THREE CENTURIES OF ASSET PRICING

by

Elroy Dimson* and Massoud Mussavian**

05 January 2000

* London Business School, Sussex Place, Regents Park, London NW1 4SA, UK. Telephone +44 (0)171 262 5050. Fax +44 (0)171 724 3317. Email edimson@lbs.ac.uk

** Salomon Brothers International, Victoria Plaza, 111 Buckingham Palace Road, London SW1E 0SB, UK. Telephone+44 (0)171 721 2000. Fax + 44 171 222 7062. Email massoud.mussavian@ssmb.com
THREE CENTURIES OF ASSET PRICING

by

Elroy Dimson and Massoud Mussavian

Abstract: Theory on the pricing of financial assets can be traced back to Bernoulli’s famous St Petersburg paper of 1738. Since then, research into asset pricing and derivative valuation has been influenced by a couple of dozen major contributions published during the twentieth century. These seminal works have underpinned the key ideas of mean-variance optimisation, equilibrium analysis and no-arbitrage arguments. This paper presents a historical review of these important contributions to finance.

Keywords: Asset pricing, option pricing, arbitrage, portfolio theory, risk measurement.

JEL classification: B00, G11, G12, G13.

1. Introduction

Finance is the branch of economics that focuses on the capital markets. In the early life of the subject, emphasis was placed on describing the market environment and valuing individual securities. In more recent years, attention has switched towards broader aspects of valuation. Modern finance has developed methodologies for valuing a wide variety of assets whose characteristics extend across time, and which impose intricate and complex risks on investors.

In this paper we provide an account of the development of asset pricing theory, defined broadly to include the valuation of a wide range of financial assets and derivative securities. The fact that one can present this as a coherent body of theory, rather than a disjointed collection of loosely connected publications, reflects the fact that a small number of key ideas has had a pervasive impact on the development of the subject. In particular modern financial theory is founded on three central assumptions: markets are highly efficient, investors exploit potential arbitrage opportunities, and investors are rational.

The theoretical asset pricing models have been particularly amenable to empirical testing. The rapid growth of computer technology in the latter part of the twentieth century, taken together with relatively easy access to stock and bond price data, has allowed researchers to examine whether the various asset pricing models are supported by rigorous statistical studies. Moreover, in a number of cases the paradoxes revealed in the data have in turn influenced the development of new theoretical models.

At the same time, asset pricing theory has had a direct impact on the world of business and finance. For example, the growth of derivatives market would not have been possible had it not been for the valuation models derived by academics. The large amounts of capital invested using these models has created additional pressures for researchers to discuss, develop and test these theories. The literature of asset pricing has therefore been crucial to advances in our understanding and in the development of financial markets.
The format of this paper is as follows. In section 2 we identify some of the earliest contributions to two themes: the tradeoff between risk and return, and the impact of arbitrage on asset pricing. In section 3 we turn to portfolio theory, discussing the consequences of co-movement between securities' returns, and how investors should incorporate risk into their investment decisions. These considerations give rise to the major models of asset pricing, which are discussed in section 4. In section 5 we turn to option pricing, and section 6 concludes.

2. Historical Origins

We can trace many of the issues addressed in modern finance back to the remarkable paper presented to the Imperial Academy of Sciences in St Petersburg by Daniel Bernoulli (1738). This paper, originally published in Latin and translated into German in 1896, was referenced widely in the fields of mathematics, logic and, subsequently, economics, but was not available in English until the 1950s. Nevertheless, Bernoulli addresses a series of issues that are at the core of modern financial economics.

Bernoulli examines the proposition that “expected values are computed by multiplying each possible gain by the number of ways in which it can occur, and then dividing the sum of these products by the total number of possible cases”. He rejects this approach because it fails to consider the range of possible outcomes that might occur. Instead he argues that “the determination of the value of an item must not be based on its price, but rather on the utility it yields”. Bernoulli suggests that increases in wealth will result in an increase in utility which is inversely related to the quantity of goods in an individual's possession, and this enables him to demonstrate the tradeoff between expected changes in wealth, and the risk associated with such an opportunity.

During the eighteenth and nineteenth centuries, Bernoulli's concept of utility was regarded as the province of mathematicians rather than economists. By that time, of course, economists such as Bentham had independently developed utility theory as a central element of economics. From the late nineteenth century onwards, Bernoulli's idea of decreasing marginal utility became central to economics, notably in the works of Jevons, Menger, Walras and Marshall. Bernoulli also introduced the concept of maximisation of expected utility. However, despite endorsement by Laplace and others, Bernoulli's approach had little impact on the economics of decision making under risk until the development of expected-utility theory by von Neumann and Morgenstern (1944, 1947) and Savage (1954).

The concept of risk is now pervasive in economics, and especially in financial economics. However this has not always been the case. Knight (1921) makes a distinction between risk and uncertainty. When the randomness facing an individual can be expressed in terms of numerical

---

1 A selection of seminal articles in the areas of asset pricing and derivative valuation are identified by being printed in bold typeface. These articles are to be published in Volume II of Elroy Dimson and Massoud Mussavian (eds) Foundations of Finance, Dartmouth Publishing Company, 1999, forthcoming. Further articles, identified by italicised bold typeface, are included in Volumes I and III.
probabilities, whether these are objective or reflect the individual's subjective beliefs, the situation is said to involve risk. When probabilities cannot be assigned to alternative outcomes, then the situation is said to involve uncertainty.

In a setting that embraces both risk and uncertainty, Arrow and Debreu developed a model of general equilibrium that has been fundamental to economics and finance. Their work starts with a series of papers, notably Arrow (1951), Debreu (1951) and Arrow-Debreu (1954). Arrow and Debreu assume that markets are complete - that is, there are as many markets as goods - and this provides a framework for analysing general equilibrium. Since each good is defined by attributes such as its physical characteristics, its location, the date that it becomes available and the state of nature when it is available, the Arrow-Debreu model might be seen as one in which Pareto efficient outcomes could only occur with an almost infinite number of markets. However, Arrow was not satisfied with a framework that could be applied only when markets are perfectly complete.

