INVITED PAPER

ASSET PRICING MODELS AND INSURANCE RATEMAKING

By J. David Cummins

The Wharton School of the University of Pennsylvania, Philadelphia, USA

Abstract

This paper provides an introduction to asset pricing theory and its applications in non-life insurance. The first part of the paper presents a basic review of asset pricing models, including discrete and continuous time capital asset pricing models (the CAPM and ICAPM), arbitrage pricing theory (APT), and option pricing theory (OPT). The second part discusses applications in non-life insurance. Among the insurance models reviewed are the insurance CAPM, discrete time discounted cash flow models, option pricing models, and more general continuous time models. The paper concludes that the integration of actuarial and financial theory can provide major advances in insurance pricing and financial management.

Keywords

Asset pricing models; financial pricing models; insurance pricing; portfolio theory; option pricing; capital asset pricing model; arbitrage pricing theory; discounted cash flow models; insurance default risk.

Introduction

This paper examines insurance pricing using the techniques of financial economics. Financial economics focuses on topics such as investment decision making, the pricing of assets, and corporate financial management. These topics are important in insurance because insurance companies must choose financial and investment strategies and make decisions about projects (policies). Financial theory views insurance policies as financial instruments that are traded in markets and whose prices reflect the forces of supply and demand. Models of insurance pricing based on financial concepts are called financial pricing models.

Traditional actuarial models tend to reflect a supply-side orientation and often assume implicitly that prices are set more or less unilaterally by the insurance company, sometimes to satisfy an exogenously imposed ruin constraint. More modern actuarial models (e.g., Borch (1974) and Buhlmann (1980, 1984)) recognize the role of supply and demand in determining price, usually by modeling a market in which buyers and sellers of insurance are risk-averse utility maximizers.

The primary difference between actuarial supply-demand models and insurance financial pricing models is that the latter place more emphasis on the behavior of owners of insurance companies and the role of financial markets in determining investor behavior. Usually, the assumption is that owners are diversified investors who hold insurance company shares as part of broad-based portfolios. The ability to diversify through financial markets means that they will be less concerned with risk that can be eliminated through portfolio diversification and more concerned with the risk that cannot be eliminated. Diversifiable risk is risk specific to individual securities and is often called *non-systematic risk*. Non-diversifiable risk, known as *systematic risk*, is risk that is common to all securities.

The most basic asset pricing models such as the capital asset pricing model (CAPM) imply that investors receive a reward or risk premium (risk loading in actuarial terminology) for bearing systematic risk but do not receive a reward for bearing unsystematic risk because the latter can be "costlessly" eliminated through diversification. Because investors are diversified, basic financial theory implies that the rates of return on widely-held insurance companies do not reflect unsystematic risk so that it would be incorrect to model insurer behavior using utility theory, which typically does not distinguish among types of risk. More recent theory has begun to moderate the rather stark conclusions of the CAPM, recognizing the importance of risk of ruin and the implications of incomplete diversification (e.g., GREENWALD and STIGLITZ (1990)). However, the fundamental insights of asset pricing theory carry through: even imperfect diversification is a powerful economic reality in markets for capital as well as markets for insurance.

To summarize, actuarial models tend to focus on supply and demand in insurance markets and typically do not give much attention to the behavior of company owners beyond the assumption that they are risk averse.[1] Financial models tend to emphasize supply and demand in the capital markets and typically neglect the product market beyond the implicit assumption that insurance buyers are willing to pay more than the actuarial values for insurance.[2] There is an obvious opportunity for merging the product-market insights of actuarial models with the capital-market insights of financial models to gain a more fundamental understanding of markets for insurance.

Financial models, like actuarial models, are abstractions of reality based on assumptions about the phenomena under investigation. The assumptions used in a given field often seem strange and unrealistic to researchers from other disciplines. It is important not to get so caught up in an analysis of the

[1] These are obviously broad generalizations designed to draw general distinctions between actuarial and financial models. It would not be difficult to find exceptions to these generalizations in both the actuarial and the financial literature.

[2] For individuals, this is attributed to risk aversion, while for firms with widely traded shares it is attributed to tax advantages, market signalling, and other factors. Since widely held firms do not have utility functions under basic asset pricing models (e.g., the CAPM), such models predict that they do not buy insurance according to the same rules as individuals. See MAYERS and SMITH (1982).

assumptions underlying an abstract model that one misses the insights the model provides. Models can succeed on several different levels. The ultimate test of a model is its ability to predict. Models developed at a high degree of abstraction often prove to be extremely useful in explaining and predicting behaviour. An example from insurance economics are models of market failure due to information asymmetries and adverse selection (e.g., ROTHSCHILD and STIGLITZ (1976)). Another is the theory of modern option pricing, which is discussed in this paper. Even models that do not have high predictive power in real-world applications can be very useful if they provide insights into the understanding of complex phenomena. Thus, it is best to focus on the objectives of financial models—understanding and predicting economic phenomena—rather than primarily on the assumptions.

The objective of this paper is to provide an introduction to the principal results of asset pricing theory and its applications in insurance. Part I of the paper provides a discussion of asset pricing models. The mathematical development of each model is sketched and strengths and weaknesses are discussed. The purpose is to provide the foundations needed to understand the mathematics and intuition of the insurance applications. The reader should view the financial material presented here as no more than the basics. The literature in finance is vast and highly sophisticated. Readers interested in pursuing the topic in more detail should consult such excellent references as BREALEY and MYERS (1988) and LEVY and SARNAT (1984) for applied treatments and INGERSOLL (1987) and DUFFIE (1988) for a more rigorous mathematical approach.

Part II presents the financial models that have been proposed for the pricing of property-liability insurance. These range from early models based on the capital asset pricing model to more recent developments such as arbitrage pricing theory and option pricing models. Since the literature applying finance to insurance is growing rapidly, the review of the literature presented in this article should not be viewed as exhaustive. However, care has been taken to discuss principal results representative of the foundations of the field.

I. ASSET PRICING MODELS

1. RISK

Investment opportunities are created when firms, governments, and individuals, issue financial instruments to raise capital for production and consumption activities. Capital is provided by investors who purchase these financial instruments. Asset pricing theory studies the interaction between the supply of and demand for assets in a market context to determine asset prices and rates of return. A common feature of asset pricing models is that assets with higher risk must deliver higher expected returns in order to attract investors.

Thus, an essential feature of an asset pricing theory is its definition of risk and the formulaic relatonship between expected return and the theory's risk measure. Risk can be defined as "a property of a set of random outcomes which is disliked by risk averters (INGERSOLL, 1987, p. 114)", where a risk averter is usually defined as a person with a concave utility function. Rigorous analyses of this seemingly vague definition have been provided in the economics literature. Two of the most important papers are ROTHSCHILD and STIGLITZ (1970, 1971). One definition of risk that they study is the following: *Random variable Y is more risky than random variable X if X and Y have the same mean but every risk averter prefers X to Y.* They show that this definition and two others, i.e., (1) $Y = X +$ "noise", and (2) the distribution of Y has heavier tails than the distribution of X, essentially amount to a single definition of greater riskiness, while more conventional definitions such as variance comparisons can have significantly different implications.[3]

A detailed exploration of risk definitions would be beyond the scope of this paper. They are mentioned here to alert the reader that any theory of asset pricing incorporates, among other things, assumptions about the nature of investor preferences and the stochastic processes defining asset returns. Investors typically are assumed to be risk averse, but many theories impose stricter preference assumptions such as confining investors to particular classes of utility functions. Restrictions also are placed on the classes of stochastic processes that are admissible as descriptions of asset returns and investment opportunities. These preference and process restrictions underlie the risk-return relationships arising from the theories.

2. MARKOWITZ (MEAN-VARIANCE) DIVERSIFICATION

Much of capital market theory is based on the assumption that investors hold diversified portfolios. The first comprehensive theory of diversification was developed by HARRY MARKOWITZ (1952) and (1959). MARKOWITZ diversification forms the foundation for the capital asset pricing model (CAPM). Since MARKOWITZ diversification considers only the means and variance of asset returns, it is often called mean-variance diversification or mean-variance portfolio theory. Investors seek to form portfolios with high expected returns but are averse to risk, where risk is defined in terms of variances and covariances of returns.

Investors are assumed to make decisions on the basis of their beliefs about the means and variances of asset returns. Returns are usually defined as *holding period returns*. E.g., for a stock:[4]

$$R_{it} = (D_{it} + S_{it} - S_{i,t-1})/S_{i,t-1} \tag{1}$$

where R_{it} = the holding period return on stock i in period t,
D_{it} = the dividend on stock i in period t,
S_{it} = the price of stock i at the end of period t.

[3] This is a very cursory overview of these risk definitions and results. A rigorous treatment can be found in the cited RS articles.

[4] Similar formulas apply for other assets such as bonds.

The expected return on the stock i is denoted $E_i = E(R_{it})$. The covariance of returns between securities i and j is defined as $\mathrm{Cov}(R_i, R_j) = C_{ij}$. It is convenient to denote the variance of return as $V_i = \sigma_i^2 = C_{ii}$. For portfolios, the corresponding notation is $E_P = E(R_P)$, for the expected return, and $V_P = \sigma_P^2$, for the variance.

The assumptions underlying MARKOWITZ diversification are the following: (1) Investors are risk averse expected utility maximizers, with utility functions satisfying $U'(W) > 0$ and $U''(W) < 0$, where W = wealth and primes indicate derivatives. (2) Portfolio decision making is based on means and variances of portfolio returns, with $dU/dE_i > 0$ and $dU/\sigma_i^2 < 0$. These assumptions imply that investors prefer *efficient* portfolios, defined as portfolios with the highest return for a given level of risk (σ_P) or alternatively as the lowest risk for a given level of return (E_P). Efficient portfolios are said to *dominate* inefficient ones. E.g., portfolio A dominates portfolio B if $E_{PA} > E_{PB}$ and $\sigma_{PA} \leq \sigma_{PB}$, where PA and PB refer to portfolios A and B, respectively.

Mean-variance diversification theory requires a strong assumption about either the form of the investors' utility functions or the distribution of security returns. One assumption that has been used to justify mean-variance decision making is that investors have quadratic utility functions, i.e., $U(W) = W - bW^2/2$. Since the first derivative of this function eventually becomes negative, the additional assumption is required that outcomes are confined to the range of increasing utility. Entirely apart from this potentially troublesome assumption, quadratic utility has been discredited because it implies increasing absolute risk aversion, which most economists consider unrealistic.[5]

Because of the limitations of quadratic utility functions, mean-variance diversification is usually justified through the assumption that asset return distributions are multivariate normal.[6] Although the lognormal provides a better empirical model of security returns than the normal, the normality assumption is often adequate as an approximation. For this reason, and because of its intuitive appeal, the mean-variance model has been used extensively in practical applications.

To achieve MARKOWITZ (mean-variance) diversification, the investor solves the following optimization problem:

[5] The Pratt-Arrow coefficient of absolute risk aversion is defined as $A(W) = -U''(W)/U'(W)$. For the quadratic, $A(W) = b/(1-bW)$, which is increasing in W. Intuitively, absolute risk aversion is interpreted as the tendency of the decision maker to accept a gamble of a given size. Increasing $A(W)$ implies that the decision maker would be less likely to take an actuarially favorable gamble of a given size as wealth increases. Thus, a millionaire would be less likely to take a favorable \$ 100 gamble than a pauper (assuming equal values of b). A related concept is that of relative risk aversion, $R(A) = W * A(W)$. Intuitively, relative risk aversion indicates the decision maker's aversion to a gamble involving a specified proportion of his wealth. Decreasing absolute and constant relative risk aversion usually are considered reasonable assumptions.

[6] The normal is not the only distribution for which investor preferences can be defined solely in terms of means and variances. A wide class of distributions called *elliptical* distributions has this property. The lognormal distribution is not a member of this class. See OWEN and RABINOVITCH (1983).

Minimize Over x_i, $i = 1, 2, \ldots, N$:

$$\sigma_P = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j C_{ij}} \tag{2}$$

Subject to:

$$\sum_{i=1}^{N} x_i = 1 \tag{3a}$$

$$E_P = \sum_{i=1}^{N} x_i E_i \tag{3b}$$

where $x_i =$ the proportion of the portfolio invested in security i. Optimization is conducted by differentiating the Lagrangian,

$$L = \sigma_P + \lambda_1 \left[1 - \sum_{i=1}^{N} x_i\right] + \lambda_2 \left[E_P - \sum_{i=1}^{N} x_i E_i\right] \tag{4}$$

This yields the first-order conditions with respect to the x_i and λ_i.

Although the portfolio weights (x_i) are constrained to add to 1, we have not stipulated that they be non-negative. If $x_i < 0$, security i has been sold short. If short sales are allowed, the optimization problem is simple, involving the solution of $n+2$ linear equations in $n+2$ unknowns. If the x_i must be non-negative, the problem involves quadratic programming, which is more difficult but easily handled on a computer.

Solving the optimization problem for different levels of E_P generates the minimum variance portfolio for each level of expected return. This group of portfolios is called the *minimum variance set*. Plotting the minimum variance set in (E_P, σ_P) space yields a mean-variance transformation curve. This curve is a hyperbola (see Fig. 1). The transformation curve and the shaded area to its right represent the *feasible set* of portfolios.

