeper the yield curve, the more serious the error of a constant $\phi^*(t)$ would be.

Second, what should be the price of an asset that will pay $1/v(\boldsymbol{T})$ at time $\boldsymbol{T}$, where $\boldsymbol{T}$ is a random variable? The present value of every state-dependent cash flow is $v(t) \cdot 1/v(t) = 1$. Therefore, the certainty-equivalent value must be one unit of money. Most who believe in risk-adjusted discounting, regarding this stochastic cash flow as risky, would insist on a price of less than one unit.

Third, the limit of risk-adjusted discounting as time approaches zero is risk-neutrality. We will demonstrate this with a coin-flip example, an example that will be used in the three following points, and whose idea comes from [6:23]. In this example, at time T an asset will pay either \$120 or \$80, the amount depending on the outcome of the flip at time T of an unbiased coin. As T approaches zero, risk-adjusted discounting yields the price:

$$\lim_{T\to 0}\left(\frac{1}{2}\cdot \$120 + \frac{1}{2}\cdot \$80\right)v^*(T;r) = \lim_{T\to 0}\$100(1 + r\cdot 1)^{-T/1} = \$100$$

The reason for this undesirable property is that $\lim_{T\to 0} v^*(T;r) = v^*(0;r) = 1$, whatever r may be. Imagine that the coin will be tossed and the money paid tomorrow. An investor might be willing to pay \$95.54 today for tomorrow's outcome, whose expectation is \$100. Is it meaningful to say that the investor is seeking an expected rate of return of

4.67 percent per day (1.74×10^9 percent per year)? Would he choose the rate $r = 1.74 \times 10^9$ percent per year, and calculate $\$100(1 + r \cdot 1)^{-(1/365)/1} = \$95.54$? What else could he do, if according to the first citation, "Value … is found by discounting … at a rate which reflects the riskiness of the stream?"

Fourth, at the other end of the time line, as T grows large, the prices of cash flows can break out of their envelopes (cf. Section 3). In the coin-flip example, risk-adjusted discounting will set the price at $P(T) = \$100(1 + r \cdot 1)^{-T/1}$ for some r. The true price should be within the envelope of the present values of the minimal and the maximal outcomes, i.e., $\$80v(T) < P(T) < \$120v(T)$. If r_f = 7 percent per year and r = 12 percent per year, as in [6:23], then at T = 5 years:

$$\begin{aligned} P(5) &= \$100(1 + 12\% \cdot 1)^{-5/1} \\ &= \$56.74 \\ &< \$57.04 \\ &= \$80(1 + 7\% \cdot 1)^{-5/1} \\ &= \$80v(5) \end{aligned}$$

This fact is rarely appreciated, because the minimal outcome of the typical binomial example is zero. But just as risk-adjusted discounting adjusts too little for small T, it adjusts too much for large T. In contrast, a price that is a weighted average of the outcomes maintains at all times the same position within the envelope (Appendix D).

Fifth, the coin-flip asset appreciates at r per time period, whereas a risk-free asset appreciates only at r_f per time period. One should short the risk-free asset and put the proceeds into the risky asset in order to guarantee a profit of $r - r_f$ per time period on a net investment of zero. One might object that it is not certain that the asset will appreciate at r per time period. However, if no new information is gained as time passes (and none would be gained on the coin flip), an investor loyal to risk-adjusted discounting has no reason not to buy and sell at the price $P(T) = \$100(1 + r \cdot 1)^{-T/1}$. To use a physical analogy, if r_f is like the speed of light, risk-adjusted discounting allows for risk-free faster-than-light travel. In contrast, a price that is a weighted average of the outcomes appreciates in step with a risk-free asset (Appendix D).

And sixth, by a simple repackaging of the asset one can trick risk-adjusted discounting into yielding risk-neutral prices. Let $x_{i,\omega}^{*} = x_{i,\omega} - E_{\pi}[\boldsymbol{X}_i]$. Then $\boldsymbol{C}$ can be decomposed into $\boldsymbol{C}_1$, consisting of the amounts $x_{i,\omega}^{*}$, and $\boldsymbol{C}_2$, consisting of the amounts $E_{\pi}[\boldsymbol{X}_i]$. Now, $E_{\pi}[\boldsymbol{X}_i^{*}] = E_{\pi}[\boldsymbol{X}_i - E_{\pi}[\boldsymbol{X}_i]] = 0$. Therefore, however one adjusts the discounting function, $CEV[\boldsymbol{C}_1(t); \pi, v^{*}] = \sum_i E_{\pi}[\boldsymbol{X}_i^{*}] v^{*}(t_i) U_0(t) = \sum_i 0 \cdot v^{*}(t_i) U_0(t) = 0$. And since $\boldsymbol{C}_2$ is a non-stochastic cash flow, its value must be $\sum_i E_{\pi}[\boldsymbol{X}_i] v(t_i) U_0(t)$, as per Section 3. If risk-adjusted discounting is to be arbitrage-free, its only choice for $v^{*}(t)$ is $v(t)$:

$$\begin{aligned}
CEV[\boldsymbol{C}(t); \pi, v^{*}] &= CEV[\boldsymbol{C}_1(t) + \boldsymbol{C}_2(t); \pi, v^{*}] \\
&= CEV[\boldsymbol{C}_1(t); \pi, v^{*}] + CEV[\boldsymbol{C}_2(t); \pi, v^{*}] \\
&= 0 + \sum_i E_{\pi}[\boldsymbol{X}_i] v(t_i) U_0(t) \\
&= \sum_i E_{\pi}[\boldsymbol{X}_i] v(t_i) U_0(t)
\end{aligned}$$

For example, one could split the coin-flip asset into two components, one whose payment at time T is \$100 regardless of the flip, and the other whose payment at time T is (\$20) or \$20, depending on the flip. The price of the first component is \$100$v(T)$. The price of the second component is $\left(\frac{1}{2}\cdot(\$20)+\frac{1}{2}\cdot\$20\right)v^{*}(T;r)=\$0\cdot v^{*}(T;r)=\$0$, regardless of v^{*}. If the price of the whole asset is different from $\$100v(T)$, it will be possible to arbitrage.

