Abstract

This paper takes up the challenge by Robert S. Clarkson (AFIR 1994) in suggesting the direction for a coming revolution in finance.

The work of Borch, Gerber, Freifelder, Taylor and others is consistent with the general approach to solving problems in finance called efficient market theory. This paper applies efficient market theory to the question of the time value of money. With some rearranging, the eight axioms of Gerber (1979) become seven.

With these seven axioms, a decision rule emerges that has the following decision rules as special cases: that a firm considering among investments in securities for fixed terms should choose the investment with the highest return on investment; that a firm considering among investments in plant or equipment with limited but uncertain lifetimes should choose those which pay back the investment within the time the equipment is expected to be useful; that a firm considering among alternatives that all threaten the firm in a major way should choose the alternative that gives the maximum value for the firm under the worst possible scenario; that a firm considering among underwriting opportunities should choose those opportunities that offer the greatest rewards in excess of all costs, where all costs include an explicit estimate of the cost of capital. It is therefore a general theory of finance.

The general theory explains the failure of the central prediction of the Capital Asset Pricing Model, which is that risk can be measured by the covariance of outcomes with broad market averages, and suggests an alternative linear relationship between risk and return.

The results suggest that insurance companies that price their products using Return on Equity guidelines are producing results that are less than optimal increases in the value of the firm.
I. Introduction

Robert S. Clarkson has suggested that a revolution in the theory of finance is underway (Clarkson (1994)). He suggests that the revolution will require new foundations to replace the expected utility principle, the assumption that prices of investments reflect the variance of returns, and the assumption that prices of investments reflect the covariance between individual investment returns and broad market averages of returns. Clarkson uses both hypothetical situations and experimental evidence to criticize these pillars of financial theory.

The theory of finance, both as it is and as it will be after the coming revolution, is by its nature prescriptive. The theory of finance suggests how a decision-maker should make decisions among alternative courses of action. The theory of finance is not descriptive. It does not show why people or institutions behave as they actually do. In the long run, individuals and institutions that behave wisely will generally grow in strength and influence, although there will always be exceptions. The theory of finance will describe actual behavior to the extent that the facts being studied arise from a process that is characterized by participants who do behave in their best interest.

Fortunately, there are many practical situations that yield empirical data. We can look to the histories of specific markets around the world for evidence about whether those who follow the prescriptions of any theory of finance actually do succeed. Indeed, one of the most salient features of the revolution in the theory of finance is that reasonable armchair speculation will be increasingly subjected to empirical verification. Observations will become as important as theory. Only when this takes place will finance take its place as a historical science beside well established historical sciences such as geology and paleontology.

One of the most successful insights in the history of the theory of finance is efficient market theory. Here is Stephen A. Ross addressing the Society of Actuaries in April, 1994:

I actually trace the roots of the modern subject [of finance] back a bit further. I traced it to a wonderful, somewhat neglected article in 1937 by Cowles, who
examined what we know call the efficiency of markets....Efficient market theory lay dormant after Cowles until around the 1950s, and then it picked up steam in the 1960s and 1970s. It is the empirical basis for what we think of as modern finance. If you look closely, lurking in the background of option-pricing theory, asset-pricing models, and all of the paraphernalia of modern finance, are the fundamental intuitions of efficient market theory....[T]he thought was that the current price was really some sense of the reflection of the consensus of all of the participants in the market. As such, it incorporates all of the information that people have. (Ross (1994))

Efficient market theory suggests that Clarkson is generally correct. Because investors have a great deal of information about the entire range of possible returns on their investments, efficient market theory suggests that rewards for investors will not always vary simply in proportion to the variance of returns, and therefore that prices of investments will not generally reflect the covariance between individual investment returns and broad market averages.

In addition, the decision-maker should always choose to maximize the economic value of the firm. Whenever faced with two alternatives, one offering the highest return on equity and another offering the highest economic value of the firm, the prudent investor will choose the alternative that maximizes the value of the firm. That is, the prudent investor will not always choose the alternative that offers the higher return on equity. After a series of such decisions of any length, the strategy to maximize the value of the firm will lead to the greatest possible economic value of the enterprise, while the return-on-equity strategy will result in a lower economic value for the firm.

One implication is that insurance investors and regulators who rely on Return on Equity guidelines as they monitor the results of insurance companies will not produce regulations or investments as productive as in a free and informed market. Specifically, investments in insurance companies will perform less well than they might, and regulations will reduce the availability of insurance and increase the inefficiencies of the
market. Investments and regulations based on maximizing the economic value of the firm will maximize the efficiency of the marketplace.

2. The Coming Revolution

_Many of the decision rules in the current theory of finance contradict everyday experience. It is as if the theorists were saying, "I know that people and businesses don't make decisions this way, but they would if they were wiser." While every theory attempts to inform the reader, good theories are based on observations about the way the world works. Good theories and good data may suggest ways to improve decisions, but they can reasonably be expected to promote a greater degree of consistency among the decisions one makes in current practice, not to suggest that current practice generally reflects an ignorance of how to behave wisely._

_Two case examples illustrate the point that today's theory of finance does not rest on firm empirical footing. First, everyone who has recently worked in a manufacturing plant knows that decisions to invest in improvements in equipment are often justified in terms of a two-year or three-year payback rule: Make the investment if it will pay for itself in (say) three years, and don't make the investment if it will not. Yet in 1991, a leading textbook, Principles of Corporate Finance, by Richard A. Brealey and Stewart C. Myers, discusses payback rules for only three paragraphs, concluding, "So it is still better to use the net present value rule." (p. 77)_

_Second, many businesses survived the junk bond turmoil of the 1980's by rejecting the advice of investment bankers that they borrow at high interest rates. The investment bankers typically justified their advice by pointing out that the company had an expected rate of return on equity of, say, 18%, and arguing that the debt repayment should be discounted at that annual rate. A number of companies that borrowed at high interest rates later had trouble repaying their debts. In retrospect, a number of critics suggested that the investment bankers were irresponsible, or even criminal, for pushing their clients to do just what the theory of finance said should be done. The lessons of the junk bond_
years, as recorded in hard data about corporate successes and failures, does not support the wisdom of discounting long-term debt obligations at high discount rates.

Third, "profit" is no longer a useful measure of the return on investment. Consider, for example, McCaw Cellular, which was founded in 1982 with a capital of US$40 million and sold to AT&T in 1994 for US$16,000 million of assets and debt assumption. The investors in McCaw Cellular enjoyed a spectacular return on capital. In its twelve years, however, McCaw Cellular never made a book profit (until the books were set up for the last quarter before the sale). It just grew in market share and in its ability to throw off cash. (McCaw (1994)).

We will explore more examples later in this paper, primarily from life and property-casualty insurance companies in the United States. Regardless of the persuasiveness of these examples, the fundamental point remains true, for it is part of a worldwide shift toward information management. As the historians of science point out, science does not rest on experimentation alone. Observations of all kinds are the benchmark of science, and the most important observations are those that are surprising. As Kitcher (1982) wrote, "Theories win support by producing claims about what can be observed, claims that would not have seemed plausible prior to the advancement of the theory, but that are in fact found to be true when we make the appropriate observations." (p. 35)

Management Should Seek to Maximize the Value of the Firm

The example of McCaw Cellular illustrates the importance of maximizing the value of the firm. The cost of risk will increasingly be explicitly recognized as simply a cost, and the objective will be to maximize the excess of income over all costs. (Drucker (1993), p. 81)

The value of an investment in a firm is a function of the level of expected dividends and other cash benefits and the level of expected increase in the market value of the firm. In addition, the value of an investment in a firm is a function of the certainty with which the expected gains can be estimated. If two firms have the same average expected dividend stream and share value, but one is secure and the other is risky, the firm offering the higher degree of certainty is the more valuable.
This being the case, the management of the firm should seek to maximize the value of future dividends and increases in stock price while maintaining an appropriate aversion to risk. The economic value of the firm is simply the value of future dividends and increases in stock price in the context of an appropriate aversion to risk.