The solid line segment of the hyperbola in Figure 1 is called the *efficient frontier*. The frontier begins at the global minimum variance portfolio C and extends upward and to the right. Portfolios on the efficient frontier dominate all other feasible portfolios (including those on dotted segment of the hyperbola) because they provide the highest expected return for each level of risk. Consequently, mean-variance diversifiers prefer portfolios that lie on the frontier.

To this point, all assets have been assumed to be risky. Suppose that a riskless asset exists, e.g., a treasury bill.[7] Without loss of generality, the riskless asset is considered the $(N+1)$ st. Such an asset would have rate of return R_f, standard deviation of zero, and would be uncorrelated with all risky assets.

[7] The concept of a riskless asset in finance refers to an asset that will generate with certainty a known rate of return over a given period without risk of loss to the underlying capital. Riskless assets typically do not protect the investor against fluctuations in the price level (inflation).

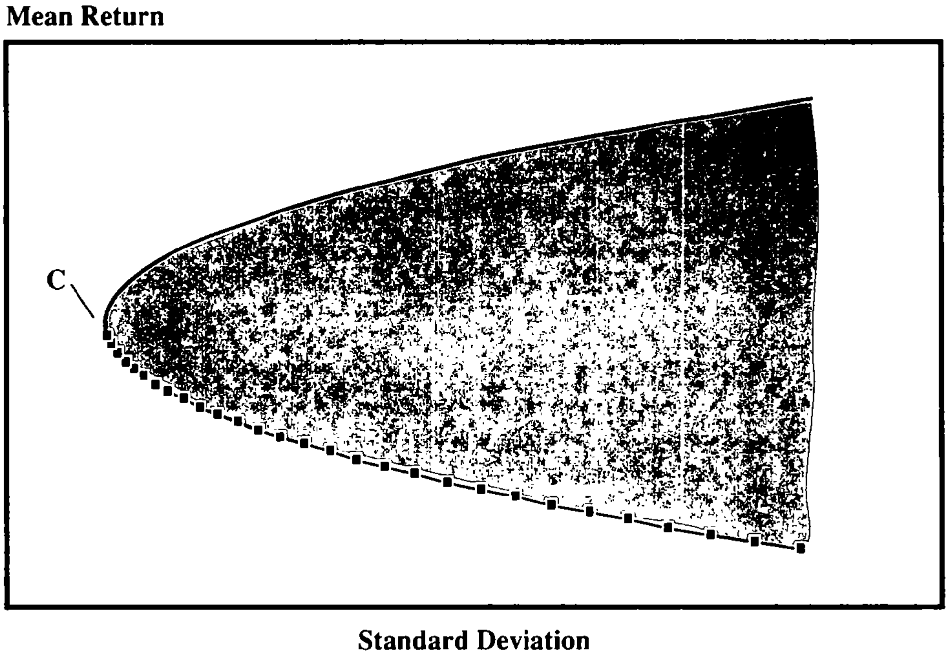

FIGURE 1. The efficient frontier.

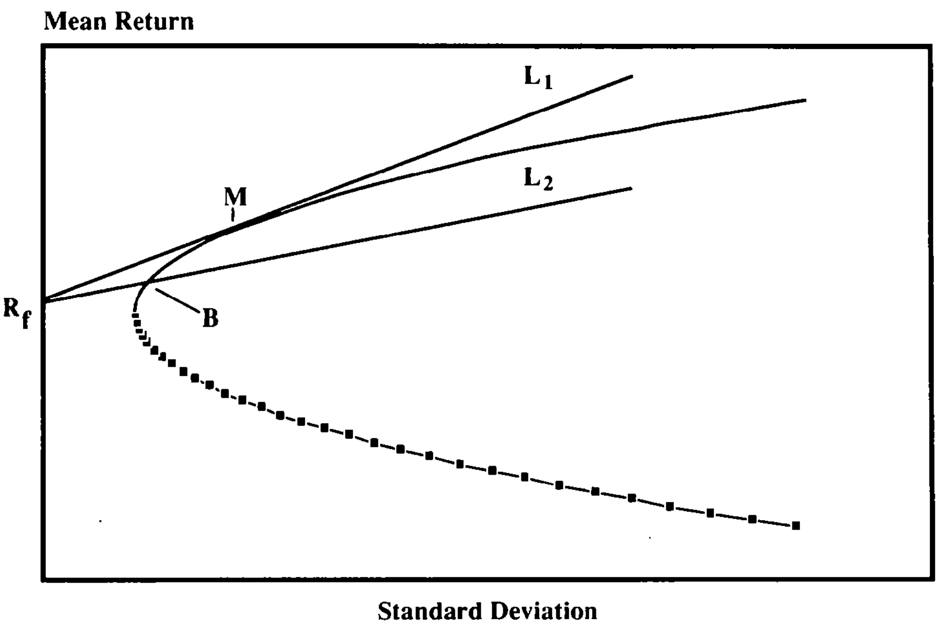

FIGURE 2. Two-fund separation.

The presence of the riskless asset leads to an imporant result called *two-fund separation*, which means that investment choices are reduced to the allocation of assets between two specific investment funds (portfolios).[8] To establish two-fund separation, consider the point R_f in Figure 2, representing the risk-return coordinates of the riskless security. Consider a portfolio (P) that is a linear combination of the riskless asset and any efficient portfolio of risky assets. Such a portfolio would have the following return:

$$R_P = \sum_{i=1}^{N} x_i R_i + \left(1 - \sum_{i=1}^{N} x_i\right) R_f \tag{5}$$

where $\sum_{i=1}^{N+1} x_i = 1 \Rightarrow x_{N+1} = 1 - \sum_{i=1}^{N} x_i$

Equation (5) can be rewritten as follows:

$$R_P = (1-\alpha) R_f + \alpha R_B \tag{6}$$

where $\alpha = \sum_{i=1}^{N} x_i$ and $R_B = \sum_{i=1}^{N} \frac{x_i}{\sum_{i=1}^{N} x_i} R_i = \sum_{i=1}^{N} w_i R_i$.

In (6), α represents the proportion of the portfolio invested in a specific efficient portfolio B, formed from the N risky assets, with rate of return R_B.

The mean and standard deviation of the portfolio P are:

$$E_P = (1-\alpha) R_f + \alpha E_B \tag{7a}$$

$$\sigma_P = \alpha \sigma_B \tag{7b}$$

Combining equations (7a) and (7b), we obtain:

$$E_P = R_f + \frac{\sigma_P}{\sigma_B} (E_B - R_f) \tag{8}$$

Thus, portfolio P can be represented in risk-return space by a straight line connecting the risk-return point of the riskless asset with the point representing the efficient risky portfolio. Such a line is represented by L_2 in Figure 2, consisting of linear combinations of the riskless asset and efficient portfolio B. Points between B and R_f represent long positions in the riskless asset (lending). Points to the right of B are achieved by taking short positions in the riskless asset (borrowing), i.e., $1-\alpha < 0$.

In choosing a risky portfolio to combine with the riskless asset, the investor will avoid dominated portfolios and choose the combination giving the highest return for each risk level. This is accomplished by choosing the efficient risky

[8] General separation theorems as well as conditions for separation when no riskless asset exists are discussed in INGERSOLL (1987).

portfolio representing the tangency between the efficient frontier and the straight line with intercept R_f. This is the line represented by R_fML_1 in Figure 2. Points along this line dominate other investment choices. Thus, investors will hold portfolios that are mixtures of the riskless asset and risky portfolio M, and investment choices are said to exhibit *two-fund separation* (the riskless asset is fund 1 and M is fund 2).

The position each investor chooses along the line R_fML_1 depends on his or her preferences for risk. Conservative investors will lend at the riskless rate, i.e., they will choose portfolios along line segment R_fM. More adventuresome investors will borrow at the riskless rate in order to buy more of the risky portfolio. They will hold portfolios along line segment ML_1. The importance of the two-fund separation property is that it reduces the investment decision to two separate steps: (1) Find the portfolio M and the line R_fML_1, and (2) select a point along the line depending upon one's risk preferences.[9]

To summarize, if investors are mean-variance decision makers, they will be interested in holding only mean-variance efficient portfolios. With no riskless asset, this implies choosing a portfolio along the efficient frontier. If a riskless asset exists and if investors can borrow and lend at the riskless rate, the efficient set becomes a straight line through $(R_f, 0)$ and tangent to the efficient frontier. Combined with some additional assumptions, these results have implications for the equilibrium expected rates of return on assets. This pricing relationship is explored in the following section.

3. THE CAPITAL ASSET PRICING MODEL (CAPM)

The capital asset pricing model (CAPM) was the first major asset pricing model, developed independently by WILLIAM SHARPE, JOHN LINTNER, and JAN MOSSIN during the 1960s. The CAPM has been extremely influential and has been subjected to exhaustive theoretical and empirical analysis, primarily by American researchers. It has been used extensively in financial decision making by institutional investors, corporate financial analysts, and others. It also has been controversial, with much of the controversy centering around the model's underlying assumptions and predictive ability.[10]

Since the 1960s, financial economics has continued to advance at a rapid rate. Numerous variants of the CAPM as well as more sophisticated asset pricing models have been developed. Noteworthy among the latter are the intertemporal capital asset pricing model (ICAPM) (MERTON 1973b)), option pricing theory (BLACK and SCHOLES (1973)), arbitrage pricing theory

[9] The selection of a point along R_fML_1 is usually illustrated as the tangency of the investor's indifference curve with this line.

[10] Considering both the quantity and quality of the research that has been done in the asset pricing field, it is not advisable for those new to the area to become bogged down in debates about the assumptions of the CAPM, even though the assumptions may seem to be tempting targets. The best approach is to read one of the excellent reviews of the CAPM literature (e.g., in LEVY and SARNAT (1984) or COPELAND and WESTON (1988)) and then to move on to the more recent literature.

(ROSS (1976, 1977)), and the consumption capital asset pricing model (BREEDON (1979)).

Although the CAPM clearly has limitations, it continues to be the most widely-used model in practical applications. As long as one is aware of the limitations and regards the CAPM as a useful source of information rather than absolute truth, it can serve as a valuable decision making tool. In addition, since the CAPM fundamentally changed the way people think about asset pricing, it is necessary to understand it in order to fully appreciate its successors.

The CAPM assumptions and the Capital Market Line (CML)

The CAPM takes mean-variance analysis one step further by seeking to determine the equilibrium asset pricing relationships that would hold in a world characterized by perfect capital markets and populated by MARKOWITZ diversifiers. Formally, the CAPM rests on the following assumptions: [11]

1. Investors are risk averse and select mean-variance (MARKOWITZ) diversified portfolios.
2. A riskless asset exists, and investors can borrow and lend unlimited amouts at the risk free rate.
3. Expectations regarding security means, variances, and covariances are homogeneous, i.e., investors agree about the parameters of the joint distribution of security returns.
4. Markets are frictionless, i.e., there are no transactions costs, taxes, or restrictions on short sales. Securities are assumed to be infinitely divisible.
5. All investors have the same one-period time horizon.
6. There is a "large number" of investors, none of whom can individually affect asset prices, and a large number of securities.

With the assumptions in hand, it is possible to draw some additional inferences from Figure 2. Specifically, since investors are assumed to be in agreement about the expected returns and the variance-covariance matrix of risky assets (homogeneous beliefs), all investors will hold portfolio M. Thus, portfolio M is the market portfolio, consisting of all risky assets. To see this, assume that asset j is not part of M. Since all investors hold M, this means that no one buys asset j. Thus, the price of asset j must adjust until it is attractive to investors. It then becomes part of the market portfolio.

The line $R_f M L_1$ is known as the *capital market line (CML)*. Portfolios along the CML are perfectly positively correlated. Since portfolio M is optimally diversified (efficient), it is said to contain no diversifiable (unsystematic) risk.

[11] Other assumptions could be added. E.g., the intitial CAPM was developed with U.S. markets in mind. For international application, one might add the assumption that transactions take place in a common currency or that exchange rate fluctuations can be costlessly and perfectly hedged. The CAPM has been shown to be quite robust with respect to the other assumptions and it might be to this assumption as well, as long as the degree of violation was not too severe.

The proportion of the ith security in the market portfolio is $x_i = V_i/\Sigma V_i$, where V_i is the total market value of security i. Because of the existence of the riskless asset and homogeneous beliefs about return distributions, separation occurs with the market portfolio (M) as the risky fund, and all investors' portfolios lie along the CML.

Derivation of the CAPM

The CAPM is derived by optimizing the following Lagrangian, which is a slight reformulation of equation (4):

$$L = \sigma_P + \lambda\left[E_P - \sum_{i=1}^{N} x_i E_i - \left(1 - \sum_{i=1}^{N} x_i\right) R_f\right] \tag{9}$$

With the introduction of the riskless asset, investors now allocate their portfolios among $N+1$ assets, the riskless asset and N risky assets. However, the sum of the portfolio proportions is still equal to 1: $\Sigma x_i + (1 - \Sigma x_i) = 1$.