'Inconsistency' is a euphemism for 'contradiction.' These six points are not like Zeno's paradoxes, which argued that change is illusory. The ancients had no convincing solutions for the paradoxes; but most continued to believe that change was real, that Zeno was a bit of a sophist, and that eventually someone would solve them. (They weren't solved until the set theory and real analysis of the nineteenth century.) These inconsistencies are not sophistries; they are to the point and are fatal to risk-adjusted discounting. Even Stephen D'Arcy, from whose idea the coin-flip example comes, senses a problem:

> "[The Certainty-Equivalent Value] is the same as the Present Value when discounted for both risk and the time value of money simultaneously. The advantage of the Certainty-Equivalent method is that the risk adjustment and the time value of money adjustment are separated, rather than combined. This makes the adjustments easier to understand and usable in situations where the combined method is not feasible." [6:23]

He tries quietly to qualify what he has said earlier: "*Because* cash flows are uncertain, they are discounted at a rate … " One might object that these six inconsistencies describe situations in which risk-adjusted discounting is not feasible. But to object thus binds one to specify the situations in which each method is feasible.

6) Implications

Section 3 established that the probability measure is what should be adjusted for risk. Section 5 showed into what inconsistencies adjusting the discount function leads. The nine following implications of these results are truly radical; they uproot much of the current financial theory and practice.

First, the price of an asset or a deal is different from the cash flow – as different as subject is from object. Since price depends on cash flow, it is circular to base price on an estimate of the price at which someone else (usually “the market”) will buy the asset or assume the deal. Current financial theory, abetted by risk-adjusted discounting, recommends for us to view the price of the S&P 500 stock index as a Wiener process whose expected growth rate is several hundred basis points greater per year than the growth rate of treasury securities. This view disposes one to accept a form of the “Greater Fool” fallacy, viz., that whether stocks are overpriced or underpriced is unimportant (really, meaningless), since their prices a year from now will probably be twelve to fifteen percent higher and in the long run we are virtually assured of making more money in stocks than in treasuries. Prognosticating the future price of a stock is no substitute for valuing the certainty equivalent of its stochastic dividend flow.

Second, rate of return, often confused with return itself (Appendix A), is ill suited to measure the value of a deal. We illustrate this with the coin-flip asset of the previous section [6:23]. Suppose that we decide to buy the $120/$80 coin-flip asset, to be determined and paid in one year, for $89.29. Suppose also that $93.46 now buys $100 one year from now. A report to management according to common financial theory might read, “The expected payoff in one year is $100. Staff priced this asset to an expected rate of return of twelve percent per year, whereas the risk-free rate is seven percent per year. Staff deems the extra five percent as sufficient to compensate for the swing of ± $20." But the theoretically correct report might read, "The present value of the outcomes is $112.15 and $74.77, and the expected present value is $93.46. So purchasing the asset for $89.29 means that the net worth of the company is either $22.86 greater or $14.52 less. The company won't know the outcome for one year; but the expected net worth now is $4.17 greater. Staff deems this $4.17 to compensate for the swing of ± $20. Staff titled the probabilities from 50%/50% to 38.84%/61.16% to arrive at the price of $89.29." The author suspects that many years must pass before most managements will prefer the second report.

Third, capital allocation is a hindrance to correct pricing. The following citation from Frank Pierson would be seconded by most financial decision-makers:

> "It's easy to say that a company prices its products to achieve at least a minimum rate of return on equity. In reality, it is not that easy. The hardest part is to allocate capital to individual contracts." [17:61]

So far, no one has presented capital allocation as a means for adjusting a probability measure; on the contrary, it is hard to imagine a connection. But unless it is somehow a means for adjusting probability measures it is at best a distraction from correct pricing, and at worst dangerous thereto. (On the possible dangers cf. [12:73-83].) The citation implies that capital allocation fits naturally with risk-adjusted discounting; but even this is doubtful. No capital was allocated when the coin-flip asset was discounted at twelve percent per year. But there is a fifty percent chance that the purchase of the asset for $89.29 will lose $9.29 next year. Does this call for a subsequent allocation of capital and a capital charge? Then the price of $89.29 would be too high. Moreover, the allocation of capital introduces cash flows into the stochastic flow that serve only to adjust its level.

Pierson himself acknowledges this with his insight that an insurance policy is a “reverse” investment [17:62]. But the sixth inconsistency of Section 5 showed how changing the level of a stochastic cash flow waylays risk-adjusted discounting. Finally, many argue that the risk of a transaction’s loss is like the risk of a bond’s default, and that capital allocation is the vehicle for repairing the default. But do the bond traders themselves include capital allocation in their pricing exercises? If they need not to allocate capital, why must others?

Fourth, (rate of) return on equity is a poor measure of performance. This follows from the first three implications; but it deserves emphasis. If a company invested $100 of its equity, and a year later received back $110, it would be better to say that the company made $10 than to say that its ROE was ten percent per year. For suppose that the company had invested $200, received back $100 almost immediately, and received $110 a year later. The company still made $10; but its ROE is unclear. Furthermore, management needs to rethink its ROE targets. Have the managers that set a goal of a twenty-percent return over the next year, when risk-free investments would give the company seven percent, evaluated whether the risks or chances expected to resolve over that year are up to the remaining thirteen? If not, then of what value are their goals?

Fifth, too much is made of the charge that insurers are overcapitalized. Indeed, many insurers have “surplus” surplus, i.e., more assets than needed for a high degree of solvency. But all money put into treasury securities is put to good use (Appendix A). The only complication is the double taxation of corporate earnings, once to the corporation and twice when distributed to its shareholders. (And double taxation is mitigated by the dividends-received deduction for corporate shareholders.) In other words, if one were content to invest in treasuries, one might have little reason for incorporating. One incorporates to protect oneself from bankruptcy, and double taxation is the "insurance premium" that the government charges for this protection. Whether it is a fair premium who can say? But one who goes to the trouble of incorporating should make sure that the corporation engages in enough risky activities to justify the double taxation. So the only drawback of overcapitalization is that shareholders may pay more taxes than needful.