Much of the theoretical structure of finance today is based on the principle that the management should seek to maximize the return on equity of the firm, where equity is reflected in accounting conventions or in the market value of the firm's stock. Maximizing return on equity is not, however, the same as maximizing the value of the firm (except in trivial cases). If two firms begin with exactly the same situation, and each is faced with the same alternatives over time, the firm that chooses the alternatives that maximize the value of the firm will always be at its maximum value, while the firm that seeks to maximize return will always be at a lesser value.

The management of the more valuable firm will be generally regarded as more successful than the management of the less valuable firm. Certainly the investors who backed McCaw Cellular from 1982 to 1994 consider its Chairman, Craig McCaw, highly successful. When we wish to develop prescriptive decision rules -- rules to follow to be successful -- it is best to base rules on the standard that one should choose alternatives that maximize the value of the firm.

In the next section we develop a general theory of finance from a set of axioms. In Section 4 we will explore several of the implications of this general theory, and cast those implications in terms of "claims about what can be observed."

3. A General Theory of Finance

The major breakthroughs in finance in the past fifty years have come through what Ross and others call efficient market theories (Ross (1994)). Several writers have developed axiomatic systems for decision-making under uncertainty. The broad field of game theory is an example of work along these lines. Borch (1962), Gerber (1979), Buhlmann and
Jewell (1979), and Taylor (1989) (and many others) have applied axiomatic systems to insurance problems. The axioms are consistent with efficient market theory. Here we extend the work of Borch, Gerber, Buhlmann, Jewell, Taylor and others, by applying efficient market theory to the problem of estimating the time value of money.

The purpose of this section is to find an optimal method to make decisions in order to maximize the value of the firm.

The axioms used to build this system of decision-making rules are:

1. There is a frontier of opportunities that are optimal for the firm, and this frontier can be expressed in terms of one or more dimensions, including perhaps the expected value and other values, including risk. The purpose of the decision-making process is to identify the opportunity frontier and to reduce the number of dimensions in which the frontier is expressed. For example, the traditional Internal Rate of Return (IRR) model for evaluating alternative investment strategies expresses each alternative as a single value, the present value at the selected discount rate.

2. The decision-maker should be averse to risk.

3. The decision-maker should never be willing to pay more to avoid a risk of loss than the amount of the loss. In practice, an insurance executive might recommend the purchase of a reinsurance contract with a premium greater than the maximum possible recovery, but this would be appropriate only if there were other considerations, such as taxes or the effects of rating agencies, that have not been considered in the calculation of premium and maximum loss.

4. The decision should be a function of the analytical process, including the data employed in reaching the decision, but small changes in the data should not lead to extraordinary changes in the indications of the appropriateness of particular decisions. For example, a small change in interest rates should not lead to a large change in the value of any alternative. A small change in
the data may, of course, imply a different ranking of the alternatives, and the alternatives may involve widely differing actions.

Put another way, the firm may behave consistently with catastrophe theory, with dramatic changes in strategy as a result of small changes in conditions, but it should not behave chaotically, with unpredictable costs and benefits arising from small changes in data or assumptions. Note that this is a more restrictive axiom than one that merely requires that the implications of changes in conditions be continuous functions of the data and the analytical process that is used.

5. Small changes in the analytical process should not lead to extraordinary changes in the apparent appropriateness of particular decisions. In particular, the results of the decision-making analysis should not be significantly affected by a small refinement in the description of any strategy, or by a small refinement in the description of the results of any strategy under a particular set of assumptions. For example, the results of an analysis of whether to expand sales in one market should not be significantly affected by refining the description of one scenario to create two similar but slightly different scenarios.

6. Small changes in the level of detail of the time scale of the analytical process should not lead to extraordinary changes in the apparent appropriateness of particular decisions. For example, the results of an analysis of whether to expand sales in one market should not be significantly affected by refining the choice of time intervals from annually to semiannually.

7. If there is no risk, but the outcomes result in flows of currency at future times, the time value of money can be determined from the current yield curve for the currency. This follows from efficient market theory, because the market for future certain cash flows exists and a decision-maker could improve the value of the firm under any alternative rule by buying and selling in the market for future cash flows.
With these axioms, we can deduce a single system of decision-making and many of the implications of this system.

As a check on the reasonableness of the results, we should expect the principle decision-making systems in existence to appear as special cases of the general system. As it happens, many popular decision-making rules appear as special cases, including the IRR approach to evaluating investment decisions and the basics of von Neumann-Morgenstern game theory.

**Axiom 1. The Opportunity Frontier**

This axiom states that every alternative being evaluated can be expressed in terms of one or more numbers that summarize the impact of the alternative on the firm. Here we use the term *firm* to denote the bundle of opportunities and commitments that comprise the entity for whom the alternatives exist. A firm might be an insurance company, a business, a non-profit organization, a public entity, or an individual. While this discussion is in terms of insurers, the results are quite general.

(I understand the economist Pareto developed the concept of the opportunity frontier, that is, that some alternatives can be excluded from consideration because they are less than optimal regardless of the decision rule used to convert two or more dimensions to a single dimension. I have not found a source.)

For example, using an IRR model, every alternative is expressed as a single number, the present value at the risk-adjusted rate of return. Classic Markowitz mean-variance analysis expresses every alternative in terms of two numbers, the mean and variance expected returns. The Capital Asset Pricing Model expresses each alternative in terms of two numbers, the expected return and the covariance with market returns. The axiom includes the idea that all of the alternatives can be ranked according to these numbers, and that one or more alternatives will have the highest ranking. (See Fig. 1.)
The axiom that there is a frontier of opportunities does not imply that the performance measures of every set of corporations follows a curve like that shown in Figure 1. Lee Barnes of Davis Bank has pointed out that the actual mean and variance performance measures of property-casualty insurance companies in the U. S. fall into two clusters, one with high return and low variance, and one with low return and high variance. (Personal communication, May, 1994.)

Leaving aside for the moment the measures to be ascribed to risk and return, attempting to prescribe behavior using the concept of the opportunity frontier begs the question of how one should make the trade-off between risk and reward. The existence of the opportunity frontier does not by itself indicate that one alternative is preferable over all of the others unless every alternative can be expressed in terms of a single number. The opportunity frontier only allows one to decide that some alternatives are not appropriate without expressing all of the alternatives as single numbers.

In combination with a rule to express every alternative as a single number, the opportunity frontier says that there exists an alternative or a set of alternatives that has a higher value than all other alternatives. A prescriptive system of decision-making -- that is, one which informs the user of which alternative to choose -- must therefore include such a rule. All prescriptive systems of decision-making employ such a rule. For example, classic Markowitz mean-variance analysis is always
supplemented in practice with a rule that relates the variance measure to the mean.

Whether or not one wishes to employ a utility function, in order to have a prescriptive system for decision-making one must adopt a set of rules for mapping the numbers that are used to measure an alternative into a single number. The balance of this paper is devoted to identifying the list of numbers that should be used to measure an alternative, and the rules that should be used to express those numbers in terms of a single number for the purpose of comparing alternatives.

**Axiom 2. Risk Aversion**

This axiom states that the decision-maker should be averse to risk. Specifically, if two alternatives are identical in the expected values of their returns at every point in time, but one of the two alternatives involves less uncertainty among the returns, that one is to be preferred.

The risk inherent in a decision is greater than the sum of the risks arising from outside the firm—the risks of fires, floods, price wars, and so on. It includes all the risks arising out of the decision-maker's ignorance. As Peter F. Drucker wrote,

> The ability to connect [to the end result] and thus to raise the yield of existing knowledge (whether for an individual, for a team, or for the entire organization) is learnable. Eventually, it should become teachable. It requires a methodology for problem definition—-even more urgently perhaps than it requires the currently fashionable methodology for "problem solving." It requires systematic analysis of the kind of knowledge and information a given problem requires, and a methodology for organizing the stages in which a given problem can be tackled—the methodology which underlies what we now call "systems research." It requires what might be called "Organizing Ignorance"—and there is always so much more ignorance around than there is knowledge. (Drucker (1993)) [italics in original]
It has long been known, and often demonstrated in readings of the early examinations of actuarial societies, that an insurer that priced its products and services in accordance with the expected value principle is mathematically certain to bankrupt itself in the long run. Any prescriptive system for decision-making should reflect an aversion to risk, or it is a prescription for disaster.

