CHAPTER 4

HOW DO WE MEASURE RISK?

If you accept the argument that risk matters and that it affects how managers and investors make decisions, it follows logically that measuring risk is a critical first step towards managing it. In this chapter, we look at how risk measures have evolved over time, from a fatalistic acceptance of bad outcomes to probabilistic measures that allow us to begin getting a handle on risk, and the logical extension of these measures into insurance. We then consider how the advent and growth of markets for financial assets has influenced the development of risk measures. Finally, we build on modern portfolio theory to derive unique measures of risk and explain why they might be not in accordance with probabilistic risk measures.

Fate and Divine Providence

Risk and uncertainty have been part and parcel of human activity since its beginnings, but they have not always been labeled as such. For much of recorded time, events with negative consequences were attributed to divine providence or to the supernatural. The responses to risk under these circumstances were prayer, sacrifice (often of innocents) and an acceptance of whatever fate meted out. If the Gods intervened on our behalf, we got positive outcomes and if they did not, we suffered; sacrifice, on the other hand, appeased the spirits that caused bad outcomes. No measure of risk was therefore considered necessary because everything that happened was pre-destined and driven by forces outside our control.

This is not to suggest that the ancient civilizations, be they Greek, Roman or Chinese, were completely unaware of probabilities and the quantification of risk. Games of chance were common in those times and the players of those games must have recognized that there was an order to the uncertainty.[1] As Peter Bernstein notes in his splendid book on the history of risk, it is a mystery why the Greeks, with their considerable skills at geometry and numbers, never seriously attempted to measure the likelihood of uncertain events, be they storms or droughts, occurring, turning instead to priests and fortune tellers.[2]

Notwithstanding the advances over the last few centuries and our shift to more modern, sophisticated ways of analyzing uncertainty, the belief that powerful forces beyond our reach shape our destinies is never far below the surface. The same traders who use sophisticated computer models to measure risk consult their astrological charts and rediscover religion when confronted with the possibility of large losses.

Estimating Probabilities: The First Step to Quantifying Risk

Given the focus on fate and divine providence that characterized the way we thought about risk until the Middle Ages, it is ironic then that it was an Italian monk, who initiated the discussion of risk measures by posing a puzzle in 1494 that befuddled people for almost two centuries. The solution to his puzzle and subsequent developments laid the foundations for modern risk measures.

Luca Pacioli, a monk in the Franciscan order, was a man of many talents. He is credited with inventing double entry bookkeeping and teaching Leonardo DaVinci mathematics. He also wrote a book on mathematics, *Summa de Arithmetica*, that summarized all the knowledge in mathematics at that point in time. In the book, he also presented a puzzle that challenged mathematicians of the time. Assume, he said, that two gamblers are playing a best of five dice game and are interrupted after three games, with one gambler leading two to one. What is the fairest way to split the pot between the two gamblers, assuming that the game cannot be resumed but taking into account the state of the game when it was interrupted?

With the hindsight of several centuries, the answer may seem simple but we have to remember that the notion of making predictions or estimating probabilities had not developed yet. The first steps towards solving the Pacioli Puzzle came in the early part of

---

[1] Chances are…. Adventures in Probability, 2006, Kaplan, M. and E. Kaplan, Viking Books, New York. The authors note that dice litter ancient Roman campsites and that the citizens of the day played a variant of craps using either dice or knucklebones of sheep.

[2] Much of the history recounted in this chapter is stated much more lucidly and in greater detail by Peter Bernstein in his books “Against the Gods: The Remarkable Story of Risk” (1996) and “Capital Ideas: The Improbable Origins of Modern Wall Street (1992). The former explains the evolution of our thinking on risk through the ages whereas the latter examines the development of modern portfolio theory.

the sixteenth century when an Italian doctor and gambler, Girolamo Cardano, estimated the likelihood of different outcomes of rolling a dice. His observations were contained in a book titled “Books on the Game of Chance”, where he estimated not only the likelihood of rolling a specific number on a dice (1/6), but also the likelihood of obtaining values on two consecutive rolls; he, for instance, estimated the probability of rolling two ones in a row to be 1/36. Galileo, taking a break from discovering the galaxies, came to the same conclusions for his patron, the Grand Duke of Tuscany, but did not go much further than explaining the roll of the dice.