Sixth, the theory of risk-adjusted probability measures trivializes opportunity cost. Suppose that you paid a fair price for a deal and are tapped out of cash. Then you hear of another opportunity. If your first deal is liquid you should be able to get your money back, less some frictional cost. Liquidity and frictional costs are the problems, not the possibility of missing an opportunity. And the more efficient a market is, the more rapidly and inexpensively one can deal in it. The unspoken fear is that the new opportunity will be a windfall and will be recognized as such by others, who will either buy it or require you to pay dearly to undo your deal. However, to lessen the chance of missing a windfall is not a sufficient reason to reduce an otherwise fair price. Moreover, one's tolerance for risk may already impound such fears. Appendix E elaborates on the subject of loads for various risks, risks that are more psychological than economic.

Seventh, there are risks that do not deserve risk loads. If the discount function does not already impound the risk, and the probability measure cannot be adjusted for it, then pricing cannot deal with the risk. For example, suppose that insurers A and B are similar except that A owns treasuries only, whereas B is twenty-percent in stocks. Many would recommend that B allocate capital for the risk of a stock-market decline. The pricing otherwise being equal, B's rates would be higher than A's. But who would knowingly pay more (or even the same) for B's policies? No, the solution is not for B to charge higher rates than A, but for B to write less insurance than A. Hence, although an insurer ought to manage all its risks, some risks are managed through decisions of quantity, not of price.

Eighth, for those risks that do not deserve risk loads, e.g., the risks of insolvency and of aimless management, dynamic financial analysis is helpful. But many hail DFA as the flowering of economic theory, and bill it as the most powerful tool for capital allocation and pricing. The theory presented herein belies these claims. Probably DFA can incorporate this theory, and eventually will do so. But due to the multivariateness of the financial universe (Appendix C), pricing must involve more than the tilting of the equiprobable scenarios of a DFA run.

And ninth, insurance data-collection and accounting arose out of the financial theory that they now perpetuate. Insurers and bureaus should collect transactions and cash flows in addition to or instead of summarized exposures, premiums, and loss triangles. As Section 3 mentioned in connection with the collective risk model, nothing is special about nominal values. And the International Accounting Standards Committee is devising a more economic form of insurance accounting [13]. For example, it recommends for all assets and liabilities be recorded at fair/market values (implying present-valued loss reserves), for every change in these values to flow into the income statement, and for profits to be booked and revised from policy inception (rather than to be earned over policy terms). Nevertheless, the committee continues to view risk-adjusted discounting as a worthwhile means of determining fair value.

This list is not exhaustive; and the necessity of some implications is debatable. But one cannot gainsay that the effects of a critique of risk-adjusted discounting are far-reaching, and touch the interests of many.

7) Conclusion

As the title suggests, the author has been more critical than constructive. All agree that *something* must be adjusted for risk. It's just that many haven't realized that that something might be other than a discount function, viz., a probability measure. It is easy and entertaining to poke holes in risk-adjusted discounting; but at present little is known about how to adjust probability measures. Theory shows that adjust them we must; but until breakthroughs occur – through utility theory, distribution transformation, or otherwise – many will content themselves with the *status quo*. Thus it has ever been in the evolution of science.

References

[1] Actuarial Standards Board, *Actuarial Standard of Practice No. 19: Actuarial Appraisals*, 1991.

[2] Bowlin, Oswald D., Martin, John D., and Scott, David F., *Financial Analysis*, Second Edition, New York, McGraw-Hill, 1990.

[3] Bühlmann, Hans, "An Economic Premium Principle," *ASTIN Bulletin*, 11(1), 1980, 52-60.

[4] Bühlmann, Hans, "Actuaries of the third kind?," *ASTIN Bulletin*, 17(2), 1987, 137-138.

[5] Casualty Actuarial Society and Society of Actuaries, *Principles Underlying Actuarial Science*, 1999.

[6] D'Arcy, Stephen P., "Introduction to the Discounted Cash Flow Approach," *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 19-25.

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

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

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

[10] Gerber, Hans U., *An Introduction to Mathematical Risk Theory*, Philadelphia, S. S. Huebner Foundation, 1979.

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

[12] Halliwell, Leigh J., "ROE, Utility, and the Pricing of Risk," *1999 CAS Spirng Reinsurance Call Papers*, 71-136.

[13] Hulse, Tony, "A Calculated Move," *The Review*, April 2000, 29-30.

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

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

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

[17] Pierson, Frank D., “Rate of Return,” *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 61-76.

[18] Van Slyke, Oakley E., “The Cost of Capital: An Axiomatic Approach,” *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 135-164.

[19] Venter, Gary G., “Premium Calculation Implications of Reinsurance without Arbitrage,” *ASTIN Bulletin*, 21(2), 1991, 223-230.

[20] Venter, Gary G., “Discussion of Dr. Shaun Wang’s Implementation of PH-Transforms in Ratemaking,” *PCAS* LXXXV (1998), 980-988.

[21] Wang, Shaun, “Implementation of PH-Transforms in Ratemaking,” *PCAS* LXXXV (1998), 940-979.

Appendix A: Dimensional Analysis and the Mechanics of Money

Physicists are trained to analyze every equation for dimensional consistency. For an equation to be dimensionally consistent each of its terms must have the same dimensionality. One can add pounds and kilograms, since the dimension of both is mass, and units of the same dimension are convertible. But one cannot add pounds and meters, for mass and length are different dimensions. The fundamental dimensions of mechanics are mass, length, and time. Dimensions can be multiplied and divided, e.g., velocity is the quotient of length and time.