Axiom 3. No Panic

This axiom states that the decision rule should not lead to an alternative that reduces the chance of a loss but at a cost greater than simply losing the full amount of the loss. As Gerber (1979), writing about the premiums to be paid to reduce risk, puts it, "The premium should not exceed the maximal possible benefit."

One immediate casualty of axiom 3 is the variance loading principle. While a correct decision rule might lead to a cost of risk that is approximately proportional to the variance in some particular factual situation, a decision rule that is simply a variance load can lead to a cost of risk that is greater than the amount at risk, as Gerber (1979) and Van Slyke (1985) have pointed out.

We shall return to the implications of this axiom on the importance of the Capital Asset Pricing Model, which presumes a variance load, in Section 4.

Axiom 4. Small Changes in the Data

The fourth axiom is that the results of the decision analysis should not be affected by small changes in the data used in the analysis supporting the decision. The implications of this axiom can be easily seen by considering a simple example.

Consider the decision between these alternatives:

1. The status quo.
2. A payment of $49 for a lottery ticket with a 50% chance of paying $100 and a 50% chance of paying nothing.
As presented, each alternative can be expressed in terms of three numbers: the chance of losing $49, the chance of losing nothing, and the chance of gaining $51. The axiom states that the single number used to express each alternative should not be unduly affected by small changes in the data.

Here are the results of applying two different rules to express each alternative as a single number:

**Rule A.** Score a 0 if the chance of losing more than $49 is more than 50%; score a 1 otherwise.

- Alternative 1: Score is 1.
- Alternative 2: Score is 1, but the slightest increase in the cost of the lottery or the probability of losing the lottery will change the score to 0.

Comparison: Alternatives are equivalent as the data is stated, but the slightest increase in the cost of the lottery or the probability of losing will cause the alternatives to be valued quite differently.

**Rule B.** Express each alternative as the expected value, less the variance of possible results divided by $2,500.

- Alternative 1: Score is 0.
- Alternative 2: Score is:

\[-49 + 0.5 \times (0 + 100) - \frac{(0.5 \times 49^2 + 0.5 \times 51^2)}{2,500} = 0.0004\]

Comparison: Alternatives are approximately equivalent as the data is stated, and a slight increase or decrease in the cost of the lottery or the probability of losing will not
affect the conclusion that the alternatives are approximately equivalent.

Note that the axiom does not say that the ranking of alternatives is not dramatically affected by the scores that the alternatives receive. Such an axiom would be useless in practice. The axiom states that the single number used to express a set of numbers that describes an alternatives is not unduly sensitive to a small change in the data.

The first scoring system is inconsistent with the axiom. The second scoring system is consistent with the axiom.

One immediate casualty of requiring the decision process to be based on axioms 1 and 4 is the decision rule that sets a threshold probability of insolvency and scores a zero or one depending whether a firm has a probability of insolvency more or less than the threshold.

Axiom 5. Small Changes in the Analytical Process

This axiom states that the numbers associated with alternatives should not be unduly affected by a small change in the way the problem is viewed. For example, consider the following two views of the two alternatives discussed in the previous section:

Alternative 1: Status quo.

Alternative 2a: Pay nothing for the lottery ticket, but encounter a 50% chance of losing $49 and a 50% chance of winning $51.

Here are the results of applying two rules to the problem of expressing the results of the alternatives as a single number:

Rule A'. Express each alternative as the expected value, less the variance of possible results divided by 100 x ($50 - cost of lottery ticket).

Alternative 1: Score is 0.
Alternative 2 : Score is :

\[-49 + (0.5 \times 100) - \frac{(0.5 \times 49^2 + 0.5 \times 51^2)}{(100 \times (50 - 49))} = \$24.01\]

Alternative 2a : Score is :

\[(0.5 \times (-49) + 0.5 \times 51) - \frac{(0.5 \times 49^2 + 0.5 \times 51^2)}{(100 \times (50 - 0))} = 0.4998\]

Rule B.  Express each alternative as the expected value, less the variance of possible results divided by $2,500.

Alternative 1 : Score is 0.

Alternative 2 : Score is 0.0004.

Alternative 2a : Score is 0.0004.

We see that Rule A' is inconsistent with the axiom, as the score is dramatically affected by merely restating the alternative. Rule B is consistent with this axiom.

One casualty of this axiom is the school of decision-making theory that prescribes decisions based on a utility function tied to the decision-maker's wealth, while tying the measure of wealth to the outcomes of the alternatives.

Other writers have criticized wealth-based decision rules. It is important not to overstate the importance of the conclusion reached here. Decision rules based on exponential utility calculations are not excluded unless they incorporate a framing of the results in terms of the decision-
maker's wealth and make wealth a function of the alternatives that are being evaluated.

Another casualty of this axiom is a set of risk loadings that is proportional to the standard deviations of the results. Changing the description of a risk to treat it as two or more risks would change the risk loading based on standard deviations, as Gerber (1979) and others have shown. (Gerber describes two principles of the calculation of risk premiums, which he calls additivity and iterativity. Each is a mathematical statement of one aspect of axiom 5. Additivity is the principle that the risk premium should be affected by two independent contingencies as the sum of the risk premiums for the two contingencies. Iterativity is the principle that the risk premium for a set of contingencies can be computed by replacing one or more contingencies with its net value (expected value less risk premium) and then evaluating the alternative as restated. These principles must hold if the number that is the result of the decision analysis is to behave reasonably in light of small changes in the analytical process.)

It has long been a premise of analytical economics that prices shall reflect the marginal cost or benefit of the good or service. For example, Willet (1901) wrote, "The most productive apportionment of capital would evidently be the one in which the marginal productivity was the same in all industries." The principle of equal marginal utility suggests that the total cost of risk for a group of independent contracts is equal to the sum of the costs of risk for each of the contracts individually. If this were not true, a profit could be made simply by buying and selling bundles of offsetting risks. (The capital asset pricing model predicts that the cost of capital for a group of independent contracts is zero. This is a reflection of that model's assumption that the size of the firm or portfolio is unchanged by the addition of a risk, and that the additional risk is accommodated by reducing the firm's share of every other risk it participates in. This conclusion of the capital asset pricing model is often misinterpreted to imply that if one increases the risk in a portfolio by undertaking an additional risk with outcomes uncorrelated with the firm's other outcomes, the cost of risk will be zero. That this is a misinterpretation can easily be seen by noting that the model for the cost of risk for the firm is that the cost of risk is proportional to the variance of the firm's results.)
Karl Borch demonstrated more than thirty years ago that under axioms 1 to 5, if and only if the cost of risk is computed using an exponential utility function, the total cost of risk for a group of independent contracts is equal to the sum of the costs of risk for each of the contracts individually. This was consistent with earlier results in the field of decision theory that showed that the only decision rule that allowed one to be willing to agree to two identical contracts in succession (for all pairs of contracts) was a rule consistent with exponential utility theory.

It is less obvious that this axiom precludes all expected utility rules that are risk averse except the exponential principle.

**Axiom 6. Small Changes in the Time Scale**

This axiom states that the number expressing each alternative should not be unduly affected by the fineness of the time scale in which the alternatives are described. An outcome that is expected to materialize at a particular date will be valued only marginally different from an outcome that is expected to materialize shortly before or after that date.

**Axiom 7. Risk-Free Returns**

This axiom states that the decision rule should be consistent with the opportunity costs for buying and selling options on amounts certain but deferred in time. For example, if an alternative involves a single payment of $100 at the end of one year, and that alternative has essentially no uncertainty about the amount that will be paid or about when it will be paid, then the alternative should be valued consistently with the value of a risk-free investment maturing for $100.