It was not until 1654 that the Pacioli puzzle was fully solved when Blaise Pascal and Pierre de Fermat exchanged a series of five letters on the puzzle. In these letters, Pascal and Fermat considered all the possible outcomes to the Pacioli puzzle and noted that with a fair dice, the gambler who was ahead two games to one in a best-of-five dice game would prevail three times out of four, if the game were completed, and was thus entitled to three quarters of the pot. In the process, they established the foundations of probabilities and their usefulness not just in explaining the past but also in predicting the future. It was in response to this challenge that Pascal developed his triangle of numbers for equal odds games, shown in figure 4.1:[3]

[3] It should be noted that Chinese mathematicians constructed the same triangle five hundred years before Pascal and are seldom credited for the discovery.

Figure 4.1: Pascal's Triangle

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

... ...

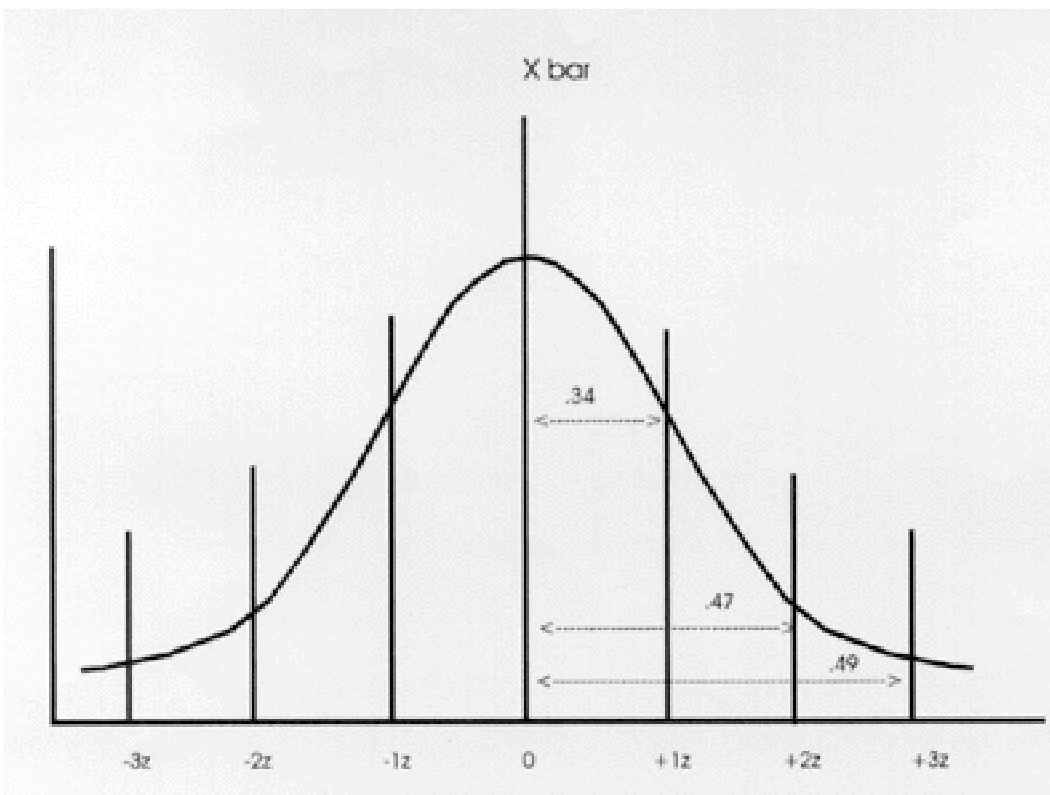
Pascal's triangle can be used to compute the likelihood of any event with even odds occurring. Consider, for instance, the odds that a couple expecting their first child will have a boy; the answer, with even odds, is one-half and is in the second line of Pascal's triangle. If they have two children, what are the odds of them having two boys, or a boy and a girl or two girls? The answer is in the second line, with the odds being ¼ on the first and the third combinations and ½ on the second. In general, Pascal's triangle provides the number of possible combination if an even-odds event is repeated a fixed number of times; if repeated N times, adding the numbers in the N+1 row and dividing each number by this total should yield the probabilities. Thus, the couple that has six children can compute the probabilities of the various outcomes by going to the seventh row and adding up the numbers (which yields 64) and dividing each number by the total. There is only a 1/64 chance that this couple will have six boys (or six girls), a 6/64 chance of having five boys and a girl (or five girls and a boy) and so on.