The Lagrangian is optimized with respect to the x_i and λ. The first-order conditions with respect to the x_i are:

$$\frac{\partial L}{\partial x_i} = \frac{\sum_{j=1}^{N} x_j C_{ij}}{\sigma_P} - \lambda[E_i - R_f] = 0 \tag{10}$$

Multiply each term in (10) by x_i and sum over all risky assets (i) to obtain:

$$\sigma_P = \lambda\left[\sum_{i=1}^{N} x_i E_i + \left(1 - \sum_{i=1}^{N} x_i\right) R_f - R_f\right] = \lambda(E_P - R_f) \tag{11}$$

In (11), we have added and subtracted R_f. Since all investors hold the market portfolio, portfolio P is actually M. Therefore, it is apparent that $1/\lambda$ can be defined as the *market price of risk*:

$$\frac{1}{\lambda} = \frac{E_m - R_f}{\sigma_m} \tag{12}$$

where E_m, σ_m = the expected return and standard deviation of return on the market portfolio, M.

The CML (linear combinations of portfolio M with R_f) has the maximum possible slope of all feasible linear combinations of efficient portfolios and the riskless asset, i.e., the largest possible increment in expected return is received for each increment in risk.

To obtain the expected return on asset i, solve equation (10) for E_i, using (12) to eliminate λ:

$$(13) \qquad E_i = R_f + \left(\frac{E_m - R_f}{\sigma_m}\right) \frac{\sum_{j=1}^{N} x_j C_{ij}}{\sigma_m}$$

$$= R_f + (E_m - R_f) \frac{C_{im}}{\sigma_m^2} = R_f + \beta_i (E_m - R_f)$$

where $C_{im} = \text{Cov}(R_i, R_m)$ and $\beta_i = \text{Cov}(R_i, R_m)/\sigma_m^2$ = the *beta coefficient* or *beta* of asset *i*. The beta, the covariability of asset *i*'s return with the market portfolio, normalized by the variance of the market portfolio, is also recognizable as the regression coefficient of R_i on R_m.

Expression (13) is the relationship known as the capital asset pricing model. Under the assumptions of the CAPM, it is an equilibrium relationship between risk and return that must hold for all traded assets.

The CAPM is plotted in Figure 3. The CAPM relationship is also known as the *security market line (SML)*, which is related to but distinct from the CML. The CML includes only efficient portfolios. Inefficient portfolios and individual securities do not lie on the CML. However, in equilibrium individual securities and all portfolios lie on the SML.

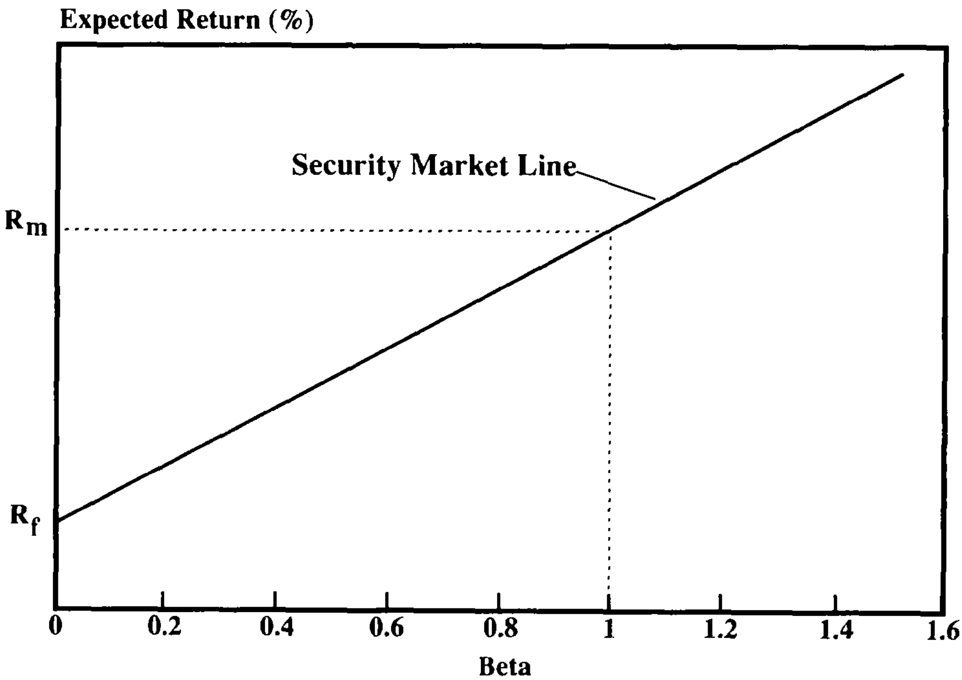

FIGURE 3. The capital asset pricing model.

Under the assumptions of the CAPM, there are several significant implications for investor behavior and corporate financial management. Among them are the following:

1. There is no market reward for diversifiable risk, i.e., the equilibrium return on a security reflects only the risk-free rate and its covariance with the market portfolio.
2. Firms should not be risk averse. Thus, under the CAPM, modeling firm decisions using utility functions will lead to incorrect results. Firms should not be concerned about diversifiable risk because investors can eliminate this type of risk by holding diversified portfolios. Rather, firms should maximize value. This is accomplished by choosing projects with positive net present value, where expected cash flows from the project are discounted at the rate appropriate for the project's systematic risk, i.e., risk adjusted discount rates defined by the CAPM.[12] Projects whose risk-return points lie above the SML will increase firm value, projects with risk-return coordinates on the SML deliver an expected return exactly appropriate for their risk, and projects below the SML will decrease firm value.

Evaluation of the CAPM

The assumptions underlying the CAPM are very strong, and nearly every one of them is violated in the real world. This of itself does not invalidate the CAPM. All economic and financial models are necessarily abstractions of reality. The question is whether the violations of the assumptions significantly affect the ability of the model to predict behavior in financial markets.

There are two primary ways to gauge the impact of assumptions: (1) Develop versions of the model where the key assumptions are relaxed, and (2) test the model against actual market data. For the most part, the CAPM has been shown to be resilient to the relaxation of assumptions. For example, a version of the CAPM (Black's zero beta CAPM, see LEVY and SARNAT (1984)) has been developed which does not require the existence of a riskless asset. The CAPM has been extended to incorporate nonmarketable assets (e.g., human capital) and non-normal return distributions. Relaxation of the homogeneity and tax assumptions also have been studied. The extensions complicate the analysis but do not change the conclusions.

Empirical tests of the CAPM have been more problematic. A careful reading of the literature suggests that the pure theoretical version of the CAPM does not agree well with reality. ROLL (1977) has argued that the CAPM may be untestable because of the impossibility of observing the true "market portfolio" consisting of all assets (not just common stocks).

[12] A risk adjusted discount rate is the discount rate that equates a project's expected future cash flows with its present market value. Riskless cash flows are discounted at the risk free rate. In a CAPM environment, cash flows with positive beta coefficients would be discounted at a rate higher than the risk free rate, i.e., positive-beta cash flows are worth less than riskless cash flows.

In spite of the apparent departures of the theoretical CAPM from observed security price behavior, the CAPM is widely used by financial professionals in the United States. Investment managers use the CAPM in portfolio selection, and corporate financial officers use it in capital budgeting (project selection). Typically, it is not used to the exclusion of other methods but as a valuable supplement to more traditional techniques.

Empirical considerations

Empirical applications of the CAPM focus on the so-called characteristic line, obtained for a particular security by regressing its holding-period returns on the returns for the "market" over some historical period. The New York Stock Exchange (NYSE) composite index, the Standard & Poor's 500 Stock Index, or some other broad-based index is used to represent the market. A five-year estimation period is common, using monthly or weekly returns in most cases.

The estimation equation for the characteristic line is:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{15}$$

where α_i, β_i = coefficients to be estimated,
R_{it}, R_{mt} = holding period returns on security i and the market index during period t, and
ε_{it} = a random error term.

Ordinary least squares estimation may be used, but more sophisticated methods are employed by some investment firms.[13]

The characteristic line can be used to partition risk into its systematic and unsystematic components. Systematic risk is risk common to all stocks, while unsystematic risk is idiosyncratic risk affecting a particular stock. Systematic and unsystematic are often used synonymously with undiversifiable and diversifiable risk. The risk partition is carried out by taking the variance of equation (15):

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{i\varepsilon}^2$$

$$1 = \rho_{im}^2 + \frac{\sigma_{i\varepsilon}^2}{\sigma_i^2}$$

where σ_i^2 = the variance of return of security i,
$\sigma_{i\varepsilon}^2$ = the variance of the random error term of security i, and
ρ_{im} = the correlation coefficient between security i and the market portfolio.

[13] These methods are designed to correct for the tendency of beta coefficients to regress toward the mean, i.e., the tendency for stocks with unusually high (low) betas in one period to have lower (higher) betas in the following period.

The estimated correlation coefficient between the returns of security i and the market partitions risk into the systematic part (ρ_{im}^2) and the unsystematic part ($1-\rho_{im}^2$). The unsystematic part serves as an index of diversifiability, i.e., the proportion of a stock's variance that can be diversified away. The average correlation coefficient with the market portfolio for NYSE stocks (ρ_{im}) is in the neighborhood of 0.5, meaning that 75 percent of total risk is potentially diversifiable. This information is clearly valuable to a portfolio manager, even one with reservations about the CAPM.

4. THE INTERTEMPORAL CAPM

As the limitations of the original CAPM became apparent, researchers developed alternative pricing models based on less stringent assumptions. One of the first models of this type was MERTON'S (1973b) intertemporal, continuous time CAPM (ICAPM). This model generalizes the CAPM in the following ways: (1) Instead of a single discrete-time planning horizon, the ICAPM postulates continuous time, intertemporal optimization by investors, who maximize the utility of lifetime consumption (and bequests). (2) In place of normal distributions, the ICAPM assumes that security returns can be described as geometric Brownian motion, so that prices are lognormal. (3) The most general form of the ICAPM incorporates an instantaneously riskless interest rate which can change over time, i.e., the investor knows with certainty the rate of interest $r(t)$ that can be earned over the next instant but future values of $r(t)$ are not known with certainty. This replaces the constant R_f of the CAPM.[14]

Although the full development of the ICAPM would be beyond the scope of this article, it is useful to summarize some of its essential features. In particular, consider the stochastic differential equation for the instantaneous return on the ith asset:

$$\frac{dS_i}{S_i} = \alpha_i dt + \sigma_i dz_i \tag{16}$$

where $S_i(t)$ = the price of asset i,
$\alpha_i(t)$ = the instantaneous expected rate of return on security i,
$\sigma_i(t)$ = the instantaneous standard deviation of return on security i,
$z_i(t)$ = standard Brownian motion (Wiener) process.[15]

[14] Lower case letters are used in this paper to represent parameters in continuous time models, while upper case letters are used for discrete time models.

[15] The intitial definition of a variable or parameter in the continuous time models indicates whether it is a constant or a function of time. Otherwise, time functionality will not be expressed explicitly. This follows the convention in the financial literature.

The equation implies that the instantaneous rate of return on an asset increases by a deterministic trend term ($\alpha_i dt$) plus a random term that can be modeled as Brownian motion. Increments of Brownian motion, $z(t+s)-z(s)$, are normally distributed with mean 0 and variance t (see KARLIN and TAYLOR (1981)). The differential $dz(t)$ is defined as:

$$dz(t) = \lim_{t\to 0} u\sqrt{\Delta t} \tag{17}$$

where u = a standard normal deviate. Equation (16) implies that asset prices are lognormally distributed, an assumption with substantial empirical support.

Sufficient statistics for the investment opportunity set at a given time point are: $\{\alpha_i, \sigma_i, \rho_{ij}\}$, where ρ_{ij} = the instantaneous correlation coefficient between dz_i and dz_j. The ρ_{ij} are constants. The instantaneous standard deviations, σ_i, and covariances, σ_{ij}, form the variance-covariance matrix Ω of risky security returns. The coefficients α_i *and* σ_i are functions of underlying state variables:

$$d\alpha_i = a_i dt + b_i dq_i \tag{18a}$$

$$d\sigma_i = f_i dt + g_i dx_i \tag{18b}$$

where $dq_i(t)$ and $dx_i(t)$ are standard Brownian motion processes, and a_i, b_i, f_i, and g_i are constants. This implies that the evolution of the market can be described by a set of Brownian motion processes, dz_i, dq_i, dx_i,

In the simplest case, the α_i, r, and Ω are constant. This leads to a continuous time version of equation (13) and two-fund separation. A more general case, involving three-fund separation, is obtained by assuming that a single state variable, the interest rate $r(t)$, describes the evolution of the investment opportunity set. Thus, $\alpha_i = \alpha_i(r)$ and $\sigma_i = \sigma_i(r)$. Furthermore, assume there exists an asset that is perfectly negatively correlated with r, e.g., a long-term bond. This asset is denoted asset n. Then it can be shown that the equilibrium rate of return on asset i ($i = 1, \ldots, n-1$) satisfies:[16]

$$\alpha_i - r = \frac{\sigma_i[\rho_{im} - \rho_{in}\rho_{nm}]}{\sigma_m(1-\rho_{nm}^2)}(\alpha_m - r) + \tag{19}$$

$$+ \frac{\sigma_i[\rho_{in} - \rho_{im}\rho_{nm}]}{\sigma_n(1-\rho_{nm}^2)}(\alpha_n - r)$$

where $\alpha_m(t)$ and $\alpha_n(t)$ are the instantaneous expected returns on assets m and n and the ρ_{jk} are the instantaneous correlation coefficients of the returns on assets j and k.