Taken together with axioms 4, 5, and 6, this implies that as uncertainty regarding the amounts to be received and regarding the timing with which the amounts are to be received approaches zero, the number generated by the decision rule should approach the sum of the set of prices for the purchase of risk-free maturities at the given times. This constrains the decision rules considerably, as they must contain references to these prices.
The mathematics of computing a value from a set of outcomes and their associated probabilities and dates of maturity constrains the decision rule still further. Specifically, when one writes that an outcome that might occur on December 31 has a value of $100 and a probability of 10%, one means that given that the eventuality has arisen, an amount of $100 will be received. The amount that will be received is certain to be received if the eventuality arises whose probability is 10%. If an amount of $100 will be received on December 31, its value today is, say, $95. (The value is indicated by the money markets.) Therefore, an outcome that might occur on December 31 for $100 with a probability of 10% is equivalent to an outcome that might occur today for (say) $95 with a probability of 10%.

This axiom has the effect of decomposing the values associated with every alternative into two effects, the time value of money and the cost of risk. The time value of money in the absence of risk is determined from the available returns on investments in risk-free securities denominated in the currency (the money) in which the returns will be paid. It is therefore a function of the currency in which the returns will be paid.

There is a body of literature about investing in currencies. This literature provides important insight into the time value of money as one component of the value associated with any decision alternative.

Actuaries and investors have often used rates of return compounded over time to reflect the time value of money. This axiom implies that the time value of money should instead be read from the yield curve of the currency in the world's money markets at the time of the decision.

A General Theory Consistent with the Axioms

As Gerber shows, only one way of evaluating a set of possible outcomes is consistent with axioms 1 to 5. That way is to summarize the value of a set of possible outcomes at a given time with the following expression, which I have called the Risk-Adjusted Value of the outcomes (Van Slyke, 1983) (although I believe John Cozzolino, now of Pace University, coined this phrase):
The Risk-Adjusted Value of the Firm

\[ RAV(c) = -c \ln E \left[ e^{\frac{x}{c}} \right] \]  

That is to say, after each outcome is scaled to reflect the time value of money, the results are averaged using weights derived by exponentiation, and the average is put back into the original units of measurement. In this expression, items of income are positive values of \( x \), and increase the Risk-Adjusted Value, and items of outgo are negative values, and decrease the Risk-Adjusted Value. Adding a 10% chance of receiving $100 to the set of possibilities increases RAV by less than $10; adding a 10% chance of paying $100 to the set of possibilities decreases RAV by more than $10.

This is a strong result. It severely limits the range of possible decision rules that can be consistent with the axioms. Yet it has been widely criticized. Perhaps the most enduring criticism is the paradox of Allais, which is cited by Clarkson. It is important, therefore, to see if the paradox of Allais means that axioms 1 to 5 are inappropriate.

*The Paradox of Allais.*

The paradox of Allais is as follows (as described by Clarkson (1994)):

Allais asked subjects to consider the following two alternatives:

Alternative A: receive £1,000,000 with certainty
Alternative B: receive £5,000,000 with probability 0.10
receive £1,000,000 with probability 0.89
receive nothing with probability 0.01

He has found that his subjects preferred alternative A to alternative B. He then asked them to consider the following:

Alternative C: receive £1,000,000 with probability 0.11
A GENERAL THEORY OF FINANCE

receive nothing with probability 0.89
Alternative D: receive £5,000,000 with probability 0.10
receive nothing with probability 0.90

He found that his subjects preferred D to C.

Clarkson suggests that the subjects are acting rationally, but that the subjects frame the alternatives differently depending on their perceptions of wealth. The subject evaluating alternatives A and B will consider his wealth to be increased by the £1,000,000 of Alternative A, and will regard Alternative B as a 1% chance of losing that wealth. The subject evaluating situations C and D will perceive no risk of loss, and nothing to avoid, in either case, and will prefer an expectation of £500,000 to an expectation of £110,000.

I suggest that the paradox does not arise from a shift in one's frame of reference. Indeed, the individual faced with a choice between A or B might be more willing to bear risk than the person faced with a choice between C or D, simply because the first decision-maker does indeed have the opportunity to have £1 million for certain. The paradox arises from the radically different values the subjects associate with the lottery between A and B and the lottery between C and D. This is clear by putting data in for the degree of risk capacity.

Assume, for example, that the decision-maker expects to earn a sum of, say £1 million in the balance of his or her lifetime, and has only modest net worth today. The decision-maker might refuse any lottery that would entail a loss of £50,000, no matter how great the expected value. We can use that information to determine the scaling constant in the decision-maker's RAV formula. For reasons that will become clear later, we call this scaling constant the decision-maker's risk capacity. The risk load (the amount by which the expected return exceeds its risk-adjusted value) for a lottery might look like this:
The subject's flinch point is perhaps £50,000, and his risk capacity £25,000. In this world, the value of Alternative A is £1,000,000, and the value of Alternative B is only £115,129.

\[
RAV = -£25,000 \ln \left[ .10e^{\frac{£5,000,000}{£25,000}} + .89e^{\frac{£1,000,000}{£25,000}} + .01 \right] = £115,129.
\]

(2)

We are not surprised to find that our subject prefers the certainty of A to the small chance of doing better with B.

When faced with alternatives C, and D, however, the decision-maker sees much less value in, and much less difference between, the alternatives he or she might realize. Again assuming a risk capacity of £25,000, the alternatives are worth:

\[
RAV = -£25,000 \ln \left[ .11e^{\frac{£1,000,000}{£25,000}} + .89 \right] = £2,913
\]

(3)
\[ RAV = -£25,000 \ln \left( \frac{\frac{£5,000,000}{£25,000} + .90}{1.10e} \right) = £2,634 \] (4)

These are nearly the same. One is nearly the same as a 17% chance of receiving a blue car worth £25,000, and the other is nearly the same as an 16% chance of receiving a red car worth £25,000. Perhaps the gamble for £5,000,000 is more thrilling, just as a red car might be more thrilling. As easy as it is for us to compute the economic advantage of taking a chance for the blue car, real decision-makers will sometimes prefer the chance for the red car. Put differently, it is no criticism of a prescriptive theory of decision-making that it does not predict financial decisions when the decisions are not financially significant.

Allais’s paradox does not appear to be a paradox when actual data is used to quantify the risk charges involved.

**Risk-Adjusted Value as a Function of Risk Capacity**

We do not need to invoke different degrees of risk aversion, as Clarkson suggests. Nor do we need to estimate the firm's risk capacity. It is sufficient to consider the Risk-Adjusted Value of each alternative under some consistent utility function which has a scale variable for risk capacity, and to plot every alternative's risk adjusted value as a function of risk capacity. The resulting graphs present a curve for each alternative which increases with risk capacity toward the expected value of the outcome and declines with risk capacity toward the value of the worst possible outcome.
Figure 3 shows the values of the four alternatives of Allais's paradox as a function of risk capacity for an exponential utility function. Note that Alternative A is preferable to B if risk capacity is less than about $420,000. Note also that Alternative D is preferable to C if risk capacity is less than the same amount. It is also clear that Alternatives A and B are worth quite different amounts if risk capacity is less than $300,000, while alternatives C and D are worth about the same amount. This makes it clear why the paradox appears: the first choice is clear, while the second choice is not clear. The two lotteries are dramatically different even though they have a preference reversal (the lines cross) at precisely the same level of risk capacity.

Figure 3 also reminds us that the paradox does not appear if the decision-maker is a multi-billion dollar corporation. The large corporation, with large resources and many large and relatively independent risks, would prefer Alternative B with its expected income of £1,390,000 over Alternative A with its expected income of £1,000,000.
In general, graphs of Risk-Adjusted Value as a function of risk capacity provide either clear guidance into the appropriateness of the alternatives, or a clear statement that the alternatives are not significantly different as they have been defined.

4. Implications of the General Theory

**More Theory**

*Attention to Detail, and the Importance of Scenarios*

One implication is that one must consider the entire distribution of outcomes. Although the formula (1) is simple to compute in practice, it demands the analyst do his homework to identify the risks and quantify the implications of the risks on the financial outcomes. As numerous writers have shown, the risk-adjusted value of the firm includes all of the information in the probability distribution of outcomes (because it is a linear transformation of the cumulant generating function).

Equation (1) only deals with the possible outcomes at a given time. When outcomes are possible over a range of times, as is usually the case, the outcomes at various times may be related to one another. For example, high inflation might affect a series of loss payments over time. An example from an investment decision is that a slow-down in the economy could affect consumer demand in a series of time periods. Interrelationships like these, if they exist, must be reflected in a set of two or more scenarios. For example, we might evaluate high inflation and low inflation scenarios, or we might evaluate high demand and low demand scenarios.