Arrow's theory of general equilibrium with incomplete asset markets is presented in Arrow (1953). He shows that, by using the temporal structure of the economy, equilibrium can be attained with a more limited number of markets. He explains how one can achieve markets that are almost complete by setting up a series of contingent claims that follow the resolution of uncertainty. This has provided a conceptual framework to underpin the theory of asset pricing.

Arrow's concept of a complete market is one in which it is possible to insure against any loss that is of concern to any individual. In an economy where it is possible to insure against the range of possible future outcomes, individuals are more likely to be willing to bear risk. He provides encouragement for investors to hold diversified portfolios, rather than putting all their eggs in one basket. Arrow’s framework set the scene for a more structured analysis of investors’ portfolio decisions.

3. Portfolio Theory and Risk Measurement

3.1 Portfolio Selection

Finance was transformed with the publication of the Markowitz (1952) article on Portfolio Selection. Ever since the days of Bernoulli, it was clear that individuals would prefer to increase their wealth, and also to minimise the risk associated with any potential gain. But could these two criteria be combined? Markowitz considers and rejects the idea that there might be a portfolio, which gives both the maximum expected return and the minimum variance. He explains that, "the portfolio with maximum expected return is not necessarily the one with minimum variance. There is a rate at which the investor can gain expected return by taking on variance, or reduce variance by giving up expected return".

The most important contribution made by Markowitz is his distinction between the variability of returns from an individual security and its contribution to the riskiness of a portfolio. He notes that "in trying to make variance small it is not enough to invest in many securities. It is
necessary to avoid investing in securities with high covariances among themselves". This insight underpins most of the papers discussed in this review.

Markowitz shows that provided we have the appropriate input data and computing power (a substantial proviso), then we can identify a set of portfolios that provide the highest possible expected return for a given level of risk, while at the same time giving the lowest level of risk for each level of expected return. These portfolios form the efficient frontier, and Markowitz shows that for any investor who only cares for the tradeoff between expected return and risk, it is economically efficient to limit choice to portfolios that fall on this frontier.

Another approach, which shares many similarities to Markowitz's framework, was the "safety first" model developed independently by Roy (1952). Roy addresses the question of how individuals can ensure a suitably small probability that their wealth will fall below some disaster level, which is essentially the same problem as that addressed by Markowitz. Roy's representation of his problem, with risk as the independent variable and expected return as the dependent variable, was adopted as standard by the finance profession. Perhaps because the Portfolio Selection article was published a few months earlier, however, it is Markowitz who is generally regarded as the "godfather" of portfolio theory.

Tobin (1958) takes Markowitz's analysis one step further by showing how to identify which efficient portfolio should be held by an individual investor. He considers how an investor should divide his or her funds between a safe liquid asset such as cash (or treasury bills) and a risky asset (a bond or equity portfolio). He shows that "the proportionate composition of the non-cash assets is independent of their aggregate share of the investment balance. This fact makes it possible to describe the investor's decisions as if there was a single non-cash asset, a composite formed by combining the multitude of actual non-cash assets in fixed proportions".

Tobin therefore proposes a framework for asset allocation that is intuitively appealing. He proposes "breaking down the portfolio selection problem into stages at different levels of aggregation - allocation first among, and then within, asset categories." The asset mix, namely the allocation to cash (or treasury bills), should reflect the degree of risk aversion or risk tolerance of the investor. The optimal portfolio of risky assets, however, should be independent of the risk preferences of the investor. This proposition, which is known as the Separation Theorem, provides a basis for identifying the efficient portfolio.

Tobin's separation theorem clarified the task of portfolio selection. But even with Tobin's contribution it was still necessary to use Markowitz's full covariance model. The data and computational requirements of this approach were onerous, particularly for applications that embrace individual securities. There are well over two thousand shares traded in the British equity market, and even more in the United States. To use the Markowitz model with 2000 securities requires estimates of over two million risk and return characteristics. This is clearly an impossible data requirement. In addition, unless the universe of securities were limited to a few dozen stocks, the computational task was beyond the most powerful computers available anywhere in the 1950s and early 1960s.
These difficulties were addressed when Sharpe (1963) devised his Simplified Model for Portfolio Analysis. Sharpe draws on an insight of Markowitz (1959) that stocks are likely to co-move with the market. His model assumes that security returns are linearly related to fluctuations in a market-wide index, with a known degree of sensitivity; and that additionally, security-specific returns are generated with a known mean and variance. With only three parameters per security, the tasks of risk measurement and portfolio optimisation are greatly simplified. Sharpe's approach is readily extended to embrace richer and more complex factor models of asset pricing (see sections 3.2 and 4.4 below).

The Markowitz full-covariance model and Sharpe's index model, together with Sharpe's development of the capital asset pricing model (see section 4.1) marked the end of the beginning of modern finance. Arrow and Tobin separately received Nobel prizes in Economics, in part for their respective contributions to the theory of finance. In 1991, the fundamental contributions of Markowitz and Sharpe were honoured by the first Nobel prize to be awarded for research in financial economics (an award shared with Miller, primarily for his contributions to corporate finance). In turn, others were able to build on these early foundations, and in 1997 the theory of asset pricing was once again recognised through the award of the Nobel prizes to Scholes and Merton for their work on valuing derivative securities (see section 5).

3.2 Risk Measurement

After Sharpe's initial work, the first major empirical study of security risk attributes was by King (1966). King's classic study examines the returns of 63 NYSE stocks between 1927 and 1960. The stocks are drawn from the tobacco, petroleum, metals, railroad, utilities and retail store sectors. Through a simple process of analysing the co-movements of the stocks' returns, King shows how share prices tend to fluctuate in line with the market, and quantifies the extent to which their variability is attributable to industry membership.

King's study triggered substantial research into the tendency of stocks to co-move with one another, for reasons other than their sensitivity to the overall market. Such models go beyond the single factor of Sharpe's index model, and encompass the tendency for companies to move together when they belong to the same industry group, are unusually small or large, have a value or growth orientation, and so on. Examples include Rosenberg and Marathe (1975), Roll and Ross (1980) and Chen, Roll and Ross (1986) - the latter articles are discussed in section 4.4 below.