Thus, in equilibrium, investors are compensated for the covariability of assets with the market portfolio and for covariability between risky assets and the interest rate. Three-fund separation is present, with the three funds being

[16] For the derivation, see MERTON (1973b).

the market portfolio (M), a portfolio that is negatively correlated with the interest rate (n), and the default-risk-free asset. The coefficients of the risk premia, $(\alpha_m - r)$ and $(\alpha_n - r)$, in (19) are the coefficients in the multiple regression of asset i's returns on those of portfolios M and n.

Both the one and two-factor versions of the ICAPM represent significant generalizations of the original CAPM. The two-factor version may be useful for pricing assets such as insurance policies that are sensitive to interest rate risk.

5. ARBITRAGE PRICING THEORY (APT)

The concept of arbitrage is very important on financial theory and practice. STEPHEN ROSS (1976, 1977) has developed a formal theory of asset pricing based on arbitrage arguments. Before considering Ross' theory, a brief general explanation of arbitrage is provided.

Consider a competitive, frictionless, and riskless capital market. Asset trading takes place at time 0, and the assets provide cash flows, v_i, at time 1. Assume that there exists a price function, $p(v)$, mapping units of time 1 cash flow into units of dollars at time zero. The following would constitute an arbitrage opportunity:

(20a) *Portfolio:* $$\sum_{i=1}^{N} x_i p(v_i) = 0$$

(20b) *Time 1 Cash Flow:* $$\sum_{i=1}^{N} x_i v_i > 0$$

where x_i = the proportion of the portfolio invested in asset i (at time 0).

Equation (20a) describes an *arbitrage portfolio*, i.e., a portfolio requiring zero net investment. Such a portfolio could be formed in any number of ways, such as borrowing and then lending the proceeds or selling one security short and buying another, etc. Equation (20b) indicates that the net cash flow from the arbitrage portfolio is positive. Thus, the portfolio constitutes a "money machine"; a net investment of zero at time 0 yields a positive cash flow at time 1.

An arbitrage opportunity also would arise if a package of assets trades for a price different from the sum of the prices of the individual assets inculed in the package. E.g., consider a package consisting of N_1 shares of asset 1 and N_2 shares of asset 2. The following would be an arbitrage opportunity:

(21) $$p[N_1 v_1 + N_2 v_2] > N_1 p(v_2) + N_2 p(v_2)$$

To exploit this opportunity, the investor would sell the package short and buy the individual assets. This would create a positive cash flow at time 0 and a cash flow of zero at time 1. Analogues to these examples also can be formulated in markets with risky securities.

Arbitrage is not a purely theoretical construct. In fact, it forms the basis for program trading and other sophisticated financial practices in the U.S. capital markets. The basic idea is that money machines cannot exist for long in competitive markets. Prices will adjust as investors shift funds to take advantage of arbitrage opportunities.

Ross's arbitrage pricing theory (APT) uses arbitrage concepts to develop an asset pricing model based on much weaker assumptions than the CAPM. Rigorous statements of the assumptions and derivation are provided in Ross (1976, 1977), Ingersoll (1987), Jarrow (1988), and other sources. Among the more important assumptions are the following:

1. Capital markets are frictionless and competitive.
2. Means and variances of asset returns are finite.
3. All economic agents hold the same beliefs regarding expected returns.
4. There are no restrictions on short sales.
5. All agents believe that returns are generated by the following linear factor model:

$$R_i = E_i + \sum_{j=1}^{K} \beta_{ij} f_j + \varepsilon_i \tag{22}$$

where f_j = the random return on factor j,

ε_i = a random error term (residual) associated with asset i.

The f_j are called *factors* and the β_{ij} are *factor loadings*. It is assumed that $E(f_j) = 0$; $E(\varepsilon_i) = 0$; $E(\varepsilon_i \varepsilon_j) = 0$, $i \neq j$; $E(\varepsilon_i f_j) = 0$; and that $E(\varepsilon_i^2) = \omega_i^2 < W^2$, where W is a finite constant. Thus, the residuals are uncorrelated and bounded.

It is essential to the development of the model that there exist a finite number (K) of factors that generate asset returns. The factors can be interpreted as variables describing the state of the economy. For example, one factor might be related to real industrial production and another to unanticipated inflation. One of the factors might represent the return on the market portfolio, but the market portfolio plays no special role in the APT.

For ease of exposition, it is assumed in the following that the linear factor model consists of only one factor, denoted f. Assume there are N assets in the market. Then, the derivation of APT can be sketched as follows (Ross (1976)):

1. Form a well-diversified arbitrage portfolio consisting of all N assets, such that

$$\sum_{i=1}^{N} x_i = x'e = 0 \tag{23}$$

where x = the vector of portfolio weights, $(x_1, \ldots x_N)$, and e = an N-dimensional vector of 1's. Each element x_i is assumed to be of order $1/N$.

2. The return on the arbitrage portfolio will be:

$$\sum_{i=1}^{N} x_i R_i = \sum_{i=1}^{N} x_i(E_i+\beta_i f) + \sum_{i=1}^{N} x_i \varepsilon_i \tag{24}$$

By the law of large numbers, the last term in (24) will be approximately zero, i.e., diversification eliminates (in some sense) non-systematic risk.[17] Assume for the moment that diversification totally eliminates non-systematic risk. This assumption is relaxed later.

3. Suppose that the arbitrage portfolio also is chosen to eliminate all systematic risk, so that:

$$\sum_{i=1}^{N} x_i \beta_i = x' B = 0 \tag{25}$$

where B = the vector of betas on the N assets in the portfolio. Short selling is typically required to generate this zero-beta portfolio. The return on such a portfolio is $\Sigma\, x_i R_i = \Sigma\, x_i E_i = x' E$, where E = the vector of expected asset returns, $(E_1, \ldots, E_N)$.

4. Since all risk has been eliminated from the portfolio, to avoid arbitrage it also must be true that:

$$\sum_{i=1}^{N} x_i E_i = x' E = 0 \tag{26}$$

If this relationship did not hold, demand for assets by arbitrage traders would reduce returns until (26) were satisfied.

Thus, any portfolio satisfying (23) and (25) must also satisfy (26). This is an algebraic statement that any vector orthogonal to e and B is also orthogonal to E. This implies that the expected return vector E can be written as a linear combination of e and B, i.e., there exist constants ρ and λ such that:

$$E_i = \rho + \lambda \beta_i \tag{27}$$

Equation (27) must hold in order to avoid arbitrage, i.e., it holds for any portfolio, not just the arbitrage portfolio (23). Thus, the coefficients in (27) can be determined by considering portfolios with different structures. For example, to obtain the intercept ρ, consider a portfolio that satisfies $x' B = 0$ and $\Sigma_i x_i = 1$. Multiply both sides of (27) by x_i and sum over all i. It is clear that ρ represents the return on a zero-beta portfolio. If a riskless asset exists, ρ is its return.

To determine λ, consider a portfolio with $\Sigma\, x_i = 1$ and $x' B = 1$.

[17] This result does not require a distributional assumption about ε_i beyond the existence and boundedness of its variance.

Multiplying (27) by x_i and summing yields the following relationship:

$$E_P = \rho + \lambda \tag{28}$$

Thus, $\lambda = E_P - \rho$. But a portfolio with beta of 1 must have the same return as the return-generating factor f. Thus, $\lambda = E(f) - \rho$, and $E_i = \rho + \beta_i[E(f) - \rho]$; so that the single factor arbitrage pricing model has the same form as the CAPM.

Generalizing to K factors, we have Ross' arbitrage pricing model:

$$E_i = \rho + \sum_{j=1}^{K} \beta_{ij}[E(f_j) - \rho] = \rho + B_i' \Lambda \tag{29}$$

where $\Lambda' = [E[f_1) - \rho, \ldots, E(f_K) - \rho]$

$B_i' = (\beta_{i1}, \ldots, \beta_{iK})$

In actuality, the APT relationship (29) is not exact and is more appropriately written as: $E_i \approx \rho + B_i' \Lambda$. The arbitrage pricing theorem is based on the following limiting result (INGERSOLL (1987, p. 172) or JARROW (1988, p. 120)):

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (E_i - \rho - B_i' \Lambda)^2 = 0 \tag{30}$$

The theorem assumes the existence of a K factor linear return generating model with bounded residual risk and no arbitrage opportunities. The linear pricing model prices "most" of the assets correctly, i.e., the model applies on the average and not uniformly.

The APT is less restrictive in its assumptions than the CAPM. Its major hypotheses are that asset returns are generated by the linear factor model and that assets are priced to eliminate arbitrage opportunities. Thus, it is not necessary to assume that markets are in equilibrium. The primary assumptions are that markets are competitive and frictionless and that investors hold diversified portfolios, prefer more to less (a very weak preference relationship), and agree on expected returns and the return generating model. Investors do not necessarily agree on the probability distributions of returns so normality and mean-variance decision making are not required. Furthermore, the arbitrage pricing relationship is multidimensional rather than unidimensional in risk. The primary drawback of the APT is the difficulty of estimating factor expected returns and response coefficients. However, there is a growing literature on estimation (e.g., ROLL and ROSS (1980)).

6. OPTION PRICING THEORY

An important advance in financial modelling was the development of the BLACK-SCHOLES (1973) option pricing model (OPM). Options are derivative securities, giving their owners rights to buy and sell primary securities such as stocks and bonds. The applicability of the OPM extends far beyond the pricing of options. Many types of derivative assets, broadly classified as contingent claims, can be priced using the principles of the OPM. Among these are corporate liabilities and insurance policies.

The original version of the OPM was developed as a continuous time model using stochastic calculus. Subsequently, researchers developed an option pricing model based on a simple binomial process (see COX and RUBENSTEIN (1985)). The continuous time OPM is the limiting version of the binomial model. In this paper, the continuous time version is used. However, readers should be aware of the existence of the binomial model because it is applicable in some situations where the BLACK-SCHOLES model cannot be used.

The primary types of options are the so-called European and American calls and puts. European options can be exercised only at a fixed date in the future (the exercise date), while American options can be exercised at any time between the option's inception date and exercise date. For ease of presentation, this discussion focuses on European options.

A European call on common stock gives the holder the right to purchase the stock for a specified price (the *exercise price*) at a specified date, the *exercise date*. The value of the call can be expressed as $C(S, \tau) = C(S, \tau; K, \sigma, r)$, where S = the present stock price, K = the exercise price, r = the instantaneous risk free rate, σ = the risk parameter, and τ = the time until the exercise date. The parameter $\tau = T-t$, where T = the expiration date and t = the present time. The variables S and τ are state variables that change over time, while K, T, r, and σ are fixed parameters. The call option is defined by the following boundary conditions: $C(0, \tau; K, \sigma, r) = 0$ and $C(S, 0; K, \sigma, r) = \text{Max}\,(S-K, 0)$. Thus, at expiration, the call value is equal to $S-K$ if the stock price exceeds the exercise price and 0 otherwise.

It is interesting to examine the reasoning that leads to the condition $C(0, \tau; K, \sigma, r) = 0$. An important assumption underlying the theory of rational option pricing (e.g., MERTON (1973)) is that options are priced so that they are neither dominant nor dominated securities. Security A dominates security B if A's return is at least as large as B's in all states of the world and larger than B's in at least one state of the world. This implies that an American call is worth at least as much as an otherwise identical European call since the American call has less restrictive exercise privileges. Likewise, if calls A and B are identical in every way except in exercise price, then $K_A < K_B$ implies $C(S, \tau; K_A, \sigma, r) \geq C(S, \tau; K_B, \sigma, r)$; and, if calls A and B are identical except for time to maturity, $T_A > T_B$ implies $C(S, T_A-t; K, \sigma, r) \geq C(S, T_B-t; K, \sigma, r)$.

These results permit one to derive the condition $C(0, \tau; K, \sigma, r) = 0$. Because a stock is in effect a perpetual Amerian call with an exercise price of 0, stock A is worth at least as much as a (American or European) call on A with a non-zero exercise price and a finite time to maturity. Thus, if the value of the stock is 0, the value of the option must be zero.[18]

[18] Negative values for the option are ruled out by the terms of the option contract.

The value of the call depends upon the underlying stock price. BLACK and SCHOLES modeled the stock price as a diffusion process:

$$\frac{dS_i}{S_i} = \alpha_i dt + \sigma_i dz_i \tag{31}$$

where α_i = the instantaneous expected rate of return on security i,

σ_i = the instantaneous standard deviation of return on security i,

$z_i(t)$ = standard Brownian motion (Wiener) process for security i, and

$S_i(t)$ = the price of security i.

In contrast to the ICAPM, the BLACK-SCHOLES model treats α_i and σ_i as constants. Suppressing the parameters and the subscript i, the value of a general option on a stock with price S can be written as $F(S, t)$. The option pricing derivation begins with the formation of a portfolio (often called the *hedge portfolio*) consisting of the stock and the option:

$$P = F + nS \tag{32}$$

where n = the number of units of the stock included in the portfolio (n not necessarily an integer).