Taken together with the observation that one must consider the entire distribution of outcomes, the need to consider a range of scenarios (except in trivial cases) seems to imply a heavy burden on the analyst. One implication of the theory of finance is that successful decision-makers today do consider a range of scenarios, rather than simply the variance-adjusted rates of returns, when they make decisions. The rich literature on corporate planning suggests this is the case.
One implication of axiom 5 is that the computations of the risk-adjusted values for the outcomes for a given scenario at various times can be computed independently (Gerber (1979)). The resulting risk-adjusted values can be treated as amounts certain. Therefore, they can be expressed in current dollars and summed. The result is the risk-adjusted value of the alternative under the scenario. The range of these values across all scenarios is a probability distribution of outcomes, with each outcome expressed in current dollars and each equivalent to an "amount certain" given that the scenario comes to pass. These can be combined to compute the risk-adjusted value over all scenarios, which is the risk-adjusted value of the alternative.

Nothing in these axioms permits outcomes at various times or various scenarios to be aggregated together (except for outcomes that are certain). For example, a 10% chance of losing $100 at time t=5 under an inflationary scenario can't be aggregated with a 10% chance of gaining $100 at time t=6 under that scenario or a 10% chance of gaining $100 at time t=5 under a different scenario.

The sequence of calculations must be:

1a. Estimate RAV(j,t,c) at each time within each scenario

1b. Estimate RAV(j,c) across time for each scenario

2. Estimate RAV(c) across all scenarios

The time value of money can be considered before or after estimating RAV(j,t,c), the risk-adjusted value at a particular time for a particular scenario. It must be considered before estimating RAV(c) across all scenarios. Doing the calculations in some other order either destroys the idea of the scenario as a way to link outcomes over time or blurs the accounting for the time value of money.

Spreadsheet software makes these calculations straightforward once the data is assembled. The analyst merely sets up one spreadsheet for each scenario, with logic and parameters to reflect the various alternatives being considered. Each spreadsheet has, say, a row for each point in time at which cash might flow in or out, and a range of results for that row,
expressed either as discrete outcomes or as the parameters of a probability distribution of outcomes. One example of the discrete outcomes the analyst might use is the output of a corporate financial model such as that described by Daykin and Hey (1990).

In practice, one can usually assume fixed values for RAV(j,t) rather than compute RAV(j,t,c) from a probability distribution. Merely specifying a detailed set of scenarios and being a little pessimistic in stating the value to the firm of each result will usually produce an adequate approximation for RAV(j,t,c). These assumed values of RAV(j,t,c) can be expressed in current dollars using published rates of return for risk-free investments. Generally, the only risk-adjustment that must be done explicitly is the calculation of RAV(c) itself.

The axioms imply that the economic value of the firm is of the form in expression (1), except for a linear transformation. Therefore, the decision-maker will get the same answer if he uses expression (1) or a linear transformation of expression (1) that more nearly approximates the economic value of the firm.

**RAV Is Always to be Measured at the Margin**

Often the results of investments or underwriting commitments are not correlated with the other results for the firm. When this is the case, the risk-adjusted value of the investment or underwriting commitment is independent of the other commitments of the firm, and can be measured using the probability distributions and time value of money for the investment itself.

More generally, firms often invest in or underwrite risks with outcomes that are not independent of the results of the firm's other activities. The lack of independence could be intentional, as it is when a life insurance company decides to underwrite more life insurance, or structural, as in the case of a company whose loss experience and investment experience will rise and fall with changes in the economy. When this is the case, reasonable estimates of the risk-adjusted value using equation (1) must be made for the firm as it will be under each of the alternatives being examined. The risk-adjusted value of any
alternative is computed as the increase in the risk-adjusted value of the firm associated with pursuing the alternative.

This mechanism of measuring the marginal increase in the risk-adjusted value of the firm correctly takes into account the correlations between the results of the alternative investment or underwriting commitment and the firm's pre-existing portfolios of investments and underwriting commitments. For example, the operating profits of insurers of automobile physical damage coverage have been shown to be negatively correlated with overall economic growth, while changes in the yields on low-risk securities have been shown to be positively correlated with overall economic growth; measuring the risk-adjusted value at the margin recognizes the interaction between these two, and suggests that, all else equal, more auto physical damage insurance can be underwritten if investments are in short-term securities (with yields following new-money rates) rather than in long-term securities.

The Unit of Risk

For the study of finance to be a quantitative science, we must have a unit of measurement of risk. Units of measurement are usually named after scientists who were instrumental in the founding of the science. They should be of such a magnitude that everyday observations run from .001 units to 1,000,000 units, if practical. For example, temperature is measured in degrees Fahrenheit or degrees Centigrade; both are observed in the range from zero to 100 in everyday experience. (I have heard that the unit of measurement for beauty is the Helen, and that a millihelen is the amount of beauty required to launch one ship.)

I propose, at least for the time being, that a workable unit of risk for business finance is the risk of assuming a 10% chance of losing DM100 million in exchange for a premium of DM10 million (with the probabilities independent of the firm's other results). I further propose that the unit of risk be named the Borch, in light of Karl Borch's demonstration that risk premium will be proportional to units of risk in an efficient market (Borch (1962)) and his many contributions to the understanding of risk. This unit is of such a size that one Borch would be a large enough risk that, if undertaken without compensation, a typical company president would be dismissed by his or her Board of Directors.
The risk in underwriting $10 million of auto business would be about a milliborch, and the risk associated with the Oakland, California, firestorm would be about a kiloborch.

*The Cost of Capital Is Not the Same as the Reward to Capital*

One implication is that decision-makers should distinguish between the cost of capital and its rewards. The only cost of capital is the cost of uncertainty. A partial list of the rewards to capital are the gains arising from applying that capital in support of talent, innovation, industry, growth, and successful management of risk.

Moreover, the time value of money enters as a cost when risk-free yields are positive only in the case that the alternative being evaluated is characterized by cash outflows prior to cash inflows. If the alternative is to *borrow* money and repay it later and risk-free yields are positive, then one will be repaying the debt in cheaper units of currency. In this case, as every borrower enjoying unexpected inflationary increases in income knows, the time value of money is a benefit, not a cost. The cost is the interest on the loan.

*The Rewards to Capital Are Not Related to the Variance of Returns*

Of the rewards of capital, only the cost of risk is even remotely related to the variance of returns. The results of formula (1) are similar to a variance loading only if risks are normally distributed, which is rarely the case, or if risks are small. Rewards are generally expected to be unrelated to the variance of returns. (Moreover, Roll and Ross (1994) show that market index proxies (baskets of investments representing market averages) can be very close to the efficient frontier and still, in principle, have returns uncorrelated with betas. That is, is it nearly impossible to measure risk by looking at betas of real stock index measures. This is another demonstration that the mean-variance assumption of the CAPM is useless in practice to predict the returns on stock portfolios.)

Researchers have noted for years that stock market returns are not well represented by the Capital Asset Pricing Model. For example, the first two articles in the December, 1994, issue of The Journal of Finance
seek to explain the evidence that certain variables explain the cross-sectional variation in realized stock returns. (Lakonishok, Shleifer, and Vishny (1994) and Davis (1994)).

Cash, Currency, Risk-Adjusted Value, and Economic Value of the Firm

Nothing in the axioms requires the analysis to be done in terms of cash. The analysis could be done in terms of statutory earnings, employee count, or any other measure of success.

The time value of money is always denominated in a currency. Therefore, if the outcomes are not denominated in some government’s currency, then the risk-adjusted value of the outcomes at a given time must be converted into a currency using a linear transformation (multiplying by some constant relating units to currency) before the values at various times can be aggregated. Once aggregated, the resulting value for the scenario must be converted back into the units being used for the analysis.

In addition, the economic value of the firm is always denominated in a currency. Therefore, the economic value of the firm will be related by some linear transformation to the risk-adjusted value computed from cash.

In practice, cash is often easiest.