When securities are aggregated into a portfolio, extra-market common factors and firm-specific risks are rapidly diversified away. For diversified portfolios the most important contributor to risk is their sensitivity to market fluctuations, or beta. Since investors must be concerned with risk subsequent to making an investment, this raises the important question of whether beta can be predicted. Blume (1971) answers this in his article, On the Assessment of Risk.

Blume compares betas measured over successive, non-overlapping periods. Security betas have a correlation over these periods of some 0.5, and as stocks are aggregated into portfolios this correlation increases towards 0.99. Security risk attributes can be predicted with a modest
degree of reliability, but portfolio risk attributes can be estimated with much greater accuracy. Blume also shows how risk characteristics tend to regress towards the mean, an attribute that is crucial in a wide range of empirical applicators. Dimson and Marsh (1983) provide confirmatory evidence for the UK, and present comparisons with the results found in a variety of other markets.

The earliest empirical studies employed monthly prices for stocks traded on the New York Stock Exchange. As research databases became available for markets with a lower volume of trading, and with the development of daily and even intradaily databases, researchers became aware of the problems posed by infrequent trading. Because security prices do not reflect transactions that all occur simultaneously, at the end of a trading period, there is a tendency for risk measures to be biased, especially for infrequently traded securities. The paper by Dimson (1979) builds on earlier work (Dimson, 1974) to develop methods for estimating risk when there are frictions in the trading process. This approach, or the alternative developed by Scholes and Williams (1977), is widely used in risk estimation.

In the early years, empirical work typically assumed a constant variance of returns over time; for example, see the studies reviewed in Dimson and Mussavian (1998). A large body of empirical literature going all the way back to Mandelbrot (1963), as well as anecdotal evidence such as the 1929 and 1987 stock market shocks, indicates that risk characteristics vary over time. Even if the unconditional variances and covariances are constant, variances and covariances may be non-constant conditional on the past.

The major breakthrough in this area is by Engle (1982), who noticed that large daily (and even monthly) price moves tend to be followed by subsequent large price moves and that this phenomenon is common to many security markets. In order to capture this clustering of volatility, Engle introduced autoregressive conditionally heteroscedastic (ARCH) models to finance. In an ARCH model, the recent past gives information about the next period's forecast variance. Along with the Generalised ARCH model formulated by Bollerslev (1986) and other variations reviewed, eg, in Bollerslev, Chou and Kroner (1992), there is now a broad family of ARCH-type processes that are widely used as parsimonious models of return volatility.

4. **Asset Pricing Theory**

4.1 **The Capital Asset Pricing Model**

In the year that Markowitz's (1959) Portfolio Selection book was published, Treynor started intensive work on the theory of asset pricing. Bernstein (1992) observes of Treynor that "he is neither an academic nor a practitioner in any conventional sense. Although he took some graduate courses, he never earned a PhD. He has written many articles, but he has written no book on the theory of finance. His most important paper was never published; it is occasionally cited in academic literature, always as an 'unpublished manuscript'. His name does not even appear on my mimeographed copy of that manuscript."
This important manuscript is **Treynor's (1961)** paper, Toward a Theory of Market Value of Risky Assets, an unsigned "rough draft" which is undated\(^2\). The intention of Treynor's paper is "to lay the groundwork for a theory of market value which incorporates risk". The paper's aims are "1. to show that under our assumptions, optimal portfolio-balancing behaviour by individual investors leads to Proposition I of the famous Modigliani-Miller paper; 2. to explore the manner in which risk affects investment value; and 3. to introduce the concept of insurability. Insurable risks have a negligible effect on the cost of capital." He shows that "the risk premium per share for the investment is proportional to the covariance of the investment with the total of all the investments in the market."

Shortly after Treynor began his work on asset pricing, Sharpe also set out to determine the relationship between the prices of assets and their risk attributes. The paper published by **Sharpe (1964)** notes that "through diversification, some of the risk inherent in an asset can be avoided so that its total risk is obviously not the relevant influence on its price; unfortunately little has been said concerning the particular risk component which is relevant." Sharpe aims to use the theory of portfolio selection "to construct a market equilibrium theory of asset prices under conditions of risk" and notes that his model "sheds considerable light on the relationship between the price of an asset and the various components of its overall risk."

Sharpe submitted the initial version of his capital asset pricing model paper to the *Journal of Finance*. According to Bernstein (1992), the Journal's editor rejected it on the grounds that "his assumption that all investors made the same predictions was so `preposterous' as to make his conclusions `uninteresting'." Published (after revision) the following year, Sharpe's (1964) paper was soon supplemented by contributions from Lintner (1965) and Mossin (1966). The resulting capital asset pricing model (CAPM) is the model of investors' return expectations that was to remain dominant as a research paradigm until the 1980s.

After publication of the Sharpe, Lintner and Mossin articles, there was a wave of papers seeking to relax the strong assumptions that underpin the original CAPM. The most frequently cited modification is by **Black (1972)**, who shows how the model needs to be adapted when riskless borrowing is not available; his version is known as the zero-beta CAPM. Another important variant is by Brennan (1970), who finds that the structure of the original CAPM is retained when taxes are introduced into the equilibrium. Mayers (1972) shows that when the market portfolio includes non-traded assets, the model also remains identical in structure to the original CAPM. The model can also be extended to encompass international investing, as in Solnik (1974) and Black (1974). The theoretical validity of the CAPM has even been shown to be relatively robust if the assumption of homogenous return expectations is relaxed, as in Williams (1977). Finally, there are extensions from the classical one-period setting to a

\(^2\) This paper is now published in **Dimson and Mussavian (1999)**. Though it has been necessary to retypew the manuscript, we have made the minimum of editorial changes (and where we have done so, this is identified in a footnote). Jack Treynor, in a personal communication, advises us that the year in which the manuscript was completed is more probably 1962. But since it is universally referred to as "Treynor (1961)", e.g. at the beginning of the seminal Black-Scholes (1973) paper, we have retained 1961 as the presumed date of the manuscript.
Continuous time environment, as discussed in section 4.3 below.