Sampling, The Normal Distributions and Updating

Pascal and Fermat fired the opening volley in the discussion of probabilities with their solution to the Pacioli Puzzle, but the muscle power for using probabilities was provided by Jacob Bernoulli, with his discovery of the law of large numbers. Bernoulli proved that a random sampling of items from a population has the same characteristics, on average, as the population.[4] He used coin flips to illustrate his point by noting that the proportion of heads (and tails) approached 50% as the number of coin tosses increased. In the process, he laid the foundation for generalizing population properties from samples, a practice that now permeates both the social and economic sciences.

The introduction of the normal distribution by Abraham de Moivre, an English mathematician of French extraction, in 1738 as an approximation for binomial distributions as sample sizes became larger, provided researchers with a critical tool for linking sample statistics with probability statements. [5] Figure 4.2 provides a picture of the normal distribution.

Figure 4.2: Normal Distribution



[4] Since Bernoulli's exposition of the law of large numbers, two variants of it have developed in the statistical literature. The weak law of large numbers states that average of a sequence of uncorrelated random numbers drawn from a distribution with the same mean and standard deviation will converge on the population average. The strong law of large numbers extends this formulation to a set of random variables that are independent and identically distributed (i.i.d)

The bell curve, that characterizes the normal distribution, was refined by other mathematicians, including Laplace and Gauss, and the distribution is still referred to as the Gaussian distribution. One of the advantages of the normal distribution is that it can be described with just two parameters – the mean and the standard deviation – and allows us to make probabilistic statements about sampling averages. In the normal distribution, approximately 68% of the distribution in within one standard deviation of the mean, 95% is within two standard deviations and 98% within three standard deviations. In fact, the distribution of a sum of independent variables approaches a normal distribution, which is the basis for the central limit theorem and allows us to use the normal distribution as an approximation for other distributions (such as the binomial).

In 1763, Reverend Thomas Bayes published a simple way of updating existing beliefs in the light of new evidence. In Bayesian statistics, the existing beliefs are called prior probabilities and the revised values after considering the new evidence are called posterior or conditional probabilities.[6] Bayes provided a powerful tool for researchers who wanted to use probabilities to assess the likelihood of negative outcomes, and to update these probabilities as events unfolded. In addition, Bayes' rule allows us to start with subjective judgments about the likelihood of events occurring and to modify these judgments as new data or information is made available about these events.

In summary, these developments allowed researchers to see that they could extend the practice of estimating probabilities from simple equal-odds events such as rolling a dice to any events that had uncertainty associated with it. The law of large numbers showed that sampling means could be used to approximate population averages, with the precision increasing with sample size. The normal distribution allows us to make probability statements about the sample mean. Finally, Bayes' rule allows us to estimate probabilities and revise them based on new sampling data.

---

[5] De Moivre, A., 1738, Doctrine of Chances.

[6] Bayes, Rev. T., "*An Essay Toward Solving a Problem in the Doctrine of Chances*", **Philos. Trans. R. Soc. London 53**, pp. 370-418 (1763); reprinted in **Biometrika 45**, pp. 293-315 (1958).

The Use of Data: Life Tables and Estimates

The work done on probability, sampling theory and the normal distribution provided a logical foundation for the analysis of raw data. In 1662, John Graunt created one of the first mortality tables by counting for every one hundred children born in London, each year from 1603 to 1661, how many were still living. In the course of constructing the table, Graunt used not only refined the use of statistical tools and measures with large samples but also considered ways of dealing with data errors. He estimated that while 64 out of every 100 made it age 6 alive, only 1 in 100 survived to be 76. In an interesting aside, Graunt estimated the population of London in 1663 to be only 384,000, well below the then prevailing estimate of six to seven million. He was eventually proved right, and London's population did not exceed 6 million until three centuries later. In 1693, Edmund Halley, the British mathematician, constructed the first life table from observations and also devised a method for valuing life annuities. He pointed out that the government, that was selling life annuities to citizens at that time, was pricing them too low and was not setting the price independently of the age of the annuitant.

Actuarial risk measures have become more sophisticated over time, and draw heavily on advances in statistics and data analysis, but the foundations still lies in the work done by Graunt and Halley. Using historical data, actuaries estimate the likelihood of events occurring – from hurricanes in Florida to deaths from cancer – and the consequent losses.