Claims about What Can Be Observed: Simple Cases

Most of the following simple cases have been explored by other writers. I have attempted to give credit where I can, but regret that I have probably not found the original author of many of these conclusions.

Internal Rate of Return on an Investment

First, the internal-rate-of-return rule of traditional finance emerges as a special case of the new theory. In one of the most interesting results, the new theory predicts that in certain factual situations, the alternative that should be selected is the one with the highest internal rate of return. Interestingly, this arises only for decisions regarding investments, that is,
alternatives in which the timing of cash outflows precedes the timing of cash inflows.

Consider an investment with a fixed time horizon and increasing uncertainty subject to some upper bound. Assume also that the time value of money can be expressed as a constant rate of discount, that the uncertainty is reasonably measured by the variance of the possible outcomes, that there is a single scenario, and that investment's outcomes are independent of the other outcomes the firm will realize. Investments with these qualities are common; an investment in an industrial bond is a good example.

The decision rule in light of the risk-adjusted value is to maximize the following sum:

$$\sum_{t>0} v(t) \text{RAV}(t) > I$$

(5)

In the particular case at hand, this becomes:

$$\sum_{t>0} (1 + R_f)^{-t} \left( \mu(t) - \frac{\sigma^2(t)}{2c} \right) > I$$

(6)

where $R_f$ is the cost of money and $\mu(t)$ is the expected return at time $t$. Then if

$$\sigma^2(t) = 2c\mu(t) \left( 1 - \left( \frac{1 + R_f}{1 + IRR} \right)^t \right)$$

(7)

the decision rule is equivalent to

$$\sum_{t>0} (1 + IRR)^{-t} \mu(t) > I$$

(8)

This is, the risk-adjusted value rule says to undertake the investment that has the maximum present value at discount rate IRR. The internal-
rate-of-return rule has a long history of success when applied to straightforward investments because it is a simple and intuitively appealing approximation to the risk-adjusted value rule. It is clear now that the internal rate of return rule is not itself a first principle of finance in the sense that the seven axioms are first principles.

Payback Rules

Many other useful decision rules also arise as special cases. Payback rules are appropriate if the uncertainty is small during the payback period and increases greatly at the end of the payback period. This is a typical situation for a manufacturer, who can expect to use a given process for a known period of time, but not much longer.

Underwriting

The internal-rate-of-return decision rule emerges only if the alternative is an investment. If the alternative is an underwriting commitment which will be realized with only a negligible time delay, then the risk-adjusted value rule simplifies to an exponential risk load. This is the case studied by Borch, Gerber, and others cited above.

Game Theory

If all of the alternatives involve a very high risk, then the general theory suggests one should choose the strategy that maximizes the firm's minimum final position. This is the special case studied by von Neumann-Morgenstern game theory.

Variance Loadings

Variance loadings are appropriate only for the cost of capital, not for the time value of money. In addition, variance loadings are appropriate for the cost of capital only if the amounts at risk are small or if the outcomes are approximately normally distributed.

The risk-adjusted value at a particular time for a particular scenario may be approximated by a variance loading, plus a variance loading on the loading, if the losses are distributed with a one-tailed distribution and
that distribution is approximately a gamma distribution. For example, if the variance loading would be 20%, the risk loading would be 20% plus 20% of 20%, or 24%. For underwriting commitments, this type of distribution of outcomes is far more realistic in practice than the normal distribution that is consistent with the variance loading.

Loadings based on the semideviation or semivariance are inconsistent with one or more of the axioms set forth above.

**Claims about What Can Be Observed: Old Observations, New Theory**

*Alternatives That Reliably Lead To Growth Of Market Share Are Strongly Preferred.*

Reliable, profitable growth is discounted at only the time value of money indicated by the currency markets. This is approximately the risk-free, after-tax rate of return. Drucker (1989, p. 116) described the new reality of international trade in the 1980's in these terms: "In the transnational economy the goal is not 'profit maximization.' It is 'market maximization.'"

On the other hand, growth that is associated with increasing uncertainty and even risk of loss of capital is strongly penalized. The cost of uncertainty in the future is treated as a real cost, and is discounted at only the time value of money indicated by the currency markets.

Moreover, when growth is financed by the issuance of debt (creating an obligation that must be repaid regardless of the success of the growth strategy), the repayments of principal and interest should be discounted at only the time value of money, not at the high rate at which the firm would have to offer the lenders. Subordinated debt at high interest rates—junk bonds—may be appropriate, but not because the firm "has a high cost of capital."
The Cost of Capital in Insurance Ratemaking Is a Percentage of the Premium Underwritten; the Percentage Depends on the Riskiness of the Insurance

The general theory implied by axioms 1 to 7 implies that the appropriate standard for profit measurement is that profit should be measured in proportion to risk, which, for insurance sales, is roughly in proportion to premium volume. Profit margins in insurance rates are not generally a function of the capital structure of the insurance firm. They are a linear function of the riskiness of the kind of insurance, which can be expressed in Borch's per million dollars of premium or some similar measure.

Real Markets Manage Capacity

One implication of this theory is that every investment or underwriting commitment has associated with it a measure of risk that is approximately proportional to the variance of expected results if the amount at risk is small enough. The converse is that essentially every investment or underwriting commitment has associated with it a measure of risk that is greater than some multiple of the variance if the amount at risk is small enough.

Together with the conclusion about real markets, this suggests that in real markets the players will undertake additional investment and underwriting opportunities until they have undertaken a sufficiently large level of commitment that they are valuing their risks at greater than some multiple of the variance. The players will take on additional activities until, in some sense, they flinch at taking on more. The general theory of finance suggests that the key management question is not, "What opportunity has the best profit margin?" The key question is, "With the expected profit margins and the risks, how much growth should we attempt?"
Claims About What Can Be Observed: New Claims

Measuring the Cost of Capital in the Capital Markets

In an efficient market, the total risk to be underwritten is allocated among the players in proportion to their capacity to bear risk (Borch (1962). Each receives the same risk premium per dollar of risk underwritten, and also receives the same risk premium per dollar of risk capacity.

For data to measure the cost of capital, we can observe the following in the capital markets for a given investment or underwriting commitment:

1. the probability distribution of the possible outcomes, expressed in current dollars, and
2. the price paid.

The amount by which the expected value of the returns exceeds the price paid for the investment is the observed risk premium of the investment. The observed risk premium for every investment is proportional to the amount of risk in the investment. For example, if the observed risk premium for an investment of $1 million in corporate Aaa bonds is $5,000, the risk premium for an investment of $2 million in such bonds is $10,000.

Using the probability distribution of the possible outcomes, one can use expression 1 to estimate the risk-adjusted value as a continuous function of the risk capacity brought to bear on the investment, RAV(c). The amount by which the expected value of the returns exceeds RAV(c) is the expected risk premium. The expected risk premium is also proportional to the amount of risk in the investment. It is not necessary for the decision-maker to decide on a level of risk aversion (denoted by the risk capacity, c). It is sufficient to estimate the risk-adjusted value of the firm in light of each alternative for a range of values of risk capacity, as long as one knows the market price for risk.
Figure 4, The Cost of Risk, shows the results for a few hypothetical alternatives. Consider first the solid line labeled "One Borch." This line shows the cost of risk for a firm facing one borch of risk for a wide range of risk capacities. In this figure, the line for one borch is well known (it is computed directly in the Appendix), but the quantity of risk capacity is unknown, a sort of hyper-parameter that cannot be directly observed. What can be observed in addition to the line for a risk of one borch is the cost of underwriting such a risk, which is the observed difference between the market value of a firm with one borch of risk and the expected value of the gains and losses for the firm.

Continuing the example, we might observe an investment fund facing essentially no risk with assets with a current market value of DM250 million, and a second investment fund with the same expected cash flows but with one borch of uncertainty, and see that the second investment fund is trading at a value of only DM235 million (DM250 million minus the cost of risk of DM15 million). So one can observe the point at which the expected cost of risk (based on probabilities of outcomes) is equal to the market cost of risk. In Figure 4, this would be a line at a height of DM15 million and a risk capacity of about DM50 million.