The concepts of portfolio theory and the development of risk measurement, taken together with the capital asset pricing model, have had a major impact on the theory and practice of investment management. It is now common to view a managed portfolio as a blend of a passive portfolio (such as index fund) and an active portfolio comprising a series of bets on the relative performance of individual securities. Treynor and Black (1973) show how best to construct such portfolios by linking the CAPM with Sharpe's (1963) index model. They explain when a portfolio manager should choose to run an almost perfectly diversified index fund, and how the portfolio's diversification should vary with the prospects for the stocks in which the portfolio is invested; they also provide the first analysis to underpin market-neutral hedge funds. Modern portfolio optimisation and risk management systems are often extensions of the Treynor-Black model.

4.2 Tests of the CAPM

As was made clear by Popper (1934), scientific theories should lead to propositions that are potentially falsifiable (or verifiable) by experimental observation. The CAPM is no exception. Although the CAPM had already been developed, as a theory, in the early 1960s, two important ingredients were still lacking: a database of stock returns and the ability to process this data. Fortunately, by the late 1960s empirical analysis of the CAPM became possible after the construction of the Center for Research in Stock Price (CRSP) database at the University of Chicago, funded by stockbrokers Merrill Lynch, and the availability of increasingly cheap and powerful computer technology.

Early empirical investigations of the CAPM were based on the natural implications that arise from the theory: higher returns should be expected from stocks that have higher beta, and the relationship between expected return and beta should be linear. In the case of the Treynor-Sharpe-Lintner-Mossin CAPM, the slope of this line should be equal to the market risk premium, and the intercept should be equal to the risk-free rate. For the zero-beta CAPM, the slope should be less than the market risk premium, while the intercept should be greater than the risk-free rate. Moreover, there should be no systematic reward for bearing non-market risk, and any deviations in realised returns from the CAPM should not be predictable.

These implications provided the testable hypotheses for the early empirical studies of the CAPM. Black, Jensen and Scholes (1972) performed the earliest rigorous tests of the CAPM. The authors found that "the cross sectional plots of the mean excess returns on the portfolios against the estimated betas indicate that the relation between mean excess return and beta was linear". This is potentially consistent with the some type of CAPM. But they also find that "the intercept and the slope of the cross-sectional relation varied with different subperiods and were not consistent with the traditional form of the capital asset pricing model." This, and subsequent evidence by Fama and MacBeth (1973) and Blume and Friend (1973), although not consistent with the Treynor-Sharpe-Lintner-Mossin CAPM, could, however, be explained by the zero-beta version of the CAPM.
The Black-Jensen-Scholes (BJS) and Fama-MacBeth studies were methodological breakthroughs. Many subsequent tests of the CAPM (and of the arbitrage pricing theory, as we explain in section 4.4) employ techniques derived from these works. In principle, to test the CAPM a researcher would regress security or portfolio returns (the dependent variable) on their betas (the independent variable). However, beta is not known and can only be estimated with error, and this violates the assumptions underpinning regression. To solve this problem BJS and Fama-MacBeth construct an ingenious test involving two steps, which has consequently become known as the "two-pass" methodology. At the first pass they run a time series regression of portfolio returns on the market return, which gives estimates of portfolio betas. To gain maximum efficiency the portfolios are pre-sorted into various groups based on their beta. At the second pass, BJS regress, in cross-section, average returns on the estimated betas from the first pass. This cross sectional regression provides a test of the traditional CAPM. However, BJS note that this does not give “any direct tests aimed at explaining the existence of the beta factor”, i.e. the zero-beta CAPM. To overcome this limitation, Fama-MacBeth modify the second pass by performing cross-sectional regressions on a month-by-month basis and then taking the time-series average of the estimated risk premium. This, it turns out, allows them to test directly for the validity of the zero-beta CAPM.

A major turning point in empirical tests of the CAPM was the devastating Roll (1977) critique. Previous tests of the CAPM examine the relationship between equity returns and beta measured relative to a broad equity market index such as the S&P500. However, Roll demonstrates that the market, as defined in the theoretical CAPM, is not a single equity market, but an index of all wealth. The market index must include bonds, property, foreign assets, human capital and anything else, tangible or intangible, that adds to the wealth of mankind. Roll points out that "the portfolio used by Black, Jensen and Scholes was certainly not the true portfolio". Moreover, Roll shows that unless this market portfolio were known with certainty then the CAPM never could be tested. Finally, Roll argues that tests of the CAPM are at best tests of the mean-variance efficiency of the portfolio that is taken as the market proxy. But within any sample, there will always be a portfolio that is mean-variance efficient; hence finding evidence against the efficiency of a given portfolio tells us nothing about whether or not the CAPM is correct.

The two-pass method of BJS and Fama-MacBeth also suffers from an inherent statistical deficiency known as the errors-in-variable problem. This arises because the second-pass independent variables (ie, the betas) are themselves estimates from the first-pass regression. This gives rise to a statistical error which typically causes the estimated risk premium to be smaller in magnitude than the true risk premium. Although BJS and Fama-MacBeth reduce the impact of this bias by forming portfolios, they cannot eliminate it. A more straightforward solution is given by Gibbons (1982), who proposes a methodology that directly tests the restriction on returns imposed by the CAPM. Gibbons’ method, which is based on maximum likelihood estimation, avoids the need for separate steps. Instead, by estimating the beta and the risk premium simultaneously, Gibbons claims that “the methods suggested ... not only avoid the errors-in-variables problem, but the approach also increases the precision of parameter estimates for the risk premium”. Nevertheless this approach still rejects the CAPM.
The fallout from the Roll critique has been that subsequent tests of the CAPM are interpreted as tests of the mean-variance efficiency of the portfolio that is a proxy for the market. For example a new set of tests based on multivariate tests statistics has been introduced by Gibbons, Ross and Shanken (1989) who argue that “since the theory is equivalent to the assertion that the market portfolio is mean-variance efficient, we wish to test whether any particular portfolio is mean variance efficient”. However a number of authors have tried to tackle the Roll critique. Shanken (1987) and Kandel and Stambaugh (1987) both argue that, even though the stock market is not the true market portfolio, it must nevertheless be highly correlated with the true market. Unfortunately, even with this insight they find evidence that the CAPM does not seem to hold. Another response to the Roll critique is the use of proxies that include broader sets of assets such as bonds and property. However, Stambaugh (1982) finds that even when bonds and real estate are included into the market proxy the CAPM is still rejected.