Financial equations also must be dimensionally consistent. Money and time are the fundamental dimensions of finance. To give \$100 and to receive back \$110 is to have a return of (\$110–\$100)/\$100 = 0.10. The dimensions of the denominator cancel those of the numerator, so return is dimensionless. But to give \$100 and to receive \$110 after one year is to have a rate of return of 0.10 / 1 year, or 0.10 per year. The dimension of rate of return is time^{-1}. The failure to distinguish return from rate of return can lead to the treatment of some unit of time (usually a year) as "natural," and to financial mistakes ([12:133-135]). Checking for dimensional consistency is the corrective.

The exponential function $e^{x} = 1 + x + x^{2}/2 + x^{3}/6 + \ldots$ involves the integral powers of *x*. Dimensional consistency demands that its argument be dimensionless. The dimension of the force of interest ρ is time^{-1}; therefore, force of interest can appear in an exponential function only when multiplied by a time variable, e.g., $e^{\rho t}$. Similarly, both the base and the exponent of the commonly written formula $(1+i)^t$ must be dimensionless. Hence, its true form is really $(1+i\cdot 1_t)^{t/1_t}$, where 1_t means one unit of time.

Probability is dimensionless, as is the cumulative distribution function F. Therefore, the integral $\int g(x)dF(x)$ has the dimension of $g(x)$. The dimension of the derivative of g with respect to x is the quotient of the dimension of g and the dimension of x.

Many chide utility theory (Appendices B and C) for its unit, the "util." But one can, and perhaps should, define utility functions whose dimension is monetary. For example, a useful form of exponential utility is [12:91]:

$$u(x;\alpha)=\begin{cases}\dfrac{e^{\alpha x}-1}{\alpha} & \alpha\neq 0\\ x & \alpha=0\end{cases}$$

One can see that if $\alpha = 0$ the dimension of u is that of x, viz., monetary. But αx, being the argument of an exponential function, must be dimensionless. Therefore, the dimension of α is money^{-1}. So even when $\alpha \neq 0$ the dimension of u is money. The util in this case is a unit of money. One who understands these principles can verify that all the equations of this paper are dimensionally consistent.

Dimensional analysis reveals a common notion of rate of return, viz., the mechanistic notion that money does work. One hears, for example, that an investment that yields seven percent "works harder" than one that yields six percent. Or, in July 2000 the asking prices for $100 of treasury STRIPs (cf. Appendix E) maturing in five, ten, and twenty years were $73.20, $53.26, and $29.12. This implies that their yields to maturity were 6.44, 6.50, and 6.36 percent per year. Is it legitimate to say that money invested in the ten-year STRIP works harder than money invested in the five- and twenty-year STRIPs? May one conclude that the ten-year STRIP is the best investment of the three?

Nothing like the dimension of length in the mechanical world exists in the financial world. But we can analogize from the mechanical world to the financial according to the mapping:

$$\begin{aligned} \text{mass} &\to \text{money} \\ \text{length} &\to \text{time} \\ \text{time} &\to \text{time} \end{aligned}$$

The dimension of velocity is length/time; so the dimension of the financial analogue of velocity is time/time, or dimensionless. The dimension of acceleration is length/time^2; hence, the dimension of its financial analogue is time^{-1}. Thus, the financial analogue of acceleration is rate of return, or rate of growth. Force equals mass times acceleration; therefore, the unit of financial force is money/time. This is sensible: twice the financial force can produce either twice the money in the same time or the same money in half the time. Mechanical work equals force times distance; therefore, the unit of financial work is (money/time)×time = money. Again, this makes sense: the financial work is the making of money. Finally, power is the derivative of work with respect to time, so its unit is the unit of force times length/time. So in the financial world, the dimension of power is that of force; in fact, financial power and financial force are the same.

Let $m(t)$ represent the amount of money at time t. It is a happy coincidence that it m suggests the mechanical counterpart, mass. Let $\phi(t)$ represent the force of interest (more exactly, the forward rate of Section 2). Financial force through time does the work of making money grow, the growth being described by the differential equation $dm(t) = \phi(t)m(t)dt$. To increase ϕ is the meaning of such language as "making one's money work harder." The adverb 'harder' even has the meanings 'with more force' and 'with more power'.

And even the dimensionless financial velocity makes some sense. For the solution of the differential equation is $m(t) = m(0)e^{\int_{u=0}^{t}\phi(u)du} = m(0)/v(t)$, where v is the discount function. The dimensionless v (cf. Section 2) is somewhat analogous to velocity. So the mechanical analogy works well.

However, the analogy has mislead financial theory into looking for risk-adjusted discount functions, functions according to which money "works harder." In the example above, it is doubtful that money works harder in the ten-year STRIP. Investors in July 2000 were indifferent to paying \$73.20 to receive \$100 five years later and paying \$53.26 to receive it ten years later. They had reasons for their indifference to paying these amounts, especially reasons that respected their beliefs about the future of interest rates (Appendix E); but they were not asking themselves whether their money would be working hard enough at these amounts.

To the extent that money works, it works at the risk-free forward rate $\phi(t)$. The forward rate ϕ is like a variable speed of light, as explained in the fifth inconsistency of Section 5. One can say neither that the money invested in the coin-flip asset of Section 5 worked hard if the $120 is paid, nor that it idled if the $80 is paid. Luck, whether good or bad, is not the work of money. If the money works during the time of uncertainty, it works only according to the risk-free forward rate, as shown in Section 5 and Appendix D. In sum, the notion that money works in non-stochastic cash flows is doubtful; the notion that it works in stochastic cash flows is erroneous.

Appendix B: Equilibrium Pricing

The fundamental theorem of asset pricing states that a system of payouts and prices is consistent, or arbitrage-free, if and only if there are state prices ψ_ω such that the price of every asset is the average of its payouts weighted according to the state prices. In symbols, let $d_{i,\omega}$ be the payout (or, '*d*' for dividend) of one unit of the i^{th} asset in state ω, and let q_i be the price of one unit of the i^{th} asset. The state ω is realized immediately, and the payouts $d_{i,\omega}$ are received immediately. Even if the payouts are cash flows extending into the future, Section 3 shows that they can be collapsed into the present by defining $d_{i,\omega}$ as $PV[C_{i,\omega}(t)]$. One can find proofs of the fundamental theorem in [9:3-4], [12:104-107], and [16:524-526].