As Freifelder (1976) clearly showed (but was inferred by the work of Borch, Gerber, and other writers), the cost of risk must be marginally the same for all units of risk. Therefore, underwriting a risk of two borchs will command twice the risk premium as underwriting the risk of one borch. The line labeled "Two Borchs" shows graphically the implications of this statement. If a third firm is just like the two we have just described except that it faces an amount of risk equal to two borchs, its value will be diminished by exactly twice the cost of risk as the firm that faces a risk equal to one borch. We expect to observe in the market that the third firm's price reflects a cost of risk of DM30 million. All actual prices of the cost of risk should appear on a single vertical line. In this example, that vertical line is at roughly the position indicated by a risk capacity of DM50 million.

The insight that "the cost of risk for an independent risk is directly proportional to the size of the risk" leads directly to the conclusion that the capital markets behave as if they had a certain level of risk capacity.
This is an extremely powerful conclusion. It says that the capital markets have a price for risk that can be observed as a linear relationship between risk and price. As many authors have shown, the capital asset pricing model's prediction of a linear relationship between variance and return has not been found in the markets. Figure 4 says that the linear relationship will emerge if deferred amounts are brought to current value using the prices in currency markets and risk is measured using probability distributions and expression (1).

Observations of capital markets will show the relationship between risk and its cost. This might be, for example, DM15 million per borch. Once this relationship has been observed, the decision-maker can decide the market share of that risky opportunity that corresponds to his or her firm's own risk-bearing capacity. As Karl Borch (1962) showed, market shares will vary in direct proportion to the several firms' willingness to bear risk.

The line "20% of c" illustrates a rather different way to view this result. Up until now, we have been interpreting Figure 4 as a portrait of the cost of risk for a particular firm. Consider for a moment the allocation of the costs of risk among a number of firms. As Borch showed, firms will undertake risks in proportion to their ability to bear risk. They will share in the total cost of risk of a venture (the market price for underwriting the whole venture) in these proportions. From firm to firm, then, the market price for the cost of risk will appear as a certain fraction (or multiple) of each firm's risk capacity. The slope of that line will be a reflection of the total capital in the world's capital markets and the ventures that are presented to those capital markets. It will therefore change slowly in response to major shifts in society (such as the introduction of the steam engine in the late nineteenth century or the introduction of computerized trading in the late twentieth century.) It may change more quickly if there is a worldwide change in the perception of risks, as happened before and after the two world wars.

According to the efficient market principle, the reward for risk should be the same in investment markets as in underwriting markets. (If it were not, capital would flow from markets with low rewards to markets with high rewards.) Actual data like that in Figure 4 can be observed in investment markets and inferred for underwriting markets.
Risks That Are Not Independent

The great insight of the capital asset pricing model is that the cost of risk will be affected by the fact that the outcomes of the aggregate of all the risks underwritten by the capital markets will be correlated with one or more economic variables. Ross (1976) shows that the outcomes can always be decomposed in terms of results that are correlated with an index built of economic statistics (or, in principal, two or more such indexes) and random noise. It is therefore important to see how the cost of risk in the capital markets will be affected by the lack of independence of outcomes.

In Figure 4, the line labeled "Double Whammy" shows the cost of risk for a risk of a 10% chance of losing DM200 million in exchange for a premium of DM20 million. (The "whammy" was popularized in the United States by cartoonist Al Capp, whose character Evil-Eye Fleegle would upset the good folks of the community of Dogpatch by casting his famous "whammy." When he was out to wreak even more havoc, Evil-Eye would cast a "double whammy.") Firm A facing a risk of a double whammy is facing the same risk as Firm B with an existing risk of one borch that then underwrites a second risk that is 100% correlated with those existing outcomes.

The shape of the line for the double whammy has two key characteristics. On the left, for small levels of risk capacity, the cost of risk is intermediate between the cost of risk for two borchs and a simple payment of DM180 million. (When the firm has a small ability to bear risk, the cost of risk is primarily a function of the worst possible outcomes. As a result, the cost of risk for two borchs and for one double whammy both approach DM180 million. This illustrates the conclusion that the strategy of von Neumann-Morgenstern game theory that the decision-maker should seek to maximize the firm's minimum final position is a special case of the actuarial cost of risk.) On the right, for large levels of risk capacity, the cost of risk is nearly equal to that of four borchs. (When the firm has a large ability to bear risk, the cost of risk is essentially a function of the variance of the outcomes, and the variance of the double whammy is the same as the variance of four independent borchs.)
The insight of the capital asset pricing model, and more precisely of the factor explanation of arbitrage pricing theory as set forth by Roll and Ross, is that actual capital markets will contain prices for risks that are correlated with some economic factor, and higher (or lower) prices for risks that are correlated more (or less) with this factor. This implies that the markets will have a price for risks like the double whammy. In Figure 4, the market that would charge a cost of risk for two borchs of DM30 million would charge a cost of risk for one double whammy of DM75 million. The extra cost of risk of DM45 million arises from the correlation of 1.00 between the first gain or loss and the second gain or loss.

The capital asset pricing model proposed a linear relationship between the cost of risk and the correlation of results with some broad market average. That relationship was expected to have an intercept of zero and a slope denoted by beta.

The actuarial model suggests that there is such a relationship, but that the intercept is a constant (the cost of risk for an uncorrelated unit of risk), and that both the intercept and the slope can be observed by the linear relationship that will emerge if expectations of deferred amounts are brought to current value using the prices in currency markets and risk is measured using probability distributions and expression (1). The intercept is not at zero, and the slope is not precisely the covariance of returns with the market factor.
This market line is the underlying cost of capital in the capital markets. As noted above, the market line predicted by the Capital Asset Pricing Model has not been found. The market line predicted by the Capital Asset Pricing Model is inconsistent with axioms 1 to 7. The general theory consistent with axioms 1 to 7 predicts that the market line of figure 5 will illustrate the cost of capital that should be recognized by the decision-maker.

This provides a direct empirical test of the usefulness of the general theory. If the cost of capital for a set of reasonably well defined investments does not follow a somewhat linear relationship, the pattern that does emerge might at least indicate the flaw in the axioms. Given the simplicity and explanatory power of the general theory, however, this seems unlikely.

Of course, actual investments will not be in fact perfectly linearly related. Observed differences between the average market line and the risk premium for individual investments and underwriting opportunities will indicate either differences in information or real opportunities for arbitrage. Further research can indicate which departures offer real opportunities for arbitrage.

Data must reflect the correct time horizon. Measuring the cost of risk in the capital markets requires the description of the probability
distributions of outcomes. These probabilities are a function of the decision-makers' time horizons, and must be estimated for the time horizons that investors have in mind. Given the large role that institutional investors have in creating the world's capital markets, the probabilities must be measured over the time frames that institutional investors actually use in evaluating investment alternatives. This time frame is usually a period of years. Therefore, we would not expect to see a clear linear relationship if the risks and rewards are estimated from data about monthly returns on investments in common stocks.

Diversifiability

So much has been said about the diversifiability of risks since the introduction of the capital asset pricing model that it is appropriate to consider the subject briefly here. Let us change the description of the problem somewhat, and consider not the risk of underwriting one or more borchos but the risk of underwriting some fraction of a borch (or any other underwriting possibility). Assume that any opportunity that can be presented can be fragmented, and that every firm can underwrite just that part of each risk that it wishes. The implications are shown in Figure 6.

Figure 6 shows the risk-adjusted cost of underwriting some fraction of each risk, with the fraction selected to result in a unit variance. (Because the units are really millions of German marks, one unit of variance is really $10^{12}$ DM$^2$, but this is immaterial to the argument.) The line for one borch is for one part in 30 of the outcomes of a borch. The same capital market that would price a borch of risk at DM15 million would price this fraction of a borch at DM166,667. And if a risk of one borch were to draw a risk capacity of DM50 million, this small risk would draw a risk capacity of DM1.67 million.

Similarly, a risk of two borchos would be divided into 42.43 parts, each with a unit variance. The outcomes of two borchos are rather more dispersed, per unit of variance, than the outcomes of one borch. As a result, at small values of risk capacity the cost of risk would be slightly greater for a part of two independent borchos than for a part of one borch, per unit of variance. At large values of risk capacity the cost of risk
would be slightly less for a part of two independent borchs than for a part of one borch.