Apart from these tests there was mounting evidence that other risk factors also affect stock returns. The factors include the price/earnings ratio (Basu, 1977), company size (Banz, 1981), book-to-market equity (Fama and French, 1992) and a variety of other systematic influences on security prices (see Dimson and Mussavian, 1998). The empirical evidence provided further motivation for research into other models of asset pricing that might more successfully explain returns, or at least indicate why in practice the CAPM did not seem to be hold. The first developments involved extending the one-period model into a multi-period framework.

4.3 Intertemporal Asset Pricing

A key assumption in Markowitz portfolio optimisation and the original CAPM is that agents make decisions for only one time period. This is clearly an unrealistic assumption since investors can and do rebalance their portfolios on a regular basis. Moreover daily movements in the prices of many assets cannot be explained by the ordinary CAPM. This limitation of the CAPM was well understood and by the late 1960s researchers were trying to determine whether the ordinary CAPM would hold in a dynamic setting. Early examples of this are the intertemporal portfolio choice and asset pricing models of Samuelson (1969), Hakansson (1970) and Fama (1970), which assume that agents make portfolio and consumption decisions at discrete time periods.

In order to construct a framework that is both more realistic and at same time, more tractable than the discrete time model, Merton (1973) developed an Intertemporal CAPM (ICAPM) by assuming that time flows continuously. The framework of continuous time turns out to be one of the major developments of modern finance, in both equilibrium asset pricing and derivative valuation. As Merton (1990) states in the introduction to his book of collected papers, “The mathematics of the continuous-time model contains some of the most beautiful applications of probability and optimisation theory. But, of course, not all that is beautiful in science need also be practical. And surely, not all that is practical in science is beautiful. Here we have both”. However Merton’s contribution does not end there; Ross (1989) states that Merton (1973) produced "the first such model that went beyond simple analogies with the static models to
introduce and approach a new phenomenon that is intrinsically intertemporal in nature."

One of Merton's key results is that the static CAPM does not in general hold in a dynamic setting and “that the equilibrium relationships among expected returns specified by the classical capital asset pricing model will obtain only under very special additional assumptions”. In particular, Merton demonstrates that an agent's welfare at any point in time is not only a function of his own wealth, but also the state of the economy. If the economy is doing well then the agent's welfare will be greater than if it is doing badly, even if the level of wealth is the same. Thus the demand for risky assets will be made up not only of the mean-variance component, as in the static portfolio optimisation problem of Markowitz (1952), but also of a demand to hedge adverse shocks to the investment opportunity set. Merton summarises his result with the following example: “An intertemporal investor who currently faces a five per cent interest rate and a possible interest rate of either two or ten per cent next period will have portfolio demands different from a single-period maximizer in the same environment or an intertemporal maximizer facing a constant interest rate of five per cent over time.” The upshot is that a CAPM will hold at each point in time, but there will be multiple betas; the number of betas will be equal to the one plus number of state variables that drive the investment opportunity set through time.

Although a major breakthrough, Merton’s analysis was at the same time disconcerting because it runs counter to the basic intuition of the CAPM, that an asset has greater value if its marginal contribution to wealth is greater. Breeden (1979), however, reconciled Merton's ICAPM with the classical CAPM by highlighting the dichotomy between wealth and consumption. In an intertemporal setting, Breeden showed that agents’ preferences must be defined over consumption and thus _always, when the value of an additional dollar payoff in a state is high, consumption is low in that state, and when the value of additional investment is low, optimal consumption is high. This is not always true for wealth, when investment opportunities are uncertain_. The implication is that assets are valued by their marginal contribution to future consumption and not wealth. Breeden's model which became known as the Consumption CAPM (CCAPM) allows assets to be priced with a single beta as in the traditional CAPM. In contrast to the latter the CCAPM’s beta is measured not with respect to aggregate market wealth, but with respect to an aggregate consumption flow and, as Breeden states, _the higher that an asset’s beta with respect to consumption is, the higher its equilibrium expected rate of return_.

One troubling feature of both Merton's ICAPM and Breeden's CCAPM still remained. Although in these papers the demand side of the capital markets was intricately built up from the microeconomic choices of consumers, the microeconomic choices of firms - which determine the supply of assets - was largely assumed away. On this point Merton (1990) himself says “as in the development of the original CAPM, the analysis emphasizes the demand side of the capital markets and thus treats as largely exogenous the dynamics of the supply curves for securities. Therefore the model does not provide all the structural equations of endogenous behavior required for a full equilibrium analysis of the system”. In other words these papers did not explicitly derive the prices in an economy with production as well as exchange.
Cox, Ingersoll and Ross (1985a) were the first to derive an explicit rational expectations equilibrium that endogenously calculates the price of assets. Indeed, versions of this paper had been circulating since the mid-1970s (see LeRoy, 1989) and Breeden (1979), amongst others, cites Cox, Ingersoll and Ross as a working paper with a 1977 date. The primitives of their model are not only the preferences of investors and the fundamental sources of risk, but also the underlying productive technology in the economy.