The Insurance View of Risk

As long as risk has existed, people have been trying to protect themselves against its consequences. As early as 1000 BC, the Babylonians developed a system where merchants who borrowed money to fund shipments could pay an extra amount to cancel the loan if the shipment was stolen. The Greeks and the Romans initiated life insurance with "benevolent societies" which cared for families of society members, if they died. However, the development of the insurance business was stymied by the absence of ways of measuring risk exposure. The advances in assessing probabilities and the subsequent development of statistical measures of risk laid the basis for the modern insurance business. In the aftermath of the great fire of London in 1666, Nicholas Barbon opened "The Fire Office", the first fire insurance company to insure brick homes. Lloyd's of London became the first the first large company to offer insurance to ship owners.

Insurance is offered when the timing or occurrence of a loss is unpredictable, but the likelihood and magnitude of the loss are relatively predictable. It is in the latter pursuit that probabilities and statistics contributed mightily. Consider, for instance, how a company can insure your house against fire. Historical data on fires can be used to assess the likelihood that your house will catch fire and the extent of the losses, if a fire occurs. Thus, the insurance company can get a sense of the expected loss from the fire and charge an insurance premium that exceeds that cost, thus earning a profit. By insuring a large number of houses against fire, they are drawing on Bernoulli's law of large numbers to ensure that their profits exceed the expected losses over time.

Even large, well-funded insurance companies have to worry, though, about catastrophes so large that they will be unable to meet their obligations. Katrina, one of the most destructive hurricanes in memory, destroyed much of New Orleans in 2005 and left two states, Louisiana and Mississipi, in complete devastation; the total cost of damages was in excess of $ 50 billion. Insurance companies paid out billions of dollars in claims, but none of the firms were put in serious financial jeopardy because of the practice of reinsuring, where insurance companies reduce their exposure to catastrophic risk through reinsurance.

Since insurers are concerned primarily about losses (and covering those losses), insurance measures of risk are almost always focused on the downside. Thus, a company that insures merchant ships will measure risk in terms of the likelihood of ships and cargo being damaged and the loss that accrues from the damage. The potential for upside that exists has little or no relevance to the insurer since he does not share in it.

Financial Assets and the Advent of Statistical Risk Measures

As stock and bond markets developed around the world in the nineteenth century, investors started looking for richer measures of risk. In particular, since investors in financial assets share in both upside and downside, the notion of risk primarily as a loss function (the insurance view) was replaced by a sense that risk could be a source of profit.

There was little access to information and few ways of processing even that limited information in the eighteenth and nineteenth centuries. Not surprisingly, the risk measures used were qualitative and broad. Investors in the financial markets during that period defined risk in terms of stability of income from their investments in the long term and capital preservation. Thus, perpetual British government bonds called Consols, that offered fixed coupons forever were considered close to risk free, and a fixed rate long term bond was considered preferable to a shorter term bond with a higher rate. In the risk hierarchy of that period, long term government bonds ranked as safest, followed by corporate bonds and stocks paying dividends and at the bottom were non-dividend paying stocks, a ranking that has not changed much since.

Given that there were few quantitative measures of risk for financial assets, how did investors measure and manage risk? One way was to treat entire groups of investments as sharing the same risk level; thus stocks were categorized as risky and inappropriate investments for risk averse investors, no matter what their dividend yield. The other was to categorize investments based upon how much information was available about the entity issuing it. Thus, equity issued by a well-established company with a solid reputation was considered safer than equity issued by a more recently formed entity about which less was known. In response, companies started providing more data on operations and making them available to potential investors.

By the early part of the twentieth century, services were already starting to collect return and price data on individual securities and computing basic statistics such as the expected return and standard deviation in returns. For instance, the Financial Review of Reviews, a British publication, examined portfolios of ten securities including bonds, preferred stock and ordinary stock in 1909, and measured the volatility of each security using prices over the prior ten years. In fact, they made an argument for diversification by estimating the impact of correlation on their hypothetical portfolios. (Appendix 1 includes the table from the publication). Nine years previously, Louis Bachelier, a post-graduate student of mathematics at the Sorbonne, examined the behavior of stock and option prices over time in a remarkable thesis. He noted that there was little correlation between the price change in one period and the price change in the next, thus laying the foundation for the random walk and efficient market hypothesis, though they were not fleshed out until almost sixty years later.[7]

At about the same time, the access to and the reliability of financial reports from corporations were improving and analysts were constructing risk measures that were based upon accounting numbers. Ratios of profitability (such as margin and return on capital) and financial leverage (debt to capital) were used to measure risk. By 1915, services including the Standard Statistics Bureau (the precursor to Standard and Poor's), Fitch and Moody's were processing accounting information to provide bond ratings as measures of credit risk in companies. Similar measures were slower to evolve for equities but stock rating services were beginning to make their presence felt well before the Second World War. While these services did not exhibit any consensus on the right way to measure risk, the risk measures drew on both price volatility and accounting information.