Let us denote the expected risk premium for an investment with a variance of one unit by \( p(c) \), and let us denote the observed risk premium for an investment with a variance of one unit by \( p \). Formally, for a set of outcomes over a range of scenarios all expressed in current dollars,

\[
p(c) = E\left[\frac{x}{\sigma}\right] + c \ln E\left[\frac{1}{e^{\sigma c}}\right]
\]

where

\[
\sigma^2 = E\left[(x - E[x])^2\right]
\]

Also,

\[
p = E\left[\frac{x}{\sigma}\right] - \frac{\text{price}}{\sigma}
\]

Note that the line for the double whammy is the same as the line for one borch. This is because the risk of the double whammy has been diversified, and the firm is only participating in one-half as much of the double whammy as it would of the borch alone.

Figure 6 also shows a line for which the cost of risk is inversely proportional to the variance. This would be the case if outcomes were normally distributed (with zero mean and unit variance). The law of large numbers says that the outcomes of a bundle of identical, independent risks will approach a normal distribution as the number of independent risks grows large.

The normal distribution of outcomes is of course not the least risky. Although not shown, many risks are characterized as having a smaller chance of catastrophic loss (for a given variance) than the normal. These have a cost of risk per unit variance less than that of the normal. A group of risks characterized by equal amounts of gain and loss (as in a coin
A GENERAL THEORY OF FINANCE

toss) has a lower cost of risks, and approaches the normal distribution from below as the number of independent risks is increased (for a large value of risk capacity).

Implications for the Cost of Risk in Insurance Contracts

There are many implications for those who make decisions about insurance, whether as underwriters, actuaries, marketers, investors, or regulators. Among the implications are:

There is a cost of risk for investors in insurance companies. That cost of risk is a function of the probabilities of receiving dividends and changes in the market value of their investments. The cost per unit of risk is determined by the worldwide capital markets.

There is a cost of risk for insurance companies when they underwrite a set of insurance contracts. That cost of risk is a function of the probabilities of gains and losses, in the context of the time value of money in risk-free investments. For a given unit of risk, the cost of risk is the same to the insurance company as to the investor. It is the risk that differs from the point of view of the investor, not the cost per unit of risk.

If the cost of risk in underwriting $20 million of auto insurance is $1 million, the cost of risk in underwriting $40 million of auto insurance is $2 million. That is, the cost of risk is a percentage of the premium, and the percentage varies from one kind of insurance to the next depending on the riskiness of the kind of insurance. (The same statements can be said about the markets for commodities, or common stocks, or any other type of risk.)

The cost of risk in underwriting $20 million of auto insurance is independent of the capital structure of the firm. It is the same for a company with $10 billion of net worth as for the company with $10 million of net worth. The cost of risk per dollar of premium is set by the worldwide capital markets, not by the insurance companies. What varies from firm to firm is the financial ability to bear risk, and hence the appropriate level of premium volume.
Lines of insurance whose outcomes are correlated with broad market-baskets of possible investments have a higher cost of risk than lines of insurance whose outcomes are independent of such broad market results. Lines of insurance that have favorable results when the broad market results are unfavorable act to hedge such investments, and have a lower cost of risk.

Reinsurance may improve the value of the firm by reducing the firm's cost of risk.

The goal of management is to have the expected value of income exceed the sum of all costs, including the cost of risk. By this measure, many companies have been losers for decades.

Measuring the cost of capital, like measuring any other cost, is an important management function. The firm that measures the cost of capital and underwrites only those insurance products that provide for the cost of capital will outperform the company that does not over the long run.

*Expectations for Improved Risk-Bearing Ability*

The arrival of a new chief executive officer might increase or decrease the market value of a firm. The actual movement of the firm's stock price will depend on the investment community's expectations for the new CEO. Most investment analysts would consider the CEO's ability to increase the expected level of income and reduce the uncertainties associated with the income streams. The general theory of finance suggests that there is another way in which the new CEO can increase the economic value of the firm: he can increase the firm's ability to bear risks. To the extent that analysts today do not recognize this as much as they should, the new theory of finance suggests there are arbitrage opportunities.

*Markets Without Transaction Costs Diversify Risks*

Much has been written on the implications of competitive markets for goods and services on the diversification of risk. In the absence of transaction costs, markets will diversify risks until every firm holds an
infinitesimally small portion of every risk. This process will continue until all risks are fragmented down to the equilibrium level. Venter (1991) has shown that this result holds for underwriting risks as well as investment risks.

Real Markets Absorb Risks Great Enough to Have Risk Costs Greater than Implied by the Variance Principle

Markets do not, in fact, diversify risks as predicted by Venter. Further thought about the efficient market shows why they do not. The key is that there are always transaction costs. This means that it will always be in each firm's interest to hold a slightly greater amount of risk, and expect a slightly greater amount of reward, so long as the reward is less than the transaction cost required to eliminate the risk.

There are two ways a firm can hold a slightly greater amount of risk. First, if the player perceives that risk is proportional to variance, he or she can undertake more of the business. Second, the player can undertake investments for which the risks are perceived to be greater than indicated by the variance alone. Such risks would never be undertaken in the absence of transaction costs because they are not diversifiable. In the presence of transaction costs, however, risks will not be fully diversified anyway, and there is no obstacle to undertaking risks that are not fully diversifiable.

In a real market in which there are transaction costs and in which there are opportunities which are perceived to have a risk that is greater than in proportion to their variance, these high-risk opportunities will in fact be undertaken. That is, real markets generally will be characterized by investments and underwriting commitments that have a risk measure more risk-averse than the variance of their expected returns.

Non-Diversifiable Risks Will Have High Transaction Costs

An empirical implication of this is that markets with high transaction costs should be observed to have relatively high levels of non-diversifiable risks, and that markets in non-diversifiable risks should have relatively high transaction costs. As a first-order approximation of empirical data, we point out that London reinsurance brokers have high
transaction costs and arrange transfers of highly unusual risks, and that
direct sellers of non-medical life insurance have low transaction costs and
sell highly diversifiable risks.

5. Conclusion

This paper has suggested that a revolution in the theory of finance is
underway. Leading observers such as Peter Drucker have said that some
of its most surprising qualities are already present in the behavior of
firms around the world. One principle of the revolution is that
observations must be used to check the theories. Another principle is that
management should focus on maximizing the economic value of the
firm.

A new, general theory of finance can be built on the structures
erected by the giants of the actuarial community. Explicit recognition of
probability distributions will replace measures such as the variance, the
probability of ruin, and the semivariance, which are only appropriate in
special cases that often are not relevant. One result of an axiomatic
approach is that the time value of money is distinguished from the cost
of risk. The rewards to capital are many; the cost of capital is a function
of the capital markets, and not of the capital structure of the firm.

This general theory of finance makes "claims about what can be
observed, claims that would not have seemed plausible prior to the
advancement of the theory, but that are in fact found to be true when we
make the appropriate observations," as Philip Kitcher has said a science
must do. Perhaps its greatest strength today is that it explains a wide
range of observations which previous theories explained only by citing
special cases. It is also important in that it implies that data from the
capital markets can be used to directly calculate the cost of capital for a
unit of risk.

Perhaps this theory is wrong. Perhaps there is an error in the
mathematics, or some other axioms apply. That is of little consequence.
All scientific theories are correct only in the context of the knowledge of
the time. What is of consequence is that there is a revolution in the theory
of finance underway, and powerful tools are available to explore and test
the implications of any new theory. Many observers have reported data
that are inconsistent with the established theories of finance, including the return on equity approach to capital budgeting and insurance ratemaking and the Capital Asset Pricing Model for portfolio management. Now it is time to test the new theory with further predictions and observations. As we do, we can expect to find many claims that would not have seemed plausible, and if the theory is correct they will be found to be true.
Fig. 4  The Cost of Risk
Fig. 6 Diversification:
The Cost of Risk for Unit Variance

![Graph showing the cost of risk for unit variance with different risk capacities.](image-url)
BIBLIOGRAPHY