To study the dynamics of the portfolio, it would be useful to obtain the derivative of (32). However, this cannot be done using conventional calculus because Brownian motion is not differentiable. What is needed is a method for differentiating functions of Brownian motion. Such a method is provided by a branch of mathematics known as stochastic calculus. In particular, a result known as Ito's lemma has found extensive application in finance.

A heuristic derivation of Ito's lemma which leads into the derivation of the option pricing formula is provided here (a more formal derivation is provided in KARLIN and TAYLOR (1981)). First expand $F(S, t)$ using the Taylor expansion:

$$dF = F_S dS + F_t dt + \frac{1}{2} F_{SS}(dS)^2 + F_{St} dS\, dt + \frac{1}{2} F_{tt}(dt)^2 + \ldots \tag{33}$$

where subscripts indicate partial derivatives. The key to the analysis is the order relationships: $[dz_i]^2 \approx dt$ and $(dt)^k \approx 0$, $k > 1$. Using these relationships and substituting (31) into (33) yields:

$$dF = \left(F_S \alpha S + F_t + \frac{1}{2}\sigma^2 S^2 F_{SS} \right) dt + F_S \sigma S dz \tag{34}$$

Expression (34) is called the Ito transformation formula.

Now take the derivative of (32): $dP = dF + ndS$ and substitute for dF from (34) and dS from (31) to obtain:

(35) $$dP = \left[(n+F_S)\alpha S + F_t + \frac{1}{2}\sigma^2 S^2 F_{SS}\right] dt + (n+F_S)\,\sigma S dz$$

This stochastic differential equation can be reduced to a non-stochastic differential equation by choosing the number of shares n in the portfolio to equal $-F_S$. This choice eliminates the risk term from (35), creating a *riskless hedge.*[19] Since the resulting portfolio has no risk, it must earn only the risk-free rate in order to avoid arbitrage. So $dP = r[F+nS]dt = r[F-F_S S]dt$, where r = the instantaneous risk-free rate (assumed constant) and the relationship $n = -F_S$ has been used again. Thus, (35) becomes:

(36) $$0 = rF - rF_S S + F_\tau - \frac{1}{2}\sigma^2 S^2 F_{SS}$$

The time to expiration of the option, $\tau = T-t$ has been substituted for t in (36) accounting for the change in the sign of the time derivative term.

Equation (36) is the differential equation for BLACK-SCHOLES option pricing. It is solved for F, subject to appropriate boundary conditions, to obtain the option pricing formula.

In deriving (36) nothing has been assumed about the nature of the option being priced. In fact, (36) is quite a general result and applies to various types of options. Specific cases are generated by the appropriate formulation of boundary conditions. For a European call option, the boundary condition $F(S, 0) = \text{Max}\,(S-K, 0)$ yields the BLACK-SCHOLES call option formula:

(37) $$F(S, \tau) = SN(d_1) - Ke^{-r\tau}N(d_2)$$

where $$d_1 = \frac{\ln(S/K) + \left(r + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}$$

$$d_2 = d_1 - \sigma\sqrt{\tau}$$

Equation (37) is obtained by forming a hedge portfolio that eliminates all risk. This can be done because both the stock and the option are determined by the same stochastic process, equation (31). The concept of hedging is extremely important and has been widely used in financial markets in the U.S. It should be noted that the formation of a riskless hedge in developing the option price formula does not mean that the option is riskless. In fact, the option is more risky than the underlying common stock.

Equation (37) is remarkable in a number of respects. One of its most important properties is that the option can be priced on the basis of only a few parameters: r, σ, τ, and K.[20] The option price does not depend on the rate of return on the stock (α_i), the beta of the stock, or the market risk premium.

[19] The risk term is the term involving dz.

[20] The intial stock price $S(O)$ is also required.

Thus, it avoids many of the estimation problems that have plagued applications of the CAPM.

An important alternative method for deriving options formulae has been developed by Cox and Ross (1976). This is called the risk-neutral valuation methodology. The risk-neutral approach is based on the following argument: Since the hedge portfolio has no risk, it should have the same value in an economy where investors are risk averse as in an economy where investors are risk-neutral. Thus, valuation of the option can take place in the environment where the calculations are easiest, i.e., in the risk-neutral environment. This led Cox and Ross to propose the following valuation formula, which gives the BLACK-SCHOLES result:

$$C(S, \tau) = E[C_T]\, e^{-r\tau} = E[\mathrm{Max}\,(0, S_T - K)]\, e^{-r\tau} \tag{38}$$

where t = the present time, T = the option expiration date, $\tau = T - t$, $C(S, \tau)$ = the option price at time t, and S_T = the random variable, stock price at time T.

Since the stock has been assumed to follow geometric Brownian motion, the expected value in (38) can be evaluated with respect to a lognormal distribution. Because of the risk neutrality argument, the appropriate lognormal is one with location parameter $= (r - \sigma^2/2)\tau$ and risk parameter $= \sigma^2 \tau$. Simplifying the following integral produces the BLACK-SCHOLES formula:

$$C(S, \tau) = e^{-r\tau} \int_K^{\infty} (S_T - K) \times \frac{1}{S_T \sigma \sqrt{\tau}\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln(S_T/S) - (r - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right)^2} dS_T \tag{39}$$

where $C(S, \tau) = C(S, \tau; K, \sigma, r)$. This method does not imply that either the option or the stock is riskless; it merely suggests that if a perfect hedge is possible, the option price can be calculated as if the mean rate of increase in the stock price is the risk free rate.

The BLACK-SCHOLES formula and its progeny have been widely used in U.S. financial markets. It forms the foundation for much of the options trading on the Chicago Board Options Exchange (CBOE), the world's second largest securities exchange.

II. INSURANCE PRICING MODELS

7. FINANCIAL MODELS

Since insurance companies are corporations and insurance policies can be interpreted as a specific type of financial instrument or contingent claim, it seems natural to apply financial models to insurance pricing. Financial theory views the insurance firm as a levered corporation with debt and equity capital.[21] The insurer raises debt capital by issuing insurance contracts.

Insurance contracts are roughly analogous to the bonds issued by non-financial corporations. Bonds tend to have fixed coupon payments and a fixed maturity date. Insurance policies are more risky than conventional bonds because both the payment time and amount are stochastic. In addition, the payout period for most policies does not have a fixed time limit. Thus, insurance pricing poses some difficult problems that are not present when dealing with conventional financial instruments.

Financial theory views the insurance underwriting and pricing decision as corporate capital budgeting. In capital budgeting (also called capital project appraisal), firms accept or reject projects based on decision criteria such as the net present value (NPV) or internal rate of return (IRR) rules (see BREALEY and MYERS (1988) or COPELAND and WESTON (1988)). These decision rules focus on the amount and timing of the anticipated cash flows from the candidate projects. The rate of return targets are based on market risk-return relationships such as the CAPM or APT. Thus, policy prices reflect equilibrium relationships between risk and return or, minimally, avoid the creation of arbitrage opportunities.

Much of the impetus for the development of insurance financial models has been provided by rate regulation in the United States. During the late 1960s and early 1970s, regulators in several states (e.g., New Jersey and Texas) began to require that rates in regulated lines, particularly private passenger automobile insurance, reflect the investment income that insurers earn on policyholder funds between the premium payment and loss payment dates. Accounting models were developed to measure the investment income attributable to a given policy block, and such models are still used in many jurisdictions.[22] Later regulatory attention has focused on financial models, providing a market value rather than a book value measure of the fair rate of return to equity capital. The first use of financial models in regulation occurred in Massachusetts in 1976.

The value of insurance financial models extends beyond the regulatory arena. The models are designed to estimate the insurance prices that would pertain in a competitive market. Charging a price at least as high as the competitive price (reservation price) increases the market value of the company. Charging a lower price would reduce the company's market value. Thus, financial models and financial prices are among the key items of information that insurers should have at their disposal when making financial decisions about tariff schedules, reinsurance contract terms, etc. This information should serve as a complement to and not a substitute for actuarial information. An objective of research in this area is to develop a unified theory of insurance pricing that combines elements of actuarial and financial theory.

[21] A levered (or leveraged) corporation is one that finances its operations in part through debt capital, i.e., borrowed funds. The concept of leverage is discussed in more detail below.

[22] For a discussion of accounting models see CUMMINS and CHANG (1983).

This section discusses the principal financial models that have been proposed for pricing insurance contracts. The earliest models, based on the CAPM, provide important insights but are too simple to be used in most realistic situations. More promising are discrete and continuous time discounted cash flow (DCF) models. Both CAPM and DCF models have been used in insurance price regulation in the United States. The most recent work on insurance pricing has focused on option pricing, arbitrage pricing theory, and more advanced models.

8. A SIMPLE CAPM FOR INSURANCE PRICING

The insurance CAPM was developed in COOPER (1974), BIGER and KAHANE (1978), KAHANE (1979), FAIRLEY (1979), and HILL (1979). This model gives prices that would obtain in a world characterized by perfect capital markets and competitive insurance markets. The derivation begins with a simple model of the insurance firm:

$$Y = I + \Pi_U = R_A A + R_U P \tag{40}$$

where Y = net income,

I = investment income (net of expenses),

Π_U = underwriting profit (premium income less expenses and losses),

A = assets,

P = premiums,

R_A = rate of investment return on assets, and

R_U = rate of return on underwriting (as proportion of premiums).

Equation (40) can be expressed as a return on equity as follows:

$$R_W = \frac{Y}{W} = R_A \frac{A}{W} + R_u \frac{P}{W} \tag{41}$$

where W = equity (policyholders' surplus), and R_W = the rate of return on equity. Recognizing that assets are the sum of liabilities and surplus, one obtains:

$$R_W = R_A \left(\frac{L}{W} + 1 \right) + R_U \frac{P}{W} = R_A (ks + 1) + sR_U \tag{42}$$

where $s = P/W$ = the premiums-to-surplus ratio, and

$k = L/P$ = the liabilities-to-premiums ratio (funds generating factor).

Equation (42) indicates that the rate of return on equity for an insurer is generated by leveraging the rates of investment return and underwriting return.[23] The leverage factor for investment income is $(ks+1)$, a function of the premiums-to-surplus ratio and the funds generating factor. The latter approximates the average time between the policy issue and claims payment dates. The underwriting return is leveraged by the premiums-to-surplus ratio.

Equation (42) can be written in an interesting way as follows:

$$R_W = R_A + s(R_A k + R_U) \tag{43}$$

The insurer has the option of not writing insurance (choosing $s = 0$). In this case, it will be an investment company, investing its equity at rate R_A. If $s > 0$, (43) shows that writing insurance at a negative underwriting profit will increase return on equity as long as $-R_U < kR_A$.

Equation (42) is essentially an accounting model and has little economic content. The model is given economic content by assuming that the insurer's equity and assets are priced according to an equilibrium pricing model, in this case the CAPM. According to the CAPM, the equilibrium rate or return on any asset is given by equation (13), where the risk premium for any cash flow stream i is measured by its beta coefficient: $\beta_i = \mathrm{Cov}(R_i, R_m)/\mathrm{Var}(R_m)$.

Since the covariance is a linear operator, the insurer's equity beta can be obtained from equation (42) as:

$$\beta_W = \beta_A(ks+1) + s\beta_U \tag{44}$$

where β_E, β_A, and β_U = the betas of the insurer's equity, assets and underwriting return, respectively.

The equilibrium rate of return on the insurer's equity is:

$$E_W = R_f + \beta_W(E_m - R_f) = R_f + [\beta_A(ks+1) + s\beta_U](E_m - R_f) \tag{45}$$

where $E_W = E(R_W)$. The equilibrium underwriting profit is obtained by equating (45) and the expected value of (42) and solving for $E_U = E(R_U)$. (The CAPM relationship is also substituted for $E(R_A)$ in (42)). The result is:

$$E_U = -kR_f + \beta_U(E_m - R_f) \tag{46}$$

Equation (46) is often called the *insurance CAPM*.

The first term of (46), $-kR_f$, represents an interest credit for the use of policyholder funds. The second component of E_U is the insurer's reward for risk-bearing: the underwriting beta multiplied by the market risk premium. If underwriting profits are positively correlated with the market, the insurer will earn a positive risk loading as compensation for bearing systematic risk. There

[23] Leverage refers to the capital structure of the firm, i.e., its mix of debt and equity. A firm with a large amount of debt relative to equity is said to be highly leveraged. Leverage can be used to increase return on equity (ROE) because debt charges are fixed, i.e., they do not vary with the fortunes of the firm. If the firm uses the funds raised by borrowing to buy productive assets that earn more than the debt charges, the effect of the earnings in excess of the debt charges is multiplied or leveraged (e.g., by the debt to equity ratio) in its effect on ROE. Leverage also increases the firm's financial risk. In the U.K., leverage is sometimes called *gearing*. Of course, insurance leverage is a generalization of the usual concept because the underwriting profit or loss is not known in advance.

is no reward for unsystematic risk because this type of risk can be eliminated by investors through diversification. Thus, the risk of ruin is not priced and policies are treated as free of default risk.