But equilibrium is another road to the existence of state prices. The utility of an investor whose portfolio is in equilibrium is maximized. No reallocation of assets will increase the utility. Here follows a simple proof of state-price existence; other and more elaborate proofs are in [8:122], [9:5-8], and [16:147-149].

Suppose that an investor whose current wealth is w has a utility function $u(w)$. The investor is a price-taker, accepting the price q_i per unit of the i^{th} asset, which pays $d_{i,\omega}$ in state ω. The probability of state ω is π_ω. If the investor buys θ_i units of each asset, his wealth in state ω will be $w_\omega = w - \sum_i q_i \theta_i + \sum_i d_{i,\omega}\theta_i$. A rational investor will seek to maximize the expected utility $E_\pi[u(\boldsymbol{w})] = \sum_\omega \pi_\omega u(w_\omega)$.

The partial derivatives with respect to the thetas are:

$$\begin{aligned}
\frac{\partial E_\pi[u(\boldsymbol{w})]}{\partial \theta_j} &= \frac{\partial \sum_\omega \pi_\omega u(w_\omega)}{\partial \theta_j} \\
&= \sum_\omega \pi_\omega \frac{\partial u(w_\omega)}{\partial \theta_j} \\
&= \sum_\omega \pi_\omega u'(w_\omega)\frac{\partial w_\omega}{\partial \theta_j} \\
&= \sum_\omega \pi_\omega u'(w_\omega)\frac{\partial\left(w - \sum_i q_i\theta_i + \sum_i d_{i,\omega}\theta_i\right)}{\partial \theta_j} \\
&= \sum_\omega \pi_\omega u'(w_\omega)\left(-q_j + d_{j,\omega}\right)
\end{aligned}$$

At the thetas of a critical point the partial derivatives are zero:

$$\begin{aligned}
0 &= \left.\frac{\partial E_\pi[u(\boldsymbol{w})]}{\partial \theta_j}\right|_{\theta_1^*,\ldots} \\
&= \sum_\omega \pi_\omega u'(w_\omega^*)\left(-q_j + d_{j,\omega}\right) \\
q_j \sum_\omega \pi_\omega u'(w_\omega^*) &= \sum_\omega \pi_\omega u'(w_\omega^*) d_{j,\omega} \\
q_j &= \frac{\sum_\omega \pi_\omega u'(w_\omega^*) d_{j,\omega}}{\sum_\omega \pi_\omega u'(w_\omega^*)} \\
&= \sum_\omega \psi_\omega d_{j,\omega}
\end{aligned}$$

where $w_\omega^* = w - \sum_i q_i\theta_i^* + \sum_i d_{i,\omega}\theta_i^*$, and $\psi_\omega = \dfrac{\pi_\omega u'(w_\omega^*)}{\sum_\omega \pi_\omega u'(w_\omega^*)}$.

It can be shown that if u is a risk-averse utility function (i.e., having positive first derivative and negative second derivative), this critical point is unique and is a maximum. The state prices constitute a probability measure. And ψ is equivalent to π (cf. Section 3), since $\psi_\omega = 0$ if and only if $\pi_\omega = 0$.

Prices are allocated to states according to the product of the real probabilities and the marginal utilities. If u were a linear function of w, u' would be constant and the state prices would be the real probabilities. But a linear utility function indicates risk neutrality. When u is risk-averse, the state prices are the real probabilities if and only if w_ω^* is the same for all ω. So equilibrium considerations show that the real state probabilities are serviceable for pricing only if the investor is risk-neutral or the outcome is determined.

Appendix C: Utility Theory and Distribution Transformation

We have seen that linearity is the key to consistent pricing, i.e., pricing in which there is no allowance for arbitrage. One must model the states ω of the universe and assign state prices ψ_ω to them. Then the value of a deal whose present value in state ω is x_ω must be $\sum_\omega x_\omega \psi_\omega$, or in probabilistic terms, $E_\psi[\boldsymbol{X}]$. The probability measure ψ does not have to be the real probability measure; in fact, rarely is it the real one.

So the risk pure premium for insurance against an immediate loss $\boldsymbol{X}$ whose cumulative density function is F should be not $E_F[\boldsymbol{X}] = \int x dF(x)$, but $E_{F^*}[\boldsymbol{X}] = \int x dF^*(x)$, where F^* is some transformed, or "tilted," distribution. This was recognized by Hans Bühlmann [3], and before him by Gerard Debreu and Karl Borch (references in [18:156]). But Gary Venter [19] popularized the idea, and recently Sean Wang [21] has recommended one transform in particular, the proportional-hazards transform.

On the other hand, utility theory has long played a role in pricing, and has exponents even today (e.g., [12] and [15]). According to utility theory one should enter only into deals that increase one's expected utility. Utility theory treats of expressions like $E_F[u(\boldsymbol{X})]$, where F is the real probability measure. It would appear that utility theory is at odds with distribution transformation. Indeed, if no translation from $E_F[u(\boldsymbol{X})]$ to $E_{F^*}[\boldsymbol{X}]$ is possible, utility theory will have to be abandoned.

Henry Panjer demonstrates that in utility theory "for small risks, the risk premium is proportional to … the variance of the loss distribution." [16:137] And what holds approximately for small risks holds exactly for normally-distributed risks under exponential utility:

$$
\begin{aligned}
u(w) &= E_f\left[u(w - P + \boldsymbol{X})\right] \\
e^{hw} &= E_f\left[e^{h(w-P+\boldsymbol{X})}\right] \\
e^{hP} &= E_f\left[e^{h\boldsymbol{X}}\right] = e^{h\mu + h^2\sigma^2/2} \\
P[\boldsymbol{X}] &= \mu + h\sigma^2/2
\end{aligned}
$$

Furthermore, Hans Gerber has proven that the only functions that possess certain desirable utility properties are exponential functions [10:73]. Venter argues that the variance principle is inconsistent with distribution transformation and allows for arbitrage. If his argument is valid, then utility theory by its association with the variance principle is condemned.