In his first edition of Security Analysis in 1934, Ben Graham argued against measures of risk based upon past prices (such as volatility), noting that price declines can be temporary and not reflective of a company's true value. He argued that risk comes from paying too high a price for a security, relative to its value and that investors should maintain a "margin of safety" by buying securities for less than their true worth.[8] This is an argument that value investors from the Graham school, including Warren Buffett, continue to make to this day.

By 1950, investors in financial markets were using measures of risk based upon past prices and accounting information, in conjunction with broad risk categories, based upon security type and issuer reputation, to make judgments about risk. There was,

---

[7] Bachelier, L., 1900, Theorie De La Speculation, *Annales Scientifiques de l'E´cole Normale Supe´rieure*,1900, pp.21–86. For an analysis of this paper's contribution to mathematical finance, see Courtault, J.M., Y. Kabanov, B. Bru and P. Crepel, 2000, Louis Bachelier: On the Centenary of the Theorie De La Speculation, Mathematical Finance, v10, 341-350.

[8] Graham, B., 1949, The Intelligent Investor; Graham, B. and D. Dodd, 1934, Security Analysis, Reprint by McGraw Hill. In "Intelligent Investor", Graham proposed to measure the margin of safety by looking at the difference between the earnings yield on a stock (Earnings per share/ Market price) to the treasury bond rate; the larger the difference (with the former exceeding the latter), the greater the margin for safety.

however, no consensus on how best to measure risk and the exact relationship between risk and expected return.

The Markowitz Revolution

The belief that diversification was beneficial to investors was already well in place before Harry Markowitz turned his attention to it in 1952. In fact, our earlier excerpt from the Financial Review of Reviews from 1909 used correlations between securities to make the argument that investors should spread their bets and that a diversified portfolio would be less risky than investing in an individual security, while generating similar returns. However, Markowitz changed the way we think about risk by linking the risk of a portfolio to the co-movement between individual assets in that portfolio.

Efficient Portfolios

As a young graduate student at the University of Chicago in the 1940s, Harry Markowitz was influenced by the work done by Von Neumann, Friedman and Savage on uncertainty. In describing how he came up with the idea that gave rise to modern portfolio theory, Markowitz explains that he was reading John Burr Williams "Theory of Investment Value", the book that first put forth the idea that the value of a stock is the present value of its expected dividends.[9] He noted that if the value of a stock is the present value of its expected dividends and an investor were intent on only maximizing returns, he or she would invest in the one stock that had the highest expected dividends, a practice that was clearly at odds with both practice and theory at that time, which recommended investing in diversified portfolios. Investors, he reasoned, must diversify because they care about risk, and the risk of a diversified portfolio must therefore be lower than the risk of the individual securities that went into it. His key insight was that the variance of a portfolio could be written as a function not only of how much was invested in each security and the variances of the individual securities but also of the correlation between the securities. By explicitly relating the variance of a portfolio to the covariances between individual securities, Markowitz not only put into concrete form what had been conventional wisdom for decades but also formulated a process by which investors could generate optimally diversified portfolios, i.e., portfolios that would maximize returns for any given level of risk (or minimize risk for any given level of return). In his thesis, he derived the set of optimal portfolios for different levels of risk and called it the efficient frontier.[10] He refined the process in a subsequent book that he wrote while he worked at the RAND corporation.[11]

The Mean-Variance Framework

The Markowitz approach, while powerful and simple, boils investor choices down to two dimensions. The "good" dimension is captured in the expected return on an investment and the "bad" dimension is the variance or volatility in that return. In effect, the approach assumes that all risk is captured in the variance of returns on an investment and that all other risk measures, including the accounting ratios and the Graham margin of safety, are redundant. There are two ways in which you can justify the mean-variance focus: one is to assume that returns are normally distributed and the other is to assume that investors' utility functions push them to focus on just expected return and variance.

Consider first the "normal distribution" assumption. As we noted earlier in this chapter, the normal distribution is not only symmetric but can be characterized by just the mean and the variance.[12] If returns were normally distributed, it follows then that the only two choice variables for investors would be the expected returns and standard deviations, thus providing the basis for the mean variance framework. The problem with this assumption is that returns on most investments cannot be normally distributed. The worst outcome you can have when investing in a stock is to lose your entire investment, translating into a return of -100% (and not $-\infty$ as required in a normal distribution).