The insurance CAPM yields important insights into the operation of insurance markets by introducing the concept of an equilibrium price for insurance and distinguishing between different types of risk. However, it is too simplistic for use in real-world applications.

Several limitations of the insurance CAPM have motivated researchers to seek more realistic insurance pricing models. One serious problem is the use of the funds generating or k factor to represent the payout tail. Discounted cash flow methods should be used to value cash flows that occur at different periods of time, but the k factor represents only a crude approximation of the DCF approach. A second problem is the assumption of no bankruptcy. It has long been recognized in actuarial science that ruin probabilities are important in insurance markets. Thus, insurance is more appropriately priced using a model that recognizes default risk. A third problem is the failure of the model to account for interest rate risk. As a practical matter, the use of underwriting betas can lead to inaccuracies because estimation errors can be quite significant (see CUMMINS and HARRINGTON (1985)). Most of the models discussed below represent an attempt to deal with one or more of these problems.

9. AN EQUILIBRIUM MODEL WITH AN INSURANCE SECTOR

The insurance CAPM prices insurance like any other financial asset; insurance risk plays no special role in this model. However, insurance risk may have characteristics which warrant separate treatment. Among other things, insurance deals with pure risk (involving a loss or no loss) rather than the speculative risk dealt with by most other financial assets.

A market equilibrium model including an insurance sector has been developed by ANDREW TURNER (1987). (A similar model is presented in ANG and LAI (1987)). Turner postulates an economy consisting of insurance companies, households, and non-insurance (productive) corporations. Households are endowed with initial wealth at time zero and maximize utility over a one-period planning horizon. Household decision variables include: (1) time zero consumption, (2) shares of stock in insurance companies and non-insurance companies, (3) real assets such as housing, and (4) insurance policies covering the real assets.

Real assets are assumed to be subject to two types of pure risk: individual risk, defined as loss events uncorrelated across real assets, and social risk, loss events that are correlated across exposure units. An individual loss might be a fire affecting only one house, while a social loss might arise from an earthquake affecting all houses in a region.

The assumptions of the TURNER model are similar to those of the CAPM, e.g., homogeneous beliefs about asset returns and variances, competitive markets, etc. He obtains the following equilibrium underwriting profit formula:

(47) $$E_{U_Z} = -R_f - (\beta_{L_Z V_m} + \beta_{L_Z V_N} - \beta_{L_Z L_N} - \beta_{X_Z X_N})(E_m - R_f)$$

where R_{U_Z} = underwriting return on insurance covering real asset Z,

V_m = value of aggregate market portfolio of traded financial assets,

V_N = value of aggregate market portfolio of real assets,

L_Z = value of aggregate social losses to real assets of type z,

L_N = value of aggregate social losses to all real assets,

X_Z = value of aggregate individual losses to real assets of type z, and

X_N = value of aggregate individual losses to all real assets.

The betas are all defined analogously to the following:

(48) $$\beta_{L_k V_m} = \mathrm{Cov}\left(\frac{L_k}{P}, \frac{V_{m1}}{V_{m0}}\right) \Big/ \mathrm{Var}\,(R_m)$$

where V_{mt} = value of market portfolio of financial assets at time t, and

P = the premium.

The betas in this model are loss rather than profit betas. Thus, their signs are opposite to the sign of the underwriting profit beta, β_U, discussed above.

The TURNER model includes a risk loading for the covariability of the social losses of asset k with the market portfolio of traded assets, the aggregate of all real assets, and aggregate social losses. In addition, there is a risk loading for the covariability of the individual losses of asset k with aggregate individual losses. The first three terms are likely to give rise to significant risk loadings, whereas the individual loss beta should be near zero.

The TURNER model defines systematic risk more broadly than the CAPM to include covariability among losses. For example, if aggregate automobile and workers' compensation losses were correlated, the TURNER model would provide a risk loading even if these losses were uncorrelated with the market portfolio of traded securities. This is an important generalization of the CAPM, which could be extended to an intertemporal, continuous time setting.

10. DISCRETE TIME DISCOUNTED CASH FLOW (DCF) MODELS

The financial models that have been most widely used in practice in the United States are discrete time discounted cash flow models. These models are based on concepts of corporate capital budgeting (see BREALEY and MYERS (1988) or COPELAND and WESTON (1988)). The two most prominent models, the MYERS-COHN model and the National Council on Compensation Insurance (NCCI) model, are analyzed in CUMMINS (1990). The following discussion focuses on MYERS-COHN.

The MYERS-COHN model is an application of MYERS' adjusted present value (APV) method (see BREALEY and MYERS (1988, pp. 443-446)). The steps in applying the APV method to analyze a project are the following: (1) Estimate the amount and timing of the cash flows expected to result from the project, (2) estimate the risk-adjusted discount rate for each flow, (3) compute the present value of the cash flows using the risk-adjusted discount rates, and (4) if the present value is greater than zero, accept the project.

Discounting each flow at its own risk-adjusted discount rate is consistent with the principle of value-additivity. The policy is priced as if the various flows could be unbundled and sold separately, avoiding the creation of arbitrage opportunities.

In insurance DCF analysis, it is important to adopt a perspective in order to avoid double counting. Flows can be measured either from the perspective of the insurer or from the perspective of the policyholder. Flows from one are the mirror image of flows from the other. The MYERS-COHN model adopts the policyholder perspective. The relevant cash flows are premiums, losses, expenses, and taxes. It would be double counting to consider profits as a flow when using the policyholder perspective. The flows from the company perspective are: surplus commitment, underwriting profits (net of taxes), and investment income, also net of taxes.

In addition to losses and expenses, the policyhoders pay the taxes arising out of the insuance transaction. The reasoning is as follows: When writing a policy the company commits equity capital to the insurance business. The owners of the company always have the alternative of not writing insurance and investing their capital directly in financial assets (stocks and bonds). They will not enter into the insurance transaction if by doing so they subject income on their capital to another layer of taxation. Thus, the policholders must pay the tax to provide a fair after-tax return.

The objective of the MYERS-COHN model is to determine the fair premium for insurance. The premium is defined as fair if the insurance company is

TABLE 1

CASH FLOWS IN TWO-PERIOD MYERS-COHN MODEL

Flow	Time:	0	1	Discount Rate
Premium		P	0	r_f
Loss		0	L	r_L
Underwriting Profits tax		0	$\tau(P-L)$	r_f, r_L
Investment balance (IB)		$P(1+\delta)$	0	
IB Tax		0	$\tau r_f P(1+\delta)$	r_f

Key: P = premiums, L = expected losses, r_f = risk-free rate of interest, r_L = risk-adjusted discount rate for losses, τ = corporate income tax rate, δ = surplus-to-premiums ratio.

exactly indifferent between selling the policy and not selling it. The insurer will be indifferent if the market value of the insurer's equity is not changed by writing the policy. Although this argument may seem to at odds with the profit motive, it is consistent with profit maximization in a competitive market. Under competition, each product sells for the price that will exactly pay for the factors of production. Thus, the premium should be sufficient to pay for the factors of production (including capital) but no more than that amount.

In order to simplify the discussion, a two-period model is considered, with flows at time 0 and 1. The model generalizes directly to multiple periods. Also for simplicity, expense flows are ignored. The MYERS-COHN cash flows are summarized in Table 1.

Premium flows occur at time 0 and loss flows at time 1. These flows are discounted at different rates (of course, the premium flow is not discounted at all in this simplified example because it occurs at time 0). Premium flows are considered virtually riskless and hence are discounted at the risk-free rate. Loss flows are obviously risky and are discounted at an appropriate risk-adjusted discount rate (RADR) (discussed below).

The underwriting profits tax is assumed to be paid at time 1. Since this is based on the difference between premiums and losses, the tax flow must be broken into two parts—the loss part and the premium part—each of which is discounted at the appropriate rate.

The other tax flow is the investment balance tax. Writing the policy creates an investment balance because the premium is paid in advance of the loss payment date and because the company commits surplus (equity capital) to the policy. The surplus and premium are invested and a tax must be paid on the investment income at time 1. It is assumed that the funds are invested at the risk-free rate, in which case the tax is discounted at that rate. MYERS has shown that the tax on risky investment income also should be discounted at the risk free rate (DERRIG (1985)).

Discounting each flow and simplifying the resulting expression leads to the following formula:

$$P = \frac{E(L)}{1+R_L} + \frac{\tau\delta P R_f}{1-\tau}\left(\frac{1}{1+R_f}\right) \tag{49}$$

where $E(L)$, P = expected losses and premium, respectively,

R_L = risk adjusted discount rate for losses,

τ = the corporate income tax rate, and

δ = the rate of surplus commitment (surplus-to-premiums ratio).

The risk adjusted discount rate for losses, R_L is equal to $R_f + \lambda$, where λ is the risk premium. A positive risk premium in the loss discount rate leads to a lower premium, while a negative risk premium is associated with higher premiums.

Consider, for example, the CAPM risk adjusted discount rate:

$$R_L = R_f + \beta_L(E_m - R_f) \tag{50}$$

where β_L = Cov $[L_1/E(L_0), R_m]$/Var (R_m), and

L_t = losses at time t.

The MYERS-COHN model is usually applied using a risk-adjusted discount rate based on the CAPM. However, it is not inherently a CAPM model, and any theoretically defensible risk adjusted discount rate could be used. To be theoretically defensible, the rate must be based on an economic model that takes into account rational behavior in a market context. For example, a rate could be obtained from the TURNER model or the APT instead of the CAPM. CUMMINS (1988b) has developed a discount rate that is appropriate for an insurer with non-zero default probability. The more stringent rules for selecting the discount rate imposed by financial modeling represent a significant difference between the MYERS-COHN approach and traditional actuarial discounted cash flow models.

A feature of the model that is obscured in the single period formula is the concept of the *surplus flow*. MYERS and COHN assume that surplus is committed to the policy when the policy is issued and gradually released as losses are paid. The surplus flow pattern has an important impact on the premium because it affects the investment balance tax.[24]

An unanswered question in this model and, indeed, in financial modeling in general, is the appropriate level of surplus commitment. Usually, the surplus-to-premium ratio is based on the company's historical average or on the average for the insurance market as a whole. Neither approach is satisfactory, particularly in a multiple-line company where lines have different risk characteristics. A solution to the surplus commitment and allocation problem would represent an important contribution to insurance financial theory.

The MYERS-COHN model is consistent with financial theory and relatively easy to apply in practice. The model is deceptively simple and avoids many subtle but important pitfalls in the definition and treatment of cash flows. It can probably be said to represent the state-of-the art in practical insurance financial models. Weaknesses include the difficulty of estimating the surplus-to-premiums ratio and the risk adjusted discount rate as well as the omission of default risk. None of these problems is inherent in the MYERS-COHN model.

11. OPTION PRICING MODELS

Like options, insurance can be interpreted as a derivative financial asset (contingent claim) where payments depend upon changes in the value of other assets. Payments under primary insurance policies are triggered by changes in

[24] Some regulatory jurisdictions use a *surplus block* rather than surplus flow approach. This usually means that all surplus is released at the end of the policy coverage period rather than being gradually released as losses are paid. The surplus block approach is unrealistic in its assumption that no surplus commitment is necessary during the loss runoff period.

value of the insured assets, while reinsurance payments depend upon the experience of the covered primary insurance policies. Both types of insurance are candidates for pricing using the option approach.

Option models have an advantage over traditional actuarial models in their strict adherence to the rules of dominance and arbitrage. To avoid arbitrage opportunities, rational option prices are based on the rule that options will be neither dominant nor dominating securities. This simple but powerful idea has far-reaching implications for the financial pricing of insurance.

An earlier section of this paper discussed the pricing of European call options. A European call was defined as a function $C(S, \tau) = C(S, \tau; K, \sigma, r)$ of the underlying stock price (S) and time to expiration (τ) such that $C(S, 0) = \text{Max}\,(S-K, 0)$, where K is the option's exercise price and the other parameters are defined as before. A European put option is a function $P(S, \tau) = P(S, \tau; K, \sigma, r)$ such that $P(S, 0) = \text{Max}\,(K-S, 0)$. An important relationship involving puts and calls on the same asset is the *put-call parity theorem*:

$$C(S, \tau) = S-[Ke^{-r\tau}-P(S, \tau)] \tag{51}$$

The theorem states that the value of a call is equal to the value of the underlying stock less the present value of the exercise price plus the value of the put.

Equation (51) provides a simple model of corporate financial structure that can be applied to insurance. Interpret S as the total market value of assets of the firm and K as the nominal value of liabilities at the exercise date (e.g., promised payments to policyholders). Then the market value of the equity holders' interest in the firm is equal to the value of the call. At the maturity of the option, the equity holders have the option of paying the value of the liabilities and keeping the net amount $S-K$. They will do this only if $S > K$ at the expiration date. If not, the equity holders will default and the policyholders become the owners of the firm, receiving assets insufficient to satisfy the outstanding liabilities.

The option to walk away from the firm if $S < K$ is a valuable right conferred upon the stockholders by the limited liability rule.[25] The value of this right is the value of the put, since $P(S, 0) = \text{Max}\,(K-S, 0)$. Thus, the net market value of the policyholders claims prior to expiration is the bracketed expression in (51): the riskless present value of the liabilities less the value of the put.