We begin with Venter's example:

> "Consider a retired couple who own two mobile homes in the same trailer park in Oklahoma and who want to purchase homeowners insurance. When the wind comes sweeping down the plain, both homes stand a chance of being damaged. An insurer may thus feel exposed to more than twice the dollar variability in results insuring both than insuring just one, and may thus want more than twice the single home premium for the two. But the market cannot charge a two trailer surcharge, because the couple could just buy separate policies. … [This] illustrates the requirement that market premiums be additive for non-independent risks, and thus rules out the variance principle. Otherwise, de-packaging of exposures could create arbitrage profits." [19:224f]

Putting the argument in symbols, we represent the loss to the first trailer as $\boldsymbol{X}_1$, and that to the second as $\boldsymbol{X}_2$. The exposures of the two trailers to loss need not be identical. According to the variance principle, an insurer selects some number h and determines a risk pure premium as the mean plus *h* times the variance. So if the insurer receives an application for the two trailers, it quotes the risk pure premium:

$$\begin{aligned} P[\boldsymbol{X}_1 + \boldsymbol{X}_2] &= E[\boldsymbol{X}_1 + \boldsymbol{X}_2] + hVar[\boldsymbol{X}_1 + \boldsymbol{X}_2] \\ &= E[\boldsymbol{X}_1] + E[\boldsymbol{X}_2] + h(Var[\boldsymbol{X}_1] + 2Cov[\boldsymbol{X}_1; \boldsymbol{X}_2] + Var[\boldsymbol{X}_2]) \\ &= (E[\boldsymbol{X}_1] + hVar[\boldsymbol{X}_1]) + (E[\boldsymbol{X}_2] + hVar[\boldsymbol{X}_2]) + 2hCov[\boldsymbol{X}_1; \boldsymbol{X}_2] \\ &= P[\boldsymbol{X}_1] + P[\boldsymbol{X}_2] + 2hCov[\boldsymbol{X}_1; \boldsymbol{X}_2] \end{aligned}$$

If the covariance is not zero, *P* is not a linear operator. But we agree with Venter that pricing operators, of which *PV* and *CEV* are instances, must be linear. Indeed, in this example of positive loss covariance between the two trailers, the retired couple ought to make separate applications for insurance.

However, to the insurer faced with covering two trailers in one park it matters not whether there are one or two named insureds. The risk premium must account for the covariance. If the insurer knew that two trailers would be insured in the park, it would charge for the first trailer $P[\boldsymbol{X}_1] = E[\boldsymbol{X}_1] + hVar[\boldsymbol{X}_1] + hCov[\boldsymbol{X}_1; \boldsymbol{X}_2]$ and for the second $P[\boldsymbol{X}_2] = E[\boldsymbol{X}_2] + hVar[\boldsymbol{X}_2] + hCov[\boldsymbol{X}_2; \boldsymbol{X}_1]$. Such a *P* operator would be linear. The example is simplistic to the point of error; it depicts each pricing exercise as happening *in vacuo*. Accustomed to univariate probability, actuaries default to the assumption that the i^{th} deal has its own sample space Ω_i. When they operate on this level, they are left only with marginal distributions, and at some point they will slip into the mistake of equating a joint distribution with the product of its marginals. This mistake is just as likely to happen with distribution transformation as with utility theory.

For example, consider the discrete joint distribution:

$$f_{X_1,X_2}(x_1,x_2)=\begin{cases}1/3 & \text{if } x_1=0 \text{ and } x_2=0\\ 1/3 & \text{if } x_1=1 \text{ and } x_2=0\\ 1/3 & \text{if } x_1=1 \text{ and } x_2=1\end{cases}$$

The marginal distributions are:

$$f_{X_1}(x_1)=\begin{cases}1/3 & \text{if } x_1=0\\ 2/3 & \text{if } x_1=1\end{cases}$$

$$f_{X_2}(x_2)=\begin{cases}2/3 & \text{if } x_2=0\\ 1/3 & \text{if } x_2=1\end{cases}$$

The product of the marginals is:

$$f_{X_1}(x_1)f_{X_2}(x_2)=\begin{cases}2/9 & \text{if } x_1=0 \text{ and } x_2=0\\ 4/9 & \text{if } x_1=1 \text{ and } x_2=0\\ 1/9 & \text{if } x_1=0 \text{ and } x_2=1\\ 2/9 & \text{if } x_1=1 \text{ and } x_2=1\end{cases}$$

The product of the marginals disagrees with the joint distribution over the possibility of the state (0, 1); so they are not equivalent probability measures (cf. Section 3). But risk-adjusted probability measures must be equivalent to the real probability measures, as Venter himself admits: "In order to avoid arbitrage the transformed distribution must give zero probability to the same events as does the original distribution." [20:984] Therefore, the product of transformed marginals will not be equivalent to the transformed joint distribution. And premiums calculated from the marginals will charge something for the state (0, 1), whereas premiums calculated from the joint distribution will charge nothing for it. An arbitrage is created, whether one is doing utility theory or transforming distributions, whenever independence is inappropriately invoked.