This framework for the first time in financial economics allowed an intriguing possibility: to link the interest rate and the yield curve to the fundamentals of the economy. Asset pricing models such as the ordinary CAPM, ICAPM and CCAPM had taken the interest rate as given exogenously. On the other hand, prior analysis of interest rates has been in terms of various ad-hoc maturity premia (Cox, Ingersoll and Ross, 1981). In a companion paper which had originally been part of the general equilibrium paper, Cox, Ingersoll and Ross (1985) argue that “while the focus of such modern and eclectic analyses of the term structure for explaining and testing the term premiums is desirable, there are two difficulties with this approach. First, we need a better understanding of the determinants of the term premiums. The previous theories are basically only hypotheses which say little more than that forward rates should or need not equal expected spot rates. Second, all of the theories are couched in ex ante terms and they must be linked with ex post realisations to be testable”. They go on to “consider the problem of determining the term structure as being a problem in general equilibrium theory” and their “approach contains elements of all of the previous theories”. Hence this model uses fundamental factors to determine the interest rate and the shape of the whole yield curve, and they are able to predict how these factors will impact the term structure.

Since the late 1970s, work in dynamic asset pricing has "with relatively few exceptions, been a mopping-up operation" (Duffie, 1996). The emphasis has largely been on weakening the underlying assumptions of the frameworks introduced by Merton (1973), Breeden (1979) and Cox, Ingersoll and Ross (1985, 1985a), and developing a general framework linking dynamic asset pricing to the original work of Arrow and Debreu.

One important insight from Merton’s ICAPM is that multiple risk factors are needed to explain asset prices. At the same time as the multiperiod framework was being developed, a number of researchers were also trying to use this insight to obtain single-period models that could better explain returns and risks. We go on to discuss the models in the next section.

4.4 Arbitrage Pricing Theory

Around the time that the shocking truth of the Roll critique was sinking in, Ross (1976) developed the arbitrage pricing theory (APT) as an alternative model that could potentially overcome the CAPM’s problems while still retaining the underlying message of the latter. The core idea of the APT is that only a small number of systematic influences affect the long term average returns of securities.

The first ingredient of Ross’s APT is a factor model. Unlike Sharpe’s (1963) single-index model, there are multiple factors that represent the fundamental risks in the economy. Multi-
factor models allow an asset to have not just one, but many measures of systematic risk. Each measure captures the sensitivity of the asset to the corresponding pervasive factor. This may seem similar to the multiple betas in Merton's ICAPM, but the resemblance to the CAPM ends here for, as Ross states, the APT “is much more an arbitrage relation than an equilibrium condition”. If the factor model holds exactly and assets do not have specific risk, then the law of one price implies that the expected return of any asset is just a linear function of the other assets’ expected return. If this were not the case, arbitrageurs would be able to create a long-short trading strategy that would have no initial cost, but would give positive profits for sure.

The intuition for the result when assets have no specific risk, is that all asset prices move in lockstep with one another and are therefore just leveraged ‘copies’ of one other. The result becomes more difficult when assets do have specific risk. In this case it is possible to form portfolios where the specific risk may be diversified away. To achieve full diversification of residual risk, however, a portfolio needs to include an infinite number of securities. With a finite set of securities, each of which has specific risk, the APT pricing restriction will only hold only approximately.

Almost from the inception of the APT, the choice of factors, number of factors and their interpretation has been hotly debated. One of the earliest empirical studies of the APT, by Roll and Ross (1980), uses factor analysis, a statistical technique that allows the researcher to infer the factors from the data on security returns. Their results indicate that there are four priced factors in the US stock market. The advantage of factor analytic techniques is that the factors determined from the data explain a large proportion of the risks in that particular dataset over the period under consideration. The drawback is that factors usually have no economic interpretation. As Roll and Ross argue, “an effort should be directed at identifying a more meaningful set of sufficient statistics for the underlying factors”.

An alternative to factor analytic techniques is to use observed macroeconomic variables as the risk factors. One of the first studies using observed factors was by Chen, Roll and Ross (1986). Their argument is that at the most basic level we can imagine that some fundamental valuation model determines the prices of assets. That is, the price of a stock will be the correctly discounted expected future dividends. Therefore the choice of factors should include any systematic influences that impact future dividends, the way traders and investors form expectations, and the rate at which investors discount future cash flows.

They find that US stock prices are significantly related to (1) changes in industrial production, (2) the spread between the yield on short-term and long-term government bonds, (3) the spread between low- and high-grade bonds, (4) changes in expected inflation, and (5) changes in unexpected inflation. The spread between short and long-term government yields is interpreted as a proxy for the business cycle, and the spread between low- and high-grade bond yields is viewed as a proxy for overall business risk in the economy. It is also interesting to note that, after accounting for these macroeconomic factors, the overall stock market index has no further impact on average individual stock returns. Similarly, and in contrast to popular opinion, after these macro factors have been considered, shocks to oil prices also have no additional impact on stock prices.
However, the APT has itself not been without controversy, with the arguments mirroring the
discussion that raged about the testability of the CAPM. Shanken (1982, 1985) asserts that for
individual securities the approximation implied by Ross' APT is so imprecise that it makes it
impossible ever to test whether the APT is true or false. Furthermore, Shanken argues that since
the expected return for any security or portfolio is related only approximately to its factor
sensitivities, to get an exact pricing relationship additional assumptions are needed. He
maintains that researchers who test the APT by assuming that the restriction holds, even for
securities, are actually testing an equilibrium form of the APT. Hence they are again confronted
with all the inherent difficulties that arise when testing the CAPM. Although Dybvig and Ross
(1985) have responded to these criticisms of the APT, the fact remains that, like the CAPM,
there are fundamental limitations to any empirical verification of the APT.

5. Valuation of Derivatives

Options and other derivatives are assets whose payoffs depend on the value of another asset.
Although there are a multitude of different and highly complex options traded in financial
markets, the most common are futures and forwards, and call and put options. Futures and
forwards are contracts traded either through an exchange or over the counter where the buyer
(seller) agrees to buy (sell) the underlying asset at some predetermined date and at some
predetermined price. A call (put), on the other hand, gives the right to exercise the option, ie to
buy (sell) the underlying asset, at some predetermined price at some future date. Both types of
derivatives have been part of economic life since for several thousand years; in The Republic,
Aristotle conveys the story of Thales the Milesian who used options to create a squeeze in the
market for olive oil pressing. The first “modern” futures exchange was in existence as early as
1730 in Japan (Schaede, 1989). Early discussions on the pricing of options and forwards
include Bachelier (1900), Keynes (1930) and Hicks (1934).