This model has some surprising implications. Because the partial derivatives of the call and the put with respect to the risk parameter (σ^2) are positive and equal (see JARROW and RUDD (1983) or INGERSOLL (1987)), increases in firm

[25] It would be interesting to modify the basic option model of the firm to take into account some additional characteristics of the insurance market. For example, regulators may take control of the firm if $S-K$ reaches some minimal (positive) value. In a multiple period model, stockholders presumably would consider the potential loss of future profits due to bankruptcy when selecting the firm's risk level, etc. Although such issues have been explored in the actuarial literature, it would be useful to reexamine them in the light of the insights provided by financial theory.

risk add to the market value of equity and correspondingly reduce the value of debt. Thus, stockholders have an incentive to increase risk because doing so increases their expected share of firm assets.

The options model can be used to obtain the market value of insurance policies subject to default risk. To do this, it is more realistic to have a version of the options model in which both assets and liabilities are stochastic. Models of this type are presented in CUMMINS (1988a) and DOHERTY and GARVEN (1986). Both assets and liabilities are assumed to follow diffusion processes:

(52a) $$\frac{dA}{A} = \alpha_A \, dt + \sigma_A \, dz_A$$

(52b) $$\frac{dL}{L} = \alpha_L \, dt + \sigma_L \, dz_L$$

where $A(t)$ = assets,

$L(t)$ = liabilities,

$dz_A(t)$, $dz_L(t)$ = possibly dependent Brownian motion processes for assets and liabilities,

α_A, α_L = instantaneous expected returns on assets and liabilities,

σ_A, σ_L = instantaneous standard deviations of return of assets and liabilities.

Insurance company debt is assumed to have a fixed maturity date τ. Its value is defined by $B(A, L, \tau)$.

B can be differentiated using Ito's lemma to yield:[26]

(53) $$dB = B_A \, dA + B_L \, dL + \left[\frac{1}{2} B_{AA} \sigma_A^2 + \frac{1}{2} B_{LL} \sigma_L^2 + B_{AL} \sigma_A \sigma_L \rho_{AL} + B_t\right] dt$$

where ρ_{AL} = the instantaneous correlation coefficient of dz_A and dz_L.

Substituting (52a) and (52b) into (53), making the changes of variables $X = A/L$ and $b = B/L$, and eliminating the risk term either by a hedging argument or by assuming that B is priced according to the ICAPM (see CUMMINS (1988a)), equation (53) is reduced to:

(54) $$r_r b = r_r b_X X + \frac{1}{2} (\sigma_A^2 + \sigma_L^2 - 2 \sigma_A \sigma_L \rho_{AL}) \, b_{XX} X^2 + b_t$$

where r_r = the real rate of interest, i.e., the difference between the nominal riskless rate of interest (r) and the anticipated inflation rate (see below).

[26] Subscripts on the function B represent partial differentiation, while subscripts on the parameters, σ and ρ, indicate whether the parameter pertains to the asset or the liability process.

Equation (54) is the BLACK-SCHOLES equation, where the optioned variable is the asset to liability ratio (X).

Defining $b(X, \tau)$ as the value of the debt (insurance liabilities), we have the boundary condition $b(x, 0) = \text{Min}(1, x)$. Solving equation (54) subject to this boundary condition and exploiting the fact that BLACK-SCHOLES options are homogeneous of degree 1 in the asset and exercise prices, yields the debt value:

(55) $$B(A, L, \tau) = AN(-d_1) + L e^{-r_f \tau} N(d_2)$$

where $d_1 = [\ln(x) + (r_f + \sigma^2/2)\tau]/(\sigma\sqrt{\tau})$,

$$d_2 = d_1 - \sigma\sqrt{\tau}, \text{ and}$$

$$\sigma^2 = \sigma_A^2 + \sigma_L^2 - 2\sigma_A\sigma_L\rho_{AL}.$$

The value of the debt is the fair value of the insurance, where the policy obligation is to pay the market value of L at time τ. Equation (52b) implies that L is lognormal, which is a reasonable assumption for insurance liabilities.

Differentiating (55) reveals that the value of liabilities is inversely related to risk and directly related to the asset to liability ratio. Thus, policies of safer companies will command higher prices in a competitive market.

As another application of option theory to insurance, consider setting retention limits and caps in excess of loss reinsurance. Excess of loss reinsurance is structured much like a call option. E.g., the ceding company's share in an excess of loss contract with no upper boundary is: $\text{Max}(Y - M, 0)$, where Y is the loss amount and M is the retention. Thus, the ceding company buys a call on the loss amount with exercise price M. If the policy has an upper limit (U), this can be considered a call written by the ceding company in favor of the reinsurer. The net value of the reinsurance is: $\text{Max}(Y - M, 0) - \text{Max}(Y - U, 0)$.

Pricing a reinsurance call illustrates some important features of option pricing theory. To use the conventional option model to value reinsurance, it is necessary to assume that the evolution of reinsurance liabilities is smooth so that a diffusion process can be used.[27] Another assumption is that trading takes place continuously. These assumptions might be approximated by stop-loss reinsurance or policies covering multiple exposure units.

If Y is the loss under the policy, the value of reinsurance can be written as $C(Y, \tau; M, \sigma, r)$, where as before $C(Y, 0; M, \sigma, r) = \text{Max}(0, Y - M)$. The hedge portfolio is $P = (1 - w)Y + wC$, where w = the portfolio weight assigned to the option. Another useful portfolio is the *replicating portfolio*, which is the combination of riskless bonds and stock that duplicates the option. This is

[27] Another possibility would be to use a *jump diffusion model*. A model where jumps follow a Poisson process has been developed by MERTON (1976).

easily obtained from the hedge portfolio: $C = P/w-[(1-w)/w]\,Y$, where P indicates holdings of riskless bonds. The weight w can be shown to be $C_Y(Y,\tau) = N(d_1)$, the partial derivative of the option value with respect to Y, where d_1 is defined above in the discussion of equation (55). The weight w is called the *hedge ratio* or *option delta.*

In the reinsurance case, the replicating portfolio would involve selling insurance (since the coefficient of $Y < 0$) and buying bonds. This could be viewed as a substitute for buying reinsurance. The difficulty with the interpretation is that it assumes that the portfolio weights can be continuously readjusted. This involves frequent trading in insurance policies, which could be difficult to accomplish.

Fortunately, discrete time hedging models are available that avoid the continuous trading assumption (see BRENNAN (1979)). A discrete time model has been applied to an insurance problem by DOHERTY and GARVEN (1986). More research is needed on the feasibility of hedging insurance portfolios and the practical importance of the continuous trading assumption in insurance. It is quite possible that adjustment of the hedge at infrequent intervals would bring about a reasonable approximation of the hedging result.

12. CONTINUOUS TIME DISCOUNTED CASH FLOW MODELS

Certainty model

Continuous time models for insurance pricing have been developed by KRAUS and ROSS (1982) and CUMMINS (1988a). To introduce this topic, consider the KRAUS-ROSS continuous-time dynamic model under conditions of certainty.

To simplify the discussion, assume that the value of losses is determined by a draw from a random process at time 0. Loss payments occur at instantaneous rate θ, while loss inflation is at exponential rate π, and discounting is at the riskless rate r. The differential equation for the rate of change in outstanding losses at time t, in the absence of inflation, is the following:

$$\frac{dC_t}{dt} = -\theta C_t \tag{56}$$

where C_t = the amount of unpaid claims at time t in real terms (i.e., not allowing for inflation). Solving this expression for C_t yields the following result: $C_t = C_0\,e^{-\theta t}$. Thus, the assumption is that the claims runoff follows an exponential decay process.

Considering inflation, the rate of claims outflow at any given time is: $L_t = \theta C_t e^{\pi t}$. The premium is the present value of losses, obtained as follows:[28]

[28] It is assumed that $r+\theta > \pi$.

(57) $$P = \int_0^\infty L_t e^{-rt}\, dt$$

$$= \int_0^\infty \theta C_0 e^{-(r+\theta-\pi)t}\, dt = \frac{\theta C_0}{r+\theta-\pi}$$

An important relationship affecting the premium is the *Fisher hypothesis*, i.e.: $1+r = (1+r_r)(1+i)$, where r_r = the real rate of interest and i = the anticipated inflation rate. This relationship states that nominal interest rates will compensate the investor for anticipated inflation and for the time value of money in an economy with no inflation (r_r). If insurance inflation (π) is the same as economy-wide inflation (i), then discounting takes place at the real rate since anticipated inflation in claims is offset by the anticipated inflation premium in the discount rate. In equation (57), π could be > or < i.

In (57), θ is the parameter of an exponential distribution. This implies that $1/\theta$ is the average time to payout, assuming no inflation. While the exponential provides some modeling insights, it is probably too simple a model to apply to actual claims runoff patterns. The author has found that the gamma distribution provides a reasonable fit to claims runoff data for automobile insurance in the United States.

The model also can be used to estimate reserves, as follows:

(58) $$C_T = \int_T^\infty \theta C_0 e^{-(\theta+r-\pi)t}\, dt = \frac{\theta C_0}{\theta+r-\pi} e^{-(\theta+r-\pi)t}$$

where C_T = the market value of reserves at time t.

Kraus-Ross uncertainty model

KRAUS and ROSS also introduce a continuous time model under uncertainty. This model is based on Ross' arbitrage pricing theory (APT). The KRAUS-ROSS model allows for market-related uncertainty in both frequency and claims inflation.

The following differential equation governs the claims process:

(59) $$\frac{dC}{dt} = \alpha(t) - \theta C(t)$$

where $\alpha(t)$ = accident frequency at time t. The frequency process affects the evolution of outstanding claims for a period of length T (the policy period). After that point, no new claims can be filed. During the entire period (0 to ∞) claims are settled at instantaneous rate θ but adjusted for inflation according to the price index $q(t)$.

The parameters $\alpha(t)$ and $q(t)$ are governed by the k economic factors of arbitrage pricing theory. In continuous time, these factors are modelled as diffusion processes:

(60) $$dx_i = m_i x_i dt + \sigma_i x_i dz_i, \qquad i = 1, 2, \ldots, k$$

The parameters are log-linear functions of the factors, e.g.:[29]

$$\log(q) = \sum_{i=1}^{k} q_i \log(x_i) + \log(q_0) \tag{61}$$

where q_0 is the price level of the average claim at policy inception, and q_i, $i = 1, 2, \ldots, k$ are coefficients (constants).

Arbitrage pricing theory implies that the value of outstanding claims at any time t, $V(x, C, t)$, where x is the vector consisting of the x_i, is governed by the following differential equation:

$$E\left[\frac{dV}{V}\right] + \left[\frac{\theta q C}{V} - r\right] dt = \sum_{i=1}^{k} \lambda_i \sigma_i \left[\text{Cov}\left(\frac{dV}{V}, \frac{dx_i}{x_i}\right) \Big/ \text{Var}\left(\frac{dx_i}{x_i}\right)\right] dt \tag{62}$$

where λ_i = the market price of risk for factor $i = (r_{mi} - r)/\sigma_i$, and

r_{mi} = the market return on a portfolio that is perfectly correlated with the ith risk factor.

The premium formula is obtained by applying the multivariate version of Ito's lemma (see INGERSOLL (1987)) and then solving the resulting differential equation. The formula is:

$$P = \left(\frac{\theta \alpha_0 q_0 L_0}{\rho + \theta}\right) \left[\frac{1 - e^{-\rho_\alpha T}}{\rho_\alpha}\right] \tag{63}$$

where $\rho = r - \pi + \sum_{i=1}^{k} \lambda_i \sigma_i q_i$

$$\rho_\alpha = r - \pi_\alpha + \sum_{i=1}^{k} \lambda_i \sigma_i (q_i + \alpha_i)$$

$$\pi = \sum_{i=1}^{k} \left[\frac{1}{2} \sigma_i^2 q_i (q_i - 1) + q_i m_i\right]$$

$$\pi_\alpha = \sum_{i=1}^{k} \left[\frac{1}{2} \sigma_i^2 (\alpha_i + q_i)(\alpha_i + q_i - 1) + (\alpha_i + q_i) m_i\right]$$

The premium given by (63) is similar to the premium for the certainty case except for the presence of the market risk loadings (λ_i terms) in the denominator. These loadings are the company's reward for bearing systematic risk. The α_i and q_i are the "beta coefficients" of the model.

[29] There is a directly analogous equation for $\alpha(t)$.

For the company to receive a positive reward for risk bearing, the risk loading term must be negative, i.e., losses must be negatively correlated with some of the market factors such that the net loading is < 0. This reduces the denominator and leads to a higher discounted value of losses.

The model requires estimates of the market prices of risk for the k risk factors as well as the beta coefficients for insurance. This would be difficult given the available data. Like most other financial pricing models for insurance, this model gives the price for an insurance policy that is free of default risk. Nevertheless, the KRAUS-ROSS model represents an important contribution to the financial pricing literature.