Two considerations in Venter's articles themselves show that he affixes the blame wrongly to the variance principle and to utility theory. The first one is the adjusted distribution $f_{\boldsymbol{X}}^{*}(x)=f_{\boldsymbol{X}}(x)(1+h(x-E_{f}[\boldsymbol{X}]))$. As long as h is not so large as to make f^{*} anywhere negative, and as long as f and f^{*} are equivalent, this is a legitimate transform. However, the adjusted mean is:

$$\begin{aligned} E_{f^{*}}[\boldsymbol{X}] &= E_{f}[\boldsymbol{X}(1+h(\boldsymbol{X}-E_{f}[\boldsymbol{X}]))] \\ &= E_{f}[\boldsymbol{X}]+hE_{f}[\boldsymbol{X}(\boldsymbol{X}-E_{f}[\boldsymbol{X}])] \\ &= E_{f}[\boldsymbol{X}]+hE_{f}[(\boldsymbol{X}-E_{f}[\boldsymbol{X}])(\boldsymbol{X}-E_{f}[\boldsymbol{X}])] \\ &= E_{f}[\boldsymbol{X}]+hVar_{f}[\boldsymbol{X}] \end{aligned}$$

Venter was aware of the problem: "Delbaen and Haezendonck … seem to have shown that a variance load is a form of adjusted probability." [20:982] His resolution of the seeming contradiction (viz., 1. Variance principle is bad; 2. Adjusted probability is good; 3. Variance principle is a subset of adjusted probability) is unclear; he claims that the contradiction arises from "inconsistent probability adjustments." Perhaps this touches on the requirement to honor joint distributions. But the fact remains that the variance load is a form of adjusted probability. And why should "inconsistent probability adjustments" plague this form of adjusted probability and not others?

The second consideration is that exponential utility theory is another form of adjusted probability, the adjustment here being the Esscher transform $f_{X}^{*}(x)=e^{hx}f_{X}(x)/E_{f}[e^{hX}]$. The adjusted mean is:

$$\begin{aligned} E_{f^*}[\boldsymbol{X}] &= E_f\left[\boldsymbol{X}e^{hX}/E_f\left[e^{hX}\right]\right] \\ &= \frac{1}{E_f\left[e^{hX}\right]}E_f\left[\boldsymbol{X}e^{hX}\right] \\ &= \frac{1}{E_f\left[e^{hX}\right]}E_f\left[\frac{\partial e^{hX}}{\partial h}\right] \\ &= \frac{1}{E_f\left[e^{hX}\right]}\frac{\partial E_f\left[e^{hX}\right]}{\partial h} \\ &= \frac{\partial \ln E_f\left[e^{hX}\right]}{\partial h} \end{aligned}$$

If $\boldsymbol{X}$ is normal with parameters μ and σ^2, $E_f\left[e^{hX}\right] = e^{h\mu + h^2\sigma^2/2}$, and $E_{f^*}[\boldsymbol{X}] = \mu + h\sigma^2$. The variance principle crops up again, although this time Venter does not deal with the seeming contradiction [20:983].

The solution is multivariate. Suppose that the true state of the universe is *n*-dimensional; each state corresponds to an $(n\times 1)$ random vector **x** whose $(n\times 1)$ mean is μ and $(n\times n)$ variance is Σ. And let h be an $(n\times 1)$ vector. The adjusted distribution of the first consideration can be made multivariate:

$$\begin{aligned} f_{\mathbf{x}}^*(\mathrm{x}) &= f_{\mathbf{x}}(\mathrm{x})\left(1 + \left(\mathrm{x} - E_f[\mathbf{x}]\right)'\mathrm{h}\right) \\ &= f_{\mathbf{x}}(\mathrm{x})\left(1 + (\mathrm{x} - \mu)'\mathrm{h}\right) \end{aligned}$$

The adjusted mean is:

$$
\begin{aligned}
E_{f^*}[\mathbf{x}] &= E_f\left[\mathbf{x}\left(1+(\mathbf{x}-\mu)'\mathrm{h}\right)\right] \\
&= E_f[\mathbf{x}] + E_f\left[\mathbf{x}(\mathbf{x}-\mu)'\right]\mathrm{h} \\
&= E_f[\mathbf{x}] + E_f\left[(\mathbf{x}-\mu)(\mathbf{x}-\mu)'\right]\mathrm{h} \\
&= E_f[\mathbf{x}] + Var_f[\mathbf{x}]\mathrm{h} \\
&= \mu + \Sigma\mathrm{h}
\end{aligned}
$$

This form assures the linearity of prices, since $E_{f^*}[\mathrm{A}\mathbf{x}] = \mathrm{A}E_{f^*}[\mathbf{x}]$. And significantly, it makes use of the covariance elements of Σ.

The Esscher transform of the second consideration also can be made multivariate. If **x** is multivariate normal with parameters μ and Σ, and h is an (n×1) vector, then $E_f\left[e^{\mathrm{h}'\mathbf{x}}\right] = e^{\mathrm{h}'\mu+\mathrm{h}'\Sigma\mathrm{h}/2}$ (cf. [12:122]). And the adjusted mean is:

$$
\begin{aligned}
E_{f^*}[\mathbf{x}] &= \frac{\partial \ln E_f\left[e^{\mathrm{h}'\mathbf{x}}\right]}{\partial \mathrm{h}} \\
&= \frac{\partial\left(\mathrm{h}'\mu + \mathrm{h}'\Sigma\mathrm{h}/2\right)}{\partial \mathrm{h}} \\
&= \mu + \Sigma\mathrm{h}
\end{aligned}
$$

The variance principle is vindicated by the appeal to true sample spaces and joint densities. Practitioners may despair of having to know everything in order to say anything; doubtless, shortcuts and approximations must be made. Practice can be, indeed, must be, excused for them. But in a theory they are inexcusable.

So distribution transformation does not contradict utility theory. Most likely, they are allies in the search for a true and clear understanding of economics, finance, and actuarial science.

Appendix D: Utility Theory and Price Appreciation

In Section 5 we saw that the price of an asset whose uncertainty is not changing remains proportional to the price of an asset whose future payment is certain. Here we will see whether exponential utility theory accords with this.

Let $v(t)$ be the discount function, and let $\boldsymbol{X}$ represent a random payment at time T with mean μ and variance σ^2. At t units of time before T the present value of $\boldsymbol{X}$ is:

$$\begin{aligned} PV[\boldsymbol{X};t] &= v(t)\boldsymbol{X} \\ E[\boldsymbol{X};t] &= v(t)\mu \\ Var[\boldsymbol{X};t] &= v(t)^2\sigma^2 \end{aligned}$$

An exponential utility is a linear transform of the function e^{hx}, where h is one's risk tolerance. In this form, a positive h indicates risk inclination and a negative h indicates risk aversion. Risk neutrality is the limiting case as h approaches zero. In Appendix C we noted that for one whose utility is exponential, the certainty equivalent value is equal to the mean plus h times the variance. Therefore, at t units of time before T, $P[\boldsymbol{X};t] = v(t)\mu + hv(t)^2\sigma^2/2$, a formula that reduces to $\mu + h\sigma^2/2$ at zero units of time before T, since $v(0) = 1$.