However the issue of how to price options remained dormant until the 1960s when Samuleson
(1965), whose interest was aroused by the rediscovered Bachelier dissertation, again considered
this problem. Uncharacteristically for Samuleson, he was unable to solve the problem. His
work and that of a number of others was lacking one final insight: assets that are subject to the
same risks must trade for the same price. If they do not, then the market would offer arbitrage
opportunities - something that cannot exist in a well-functioning market. The same concept was
used by two of Samuelson's MIT colleagues, Modigliani and Miller (1958), to prove their
famous capital structure proposition.

The breakthrough was achieved in the seminal Black and Scholes (1973) model. Black and
Scholes focus on valuing "European" options (these are options that allow exercise on only one
date), as distinct from "American" options (which have the added flexibility that the buyer can
calculate the option at any time up to the final exercise date). They develop a closed form
solution for the price of a European call option on a common stock. The underlying idea is that
an investor could exactly replicate the payoff of the option by trading at each point in time in
the stock and a riskless bond. This trading strategy should be self-financing: it should have an
initial cost, but then require no other cash inflows or outflows until the terminal date, when the payoff should exactly match the payoff of the derivative. For the market to be free from arbitrage opportunities, the cost of the replicating strategy must be the exact price of the option. Black and Scholes show that the option pricing formula must satisfy a partial differential equation with associated boundary conditions for which they are able to give a closed form solution. Moreover, they explicitly identify the trading or hedging strategy needed to replicate the option.

A startling result in the Black-Scholes analysis is that the expected rate of return of the underlying asset is completely absent from the pricing equation. Even Black and Scholes found it hard to provide a good intuition for this result, as can be seen by their comment that "The option value as a function of the stock price is independent of the expected return on the stock. The expected return on the option, however, will depend on the expected return on the stock. The faster the stock price rises, the faster the option price will rise through the functional relationship." The power of the Black-Scholes formula is that it must be true for every agent regardless of their preferences; all that is required is that at least one agent would not leave any arbitrage opportunities untouched. Shortly after the publication of the Black and Scholes formula, Merton (1973a) discussed many of the generic properties of options. His arguments, which were also based on various no-arbitrage arguments, helped to develop a deep understanding not only of European call options but of many others.

Since the Black-Scholes pricing formula holds for any agent, regardless of preferences, it must also hold for an agent who is risk-neutral. Cox and Ross (1976) were the first to use this argument to obtain the fundamental insight that the option price is the expected value of the payoff from the option, discounted at the risk-free rate. However, the probability distribution under which the expectation is taken is not the true probability distribution that investors have, but is one that is adjusted for risk. This probability distribution, known as the risk-neutral distribution, is the one that would be associated with risk-neutral investors; it equates today's price to the expected dividends and capital gains discounted at the risk-free rate.

The full ramifications of this result are explained by Harrison and Kreps (1979). Harrison and Kreps show that, under certain regularity conditions, the absence of arbitrage is equivalent to the existence of a risk-neutral probability distribution. They show that under the risk-neutral probability the price process of any asset discounted by the risk-free asset is a martingale. In mathematics a martingale, roughly speaking, is a random variable whose expected value next period is equal to its current value. Under the risk-neutral probability the expected return on any risky asset is equal to the risk-free rate, and hence investors cannot expect to make any more (or less) money on a risky assets than on the risk-free asset. For this reason Harrison and Kreps use the term martingale probability measure, and using the risk-neutral probability to price derivatives is sometimes known as martingale pricing.

The mathematical tools required by the Black-Scholes (1973) model and its extensions, as well as the general analysis of Harrison and Kreps (1979), are quite onerous. In order to facilitate the teaching of the ideas of option pricing by replication and risk-neutral probabilities, Sharpe (1978) constructed a simple binomial model. Sharpe assumes that over each time period a stock
could go up or down by some fixed percentage. With only two possible moves over each time period, the option can easily be replicated by trading in the stock and the riskless bond. Cox, Ross and Rubinstein (1979) took up this binomial approach to valuing call and put options. The resulting Cox-Ross-Rubinstein model is completely consistent with the Black-Scholes model; as the time steps are made smaller and smaller the value of the Cox-Ross-Rubinstein call converges to the Black-Scholes value. However, the Cox-Ross-Rubinstein model has the advantage that it can easily be adjusted to price other derivatives, such as American puts, which are considerably harder to evaluate in the Black-Scholes framework. This approach is immensely popular, not only in the classroom but also among practitioners.

A class of securities that has aroused considerable interest, from both academics and practitioners, is interest rate derivatives. There have been two generic approaches to pricing these derivatives, both of which differ from the equilibrium approach of Cox, Ingersoll and Ross (1985) described in section 4.3. The first and earliest approach was to take the spot interest rate and its dynamics as given. Vasicek (1977) took the spot rate as the underlying state variable and was able to derive no arbitrage restrictions on bond prices. Using a simple specification for the spot interest rate, he derived a closed form solution for the bond price. A second approach pioneered by Ho and Lee (1986) applies the concepts of option pricing more directly. They take the whole zero-coupon bond curve as given, and develop a binomial model for the dynamics of the term structure of interest rates. Over any time period, the term structure either steepens or flattens. Thus the whole complement of zero-coupon bonds are the underlying assets, and this allows Ho and Lee to derive no-arbitrage restrictions on the dynamics of the curve. The ideas of Ho and Lee are easily adaptable, and Heath, Jarrow and Morton (1992) use this approach to develop a general continuous-time framework for modelling interest rate derivatives.

Although the pricing theory was developed for options on common stock and other traded assets, as the nature of options became better understood it became clear that many other assets have option-like qualities. This suggests using option-pricing techniques to evaluate many aspects of real assets. Decisions are made in an uncertain environment, and there is usually some element of flexibility in using an asset. This means that traditional discounted cash flow analysis may be flawed, leading to myopic investment choices, undervaluation and underinvestment. Although Myers (1977) argues that flexibility in capital investment opportunities could be thought of as “growth options”, it was not until the mid-1980s that flexibility was explicitly analysed and priced using option valuation methodology.