A model with default risk

CUMMINS (1988a) has developed a continuous time, exponential runoff model that prices default risk. This model can be used to value an insurance company or a policy cohort (block of policies). Assets and liabilities are hypothesized to follow geometric Brownian motion:

$$dA = (\alpha_A A - \theta L)\, dt + A\sigma_A\, dz_A \tag{64a}$$

$$dL = (\alpha_L L - \theta L)\, dt + L\sigma_L\, dz_L \tag{64b}$$

where α_A, α_L = asset and liability drift parameters,

σ_A, σ_L = asset and liability risk (diffusion) parameters,

A, L = stock of assets and liabilities,

θ = the claims runoff parameter, and

$dz_A(t)$, $dz_L(t)$ = possibly correlated standard Brownian motion processes.

The asset and liability processes are related as follows:

$$\rho_{AL} = \mathrm{Cov}\,(dz_A, dz_L) \tag{65}$$

The model is more realistic than the standard options model since it does not have a fixed expiration date but rather allows the liabilities to run off over an infinite time horizon. In effect, it models liabilities as a perpetuity subject to exponential decay. Thus, it is a better model for long tail lines than the standard options model.

CUMMINS uses the model to obtain the market value of default risk, $D(A, L)$. Using Ito's lemma to differentiate D and then using either a hedging argument or the ICAPM to eliminate the risk terms, one obtains the confluent hypergeometric differential equation. The solution is:

$$D(x) = \frac{\Gamma(2)}{\Gamma(2+a)}\, b^a x^{-a} e^{-b/x} M(2, 2+a, b/x) \tag{61}$$

where $a = 2(r_r + \theta)/Q$,

$b = 2\theta/Q$,

$Q = \sigma_A^2 + \sigma_L^2 - 2\sigma_A\sigma_L\rho_{AL}$, and

M = Kummer's function (see ABRAMOWITZ and STEGUN (1970)).

This perpetuity model has significant potential for pricing blocks of policies subject to default risk. It poses much easier estimation problems than the KRAUS-ROSS model since one need only estimate the variance and covariance parameters of assets and liabilities rather than betas and factor risk premia.

CONCLUSIONS

This paper discusses the principal asset pricing models that have been developed in financial economics and their applications in insurance. Insurance pricing models have been developed based on the capital asset pricing model, the intertemporal capital asset pricing model, arbitrage pricing theory, and option pricing theory. The distinguishing feature of these models is that they take into account the forces of supply and demand in the capital and insurance markets. The models assume either that insurance policies are priced in accordance with principles of market equilibrium or that they are priced so that arbitrage opportunities are avoided. These are important ideas, which need to be incorporated into the actuarial approach to pricing.

Additional theoretical and empirical research is needed to develop more realistic insurance pricing models. For example, most of the models assume that interest rates are non-stochastic even though insurers face significant interest rate risk. The traditional risk-free rate is actually free only of default risk.

With few exceptions, existing financial models do not price the risk of ruin. Estimation, especially for betas and market risk premia, is a major problem given the existing insurance data. Option models and perpetuity models such as CUMMINS' cohort model may offer solutions to some of these problems, since it is possible to incorporate stochastic interest rates and since these models rely on relatively few parameters.

Financial pricing models for insurance can be expected to evolve along with financial economics in general. This suggests that multi-variate diffusion models, the consumption capital asset pricing model (BREEDON (1979)), martingale pricing models (DUFFIE 1988)), and lattice models (BOYLE (1988)) may provide promising avenues for future research.

Research into financial pricing of insurance will be greatly facilitated by more interaction between financial and actuarial researchers. Finance theory currently is more advanced in using concepts of market equilibrium, hedging, and arbitrage, but actuarial theory offers more sophisticated and realistic models of the stochastic processes characterizing insurance transactions. Both areas of expertise must be brought to bear on the problem to arrive at meaningful solutions.

REFERENCES

ABRAMOWITZ, MILTON and STEGUN, IRENE (1970) *Handbook of Mathematical Functions* (New York: Dover Publications).

ANG, JAMES, S. and LAI, TSONG-YUE (1987) Insurance Premium Pricing and Ratemaking in Competitive Insurance and Capital Asset Markets. *Journal of Risk and Insurance* **54**, 767–779.

BIGER, NIHUM and KAHANE, YEHUDA (1978) Risk Considerations In Insurance Ratemaking. *Journal of Risk and Insurance* **45**, 121–132.

BLACK, FISCHER and SCHOLES, MYRON (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**, 637–654.

BORCH, KARL (1974) *The Mathematical Theory of Insurance*. (Lexington, MA: Lexington Books).

BOYLE, PHELIM (1988) A Lattice Framework for Option Pricing With Two State Variables. *Journal of Financial and Quantitative Analysis* **23**, 1–12.

BREALEY, RICHARD, A. and MYERS, STEWART C. (1988) *Principles of Corporate Finance*, 3rd ed. (New York: McGraw-Hill Book Company).

BREEDON, D.T. (1979) An Intertemporal Asset Pricing Model with Stochastic Investment and Consumption Opportunities. *Journal of Financial Economics* **6**, 273–296.

BRENNAN, M.J. (1979) The Pricing of Contingent Claims In Discrete Time Models. *Journal of Finance* **24**, 53–68.

BUHLMANN, HANS (1980) An Economic Premium Principle. *ASTIN Bulletin* **11**, 52–60.

BUHLMANN, HANS (1984) The General Economic Premium Principle. *ASTIN Bulletin* **14**, 13–21.

COOPER, ROBERT, W. (1974) *Investment Return and Property-Liability Insurance Ratemaking* (Philadelphia: S.S. Huebner Foundation, University of Pennsylvania).

COX, JOHN, D. and ROSS, STEPHEN (1976) The Valuation of Options for Alternative Stochastic Processes. *Journal of Finance* **31**, 383-402.

COX, JOHN, D. and RUBINSTEIN, MARK (1985) *Options Markets*. (Englewood Cliffs, N.J.: Prentice-Hall).

COPELAND, THOMAS, E. and WESTON, J. FRED (1988) *Financial Theory and Corporate Policy*, 3d. ed. (Reading, MA: Addison-Wesley Publishing Company).

CUMMINS, J. DAVID (1988a) Risk-Based Premiums for Insurance Guaranty Funds. *Journal of Finance* **43**, 823–839.

CUMMINS, J. DAVID (1988b) Capital Structure and Fair Profits In Property-Liability Insurance, working paper, University of Pennsylvania.

CUMMINS, J. DAVID (1990) Multi-Period Discounted Cash Flow Ratemaking Models In Property-Liability Insurance. *Journal of Risk and Insurance* **57**, 79–109.

CUMMINS, J. DAVID and CHANG, LENA (1983) An Analysis of the New Jersey Formula for Including Investment Income In Property-liability Insurance Ratemaking. *Journal of Insurance Regulation* **1**, 555–573.

CUMMINS, J. DAVID and DERRIG, RICHARD A. (1989) *Financial Models of Insurance Solvency* (Norwell, MA: Kluwer Academic Publishers).

CUMMINS, J. DAVID and HARRINGTON, SCOTT E. (1985) Property-Liability Insurance Rate Regulation: Estimation of Underwriting Betas Using Quarterly Profit Data. *Journal of Risk and Insurance* **52**, 16–43.

CUMMINS, J. DAVID and HARRINGTON, SCOTT E. (1987) *Fair Rate of Return In Property-Liability Insurance* (Norwell, MA: Kluwer Academic Publishers).

D'ARCY, STEPHEN and DOHERTY, NEIL A. (1988) *Financial Theory of Insurance Pricing* (Philadelphia: S.S. Huebner Foundation).

D'ARCY, STEPHEN and GARVEN, JAMES R. (1988) Financial Pricing Models of Property-Liability Insurance: An Empirical Evaluation, working paper, The Pennsylvania State University.

DERRIG, RICHARD, A. (1985) The Effect of Federal Income Taxes on Investment Income in Property-Liability Ratemaking, working paper, Automobile Insurers Rating Bureau of Massachusetts, Boston, MA.

DOHERTY, NEIL, A. and GARVEN, JAMES R. (1986) Price Regulation in Property-Liability Insurance: A Contingent Claims Approach. *Journal of Finance* **41**, 1031–1050.

DUFFIE, DARRELL (1988) *Security Markets: Stochastic Models* (New York: Academic Press).

FAIRLEY, WILLIAM (1979) Investment Income and Profit Margins In Property-Liability Insurance: Theory and Empirical Tests. *Bell Journal* **10**, 192-210.

FAMA, EUGENE, F. (1977) Risk-Adjusted Discount Rates and Capital Budgeting Under Uncertainty. *Journal of Financial Economics* **4**, 3–21.

GREENWALD, BRUCE, C. and STIGLITZ, JOSEPH E. (1990) Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior. *American Economic Review* **80**, 160–165.

HILL, RAYMOND (1979) Profit Regulation in Property-Liability Insurance. *Bell Journal* **10**, 172–191.

INGERSOLL, JONATHAN, E. Jr. (1987) *Theory of Financial Decision Making* (Totowa, NJ: Rowman & Littlefield).

JARROW, ROBERT (1988) *Finance Theory* (Englewood Cliffs, NJ: Prentice-Hall).

JARROW, ROBERT and RUDD, ANDREW (1983) *Option Pricing* (Homewood, IL: Richard D. Irwin).

KAHANE, YEHUDA (1979) The Theory of Insurance Risk Premiums — A Re-examination in the Light of Recent Developments in Capital Market Theory. *ASTIN Bulletin* **10**, 223–239.

KARLIN, SAMUEL and TAYLOR, HOWARD (1981) *A Second Course In Stochastic Processes* (New York: Academic Press).

KRAUS, ALAN and ROSS, STEPHEN (1982) The Determination of Fair Profits for the Property-Liability Insurance Firm. *Journal of Finance* **33**, 1015–1028.

LEVY, HAIM and SARNAT, MARSHALL (1984) *Portfolio and Investment Selection: Theory and Practice* (Englewood Cliffs, NJ: Prentice-Hall).

MARKOWITZ, HARRY (1952) Portfolio Selection. *Journal of Finance* **7**, 77–91.

MARKOWITZ, HARRY (1959) *Portfolio Selection: Efficient Diversification of Investment* (New York: John Wiley & Sons).

MAYERS, DAVID and SMITH, CLIFFORD W. (1982) On the Corporate Demand for Insurance. *Journal of Business* **55**, 281–296.

MERTON, ROBERT, C. (1973a) Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* **4**, 141–183.

MERTON, ROBERT, C. (1973b) An Intertemporal Capital Asset Pricing Model. *Econometrica* **41**, 867–880.

MERTON, ROBERT, C. (1976) Option Prices Where Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics* **3**, 125-144.

MULLER, H.H. (1988) Modern Portfolio Theory: Some Main Results. *ASTIN Bulletin* **18**, 127–146.

MYERS, STEWART and COHN, RICHARD (1987) Insurance Rate Regulation and the Capital Asset Pricing Model, in J.D. CUMMINS and S.E. HARRINGTON, eds., *Fair Rate of Return In Property-Liability Insurance* (Norwell, MA: Kluwer Academic Publishers).

OWEN, JOEL and RABINOVITCH, RAMON (1983) On the Class of Elliptical Distributions and Their Applications to the Theory of Portfolio Choice. *Journal of Finance* **38**, 745–752.

ROLL, RICHARD (1977) A Critique of the Capital Asset Pricing Theory's Tests: Part I: On Past and Potential Testability of the Theory. *Journal of Financial Economics* **4**, 129–176.

ROLL, RICHARD and ROSS, STEPHEN A. (1980) An Empirical Investigation of the Abritrage Pricing Theory. *Journal of Finance* **35**, 1073–1104.

ROSS, STEPHEN, A. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* **13**, 341–360.

ROSS, STEPHEN, A. (1977) Risk, Return, and Arbitrage, in I. FRIEND and J. BICKSLER, eds., *Risk and Return in Finance*, v.I. (Cambridge, MA: Ballinger Publishing Co.).

ROSS, STEPHEN, A. and WESTERFIELD, RANDOLPH (1988) *Corporate Finance* (St. Louis, MO: Times Mirror/Mosby College Publishing).

ROTHSCHILD, MICHAEL and STIGLITZ, JOSEPH E. (1970) Increasing Risk I: A definition. *Journal of Economic Theory* **2**, 225–243.

ROTHSCHILD, MICHAEL and STIGLITZ, JOSEPH E. (1972) Increasing Risk II: Its Economic Consequences. *Journal of Economic Theory* **3**, 66–84.

ROTHSCHILD, MICHAEL and STIGLITZ, JOSEPH E. (1976) Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics* **90**, 629–650.

SCHWEIZER, M. and FOLLMER, H. (1988) Hedging By Sequential Regression: An Introduction to the Mathematics of Options Trading. *ASTIN Bulletin* **18**, 147–160.

TURNER, ANDREW, L. (1987) Insurance in an Equilibrium Asset Pricing Model, in J.D. CUMMINS and S.E. HARRINGTON, eds., *Fair Rate of Return in Property-Liability Insurance* (Norwell, MA: Kluwer Academic Publishers).

J. DAVID CUMMINS
Department of Insurance, The Wharton School of the University of Pennsylvania, 3641 Locust Walk, Philadelphia, PA 19104-6218 USA.