We saw that in this situation, a situation in which the uncertainty of $\boldsymbol{X}$ does not change until time T, at which time it is completely resolved, $P[\boldsymbol{X};t]$ should equal

$v(t)P[\boldsymbol{X};0] = v(t)\mu + hv(t)\sigma^2/2$. But the utility-theoretic formula for $P[\boldsymbol{X};t]$ has a quadratic factor of *v*(*t*), thus seeming not to be in accord with the stronger theory.

The explanation concerns the risk tolerance parameter *h*. One's risk tolerance is not static; rather it varies in proportion to one's wealth. If one is twice as wealthy now than before, most likely, one is prepared to accept risks twice as great. This translates into having an *h* half as large as before. Or, suppose that one lived in a country with inflation and a weak currency, the "Quant," and that over time the prices of even small purchases, such as loaves of bread, were in thousands of Quants. The people tire of these high amounts, and the government decrees that tomorrow one Quant will buy what one thousand Quants now buy. So tomorrow, everyone will drop three zeroes from prices and currency. To add three zeroes to one's *h* will allow one's utility-theoretic decisions tomorrow and thenceforth to continue undisturbed (example taken from [12:92]).

Therefore, *h* is a function of wealth. And in this example, one's wealth *t* units of time before *T* will be $w(t) = v(t)w(0)$. And since risk tolerance *h* should vary inversely with wealth *w*, i.e., $h(t)w(t) = h(0)w(0)$:

$$\begin{aligned}
P[\boldsymbol{X};t] &= v(t)\mu + h(t)v(t)^2\sigma^2/2 \\
&= v(t)\left(\mu + h(t)v(t)\sigma^2/2\right) \\
&= v(t)\left(\mu + h(t)w(t)\frac{v(t)}{w(t)}\sigma^2/2\right) \\
&= v(t)\left(\mu + h(0)w(0)\frac{v(t)}{v(t)w(0)}\sigma^2/2\right) \\
&= v(t)\left(\mu + h(0)\sigma^2/2\right) \\
&= v(t)P[\boldsymbol{X};0]
\end{aligned}$$

If we provide for the changing of risk tolerance, utility theory will accord with Section 5.

Appendix E: Interest-Rate Risk

If there is a "market" discount function, it is the discount function of the "STRIP" (Separate Trading of Registered Interest and Principal [16:17]) market. Interest-rate risk is the risk that the discount function will change to one's detriment. The discount function changes every moment, sometimes changing significantly within periods of just a few days. Changes may be beneficial, as well as detrimental. But many feel that the risk of detrimental changes is greater than the reward of beneficial ones, and desire for the price of a risky deal to include a charge for that risk. This appendix will argue that to load for interest-rate risk is illegitimate.

Let $v(t)$ be the discount function at the present time. The price of a treasury STRIP that will pay k units of money at T units of time in the future is $p(T) = v(T)k$. A moment later, when the payment is a $T - \Delta T$ units of time in the future, the price is $p(T - \Delta t) = v(T - \Delta t)k$. Therefore, the instantaneous appreciation of the STRIP is:

$$\begin{aligned}
\lim_{\Delta T \to 0} \frac{p(T-\Delta T) - p(T)}{p(T)} \cdot \frac{1}{\Delta T} &= \lim_{\Delta T \to 0} \frac{v(T-\Delta T) - v(T)}{v(T)} \cdot \frac{1}{\Delta T} \\
&= -\frac{1}{v(T)} \lim_{\Delta T \to 0} \frac{v(T-\Delta T) - v(T)}{(T-\Delta T) - (T)} \\
&= -\frac{v'(T)}{v(T)} \\
&= \phi(T)
\end{aligned}$$

The function $\phi(t) = -\frac{v'(t)}{v(t)}$, mentioned in Section 2, is the called the forward rate. At every moment the price of the STRIP is growing at the forward rate of its maturity. Though the STRIP matures in ascending order, the forward rates affect its price in descending order. Paradoxically, the forward rate is in some sense a "backward" rate.

This backwardness, or "LIFO-like" behavior, of the forward rate is the reason why one cannot pay k units of money now for the reception of $k/v(t)$ units of money whenever one should wish to receive it. The author is not aware of a risk-free asset that will pay its owner a defined $s(t)$ units of money and leaves t to the choice of the owner. But such an asset would be very useful; perhaps one can synthesize it from bonds and futures.

How would investors act if they were certain that $v(t)$ would not change? They would buy STRIPs whose maturity is at such times T as $\phi(T) = \sup\{\phi(t) : t \in \Re\}$. In other words, they would exploit the backwardness of the forward rate by buying into the highest rates. They would hold the STRIPs until their forward rates dropped, and then exchange them for STRIPs of the original maturities.

If investors were certain that $v(t)$ would not change, $\phi(t)$ would be constant; otherwise no one would buy into the lower rates. The fact that $\phi(t)$ is not constant implies that investors believe that $v(t)$ *could* change. Therefore, the current state of the $v(t)$ function impounds investors' beliefs about the future states of the $v(t)$ function. And because investors are risk-averse, they set $v(t)$ so that the reward of possible beneficial changes is sufficiently greater than the risk of possible detrimental changes. Therefore, loading for interest-rate risk is double-counting.

It is from the lack of a theory that decision-makers fabricate risks for which they can levy charges. That they sacrifice these charges when they stand in the way of a deal shows that they doubt their legitimacy. Far from being ludicrous, it is not much of a stretch to argue that one should charge for “regret risk,” the risk that one might regret having done the deal. A sound theory is the only cure for these ills.