The most obvious application to real investment decisions arises from the option inherent in developing, running and abandoning a natural resource such a mine. Traditional analysis would indicate that anyone wishing to undertake to develop a natural resource should do so whenever the net present value (NPV) of the project is positive; likewise, once the mine is in place it should be closed if the NPV becomes negative. Brennan and Schwartz (1985) show that, once the option to close the mine is taken into account the traditional analysis is flawed. They show that it is optimal to open the mine only if the NPV is larger than some (positive) value, and it is optimal to close the mine only when the NPV drops below some (negative) value. Brennan and Schwartz show that the value of the option to close the mine can be a substantial
part of its value. They also consider a general framework where a mine may not only be opened or closed, but may also be abandoned, an option which has similar properties.

Other forms of real options include the option to defer a project (McDonald and Siegel, 1986), the option to slow down the development of a project (Majd and Pindyck, 1987) and the option to abandon a project (Myers and Majd, 1990), to mention just a few. Although many of these options cannot be traded, it is now realised that many financial decisions faced by firms have option-like properties that must be taken into account. Thus the theory of derivatives has had an impact not just on the research community, but also on financial markets and businesses, with firms now being able to price and trade many different types of risk.

6. Conclusion

This paper has charted the historical development of asset pricing and derivative valuation. The roots of the subject were developed by Bernoulli (1738) who gave the first account of decision making under uncertainty. However, it was not until the twentieth century that general theories of portfolio selection and asset pricing were formulated and solved. The 1950s saw the formation of the ideas that would shape asset pricing. First, Arrow and Debreu (1954) developed their general equilibrium theory, which gave the insight that consumption in different future states could simply be viewed as different consumption goods. Second, Modigliani and Miller (1958) showed that the value of a firm is independent of the capital structure. This was the first occurrence of arguments based on the assumption of no arbitrage. Third, Markowitz' (1959) theory of portfolio optimisation showed how risk and return must be related in optimal portfolio construction.

Hard on the heels of this early work came a period of rapid growth in the understanding of how investors trade off risk and return, and how assets are priced. The work of Tobin (1958) and Sharpe (1963) not only greatly increased financial economists' understanding of portfolio selection, but also reduced the complexity of the portfolio choice problem. This research yielded insights into how and which risks should be managed. In turn, this raised the issue of how risk should be measured. Work by King (1966), Blume (1971) and Dimson (1979) explains how to measure risk, and more sophisticated measures of risk were introduced by Engle (1982).

It was only small leap from the portfolio choice problem to a model of asset pricing. Treynor (1961) and Sharpe (1964) argued that since we would expect all investors to make this tradeoff between risk and return, then in equilibrium high risk assets must compensate investors by offering higher returns. This gave rise to the capital asset pricing model, which has been the cornerstone of asset pricing. The CAPM tells us that only systematic risk, as measured by beta, is rewarded, and that the relationship between expected return and beta is linear. Black (1972) extended the CAPM to a world in which one cannot borrow at the riskless rate of interest.

After the development of the CAPM, research interest split into two directions: extending the CAPM and empirical investigation of the model. The latter was made possible by the
development of computer technology. Black, Jensen and Scholes (1972) were the first to test the CAPM rigorously using an innovative “two-pass” method, while Gibbons (1982) used classical maximum likelihood to find evidence against the CAPM. Unfortunately, as Roll (1977) shows, all these tests may be flawed because the market portfolio cannot be observed.

The major extension of the CAPM came when researchers developed models that price assets, not just over a fixed time period but intertemporally. Merton (1973) showed that the original CAPM would not in general hold in a dynamic environment and would require multiple betas. However, Breeden (1979) demonstrated that one beta would be sufficient intertemporally, provided the correct beta is measured by an asset's marginal contribution to consumption flow rather than to wealth. Multiperiod models also made it possible to develop models that endogenise interest rates and the yield curve, as shown by Cox, Ingersoll and Ross (1985).

An alternative to equilibrium pricing was to price using no-arbitrage arguments. Ross (1976) shows that in large asset markets the absence of arbitrage opportunities implies that, as in the CAPM, return is expected to be higher on assets that are riskier. Since this is based on an arbitrage argument, this model is named the arbitrage pricing theory. Empirically, the APT has been investigated by using either factor analytic techniques (Roll and Ross, 1980), or pre-specified macro-economic factory (Chen, Roll and Ross, 1986).

The final major strand of asset pricing is derivative valuation. Black and Scholes (1973) were the first to work out how to price European call options. Their argument is based on no-arbitrage arguments, and shows how to replicate the option payoff by trading in the underlying assets. Harrison and Kreps (1979) show that another way to price derivatives is to find a risk-neutral probability, and that this probability exists if and only if there are no arbitrage opportunities. A simple alternative to the mathematical techniques of Black and Scholes was offered by Cox, Ross and Rubinstein (1979), who explain how to value options using binomial techniques. Option pricing has now been applied to many different types of assets: Ho and Lee (1986) show how option pricing theory can be applied to interest rate derivatives, while Brennan and Schwartz (1985) extend the theory to price real options.

This paper has illustrated how asset pricing and derivative valuation has mushroomed from a small number of key papers. We have seen that the key ideas of underlying the seminal works on asset pricing have been mean-variance portfolio optimisation, equilibrium analysis and no-arbitrage arguments. This field of research has had a profound impact not only on economic science, but also on financial markets, institutions and businesses all over the world.

References*

Aristotle, The Republic


* In bold typeface we identify articles to be published in Volume II of Elroy Dimson and Massoud Mussavian (eds) Foundations of Finance, Dartmouth Publishing Company, 1999, forthcoming. Further articles, identified by italicised bold typeface, are included in Volumes I and III.


Hicks, J R (1934). *Value and Capital*. Oxford University Press.


Myers, Stewart C. and Saman Majd (1990). Abandonment value and project life in Frank Fabozzi (ed.) *Advances in Futures and Options*