---

[9] See the Markowitz autobiography for the Nobel committee. It can be accessed online at http://nobelprize.org/economics/laureates/1990/markowitz-autobio.html.

[10] Markowitz, H.M. 1952. "Portfolio Selection," The ***Journal of Finance, 7(l): 77-91.***

[11] Markowitz, H.M. 1959. ***Portfolio Selection: Efficient Diversification of Investments. New*** York: Wiley (Yale University Press, 1970, Basil Blackwell, 1991).

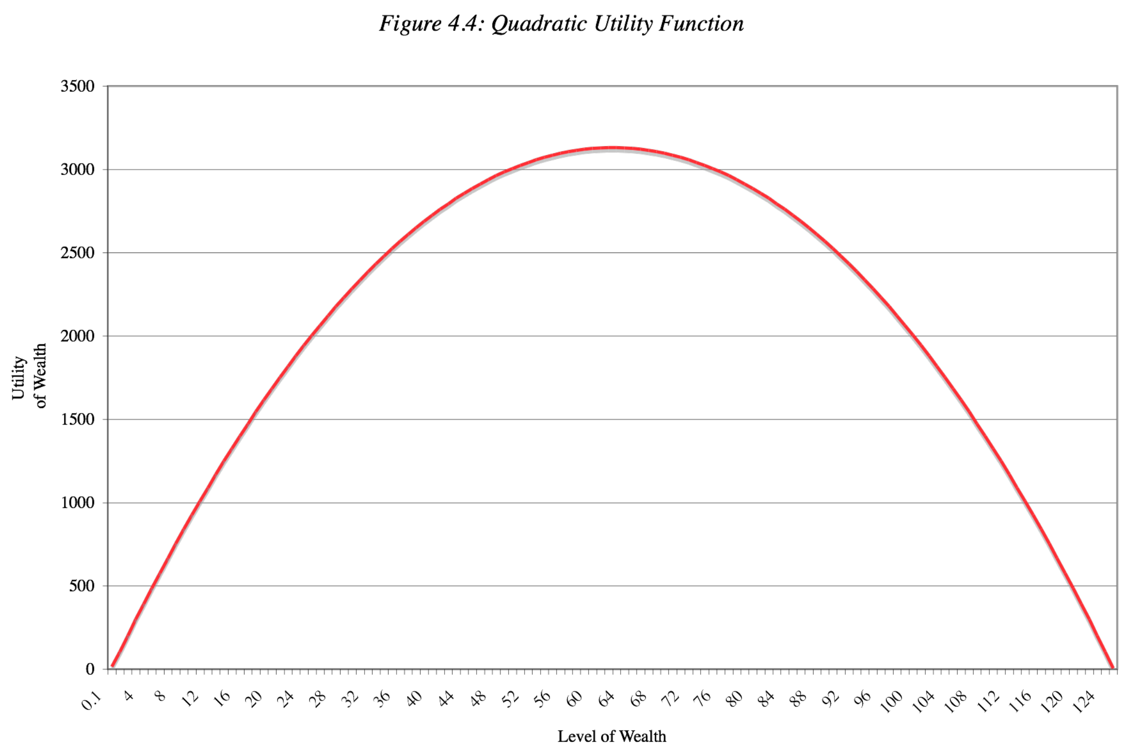
[12] Portfolios of assets that each exhibit normally distributed returns will also be normally distributed. Lognormally distributed returns can also be parameterized with the mean and the variance, but portfolios of assets exhibiting lognormal returns may not exhibit lognormality.

As for the "utility distribution" argument, consider the quadratic utility function, where utility is written as follows:

$$U(W) = a + bW - cW^2$$

The quadratic utility function is graphed out in figure 4.3:

Figure 4.3: Quadratic Utility Function



Investors with quadratic utility functions care about only the level of their wealth and the variance in that level and thus have a mean-variance focus when picking investments. While assuming a quadratic utility function may be convenient, it is not a plausible measure of investor utility for three reasons. The first is that it assumes that investors are equally averse to deviations of wealth below the mean as they are to deviations above the mean. The second is that individuals with quadratic utility functions exhibit decreasing absolute risk aversion, i.e., individuals invest less of their wealth (in absolute terms) in risky assets as they become wealthier. Finally, there are ranges of wealth where investors actually prefer less wealth to more wealth; the marginal utility of wealth becomes negative.

Since both the normal distribution and quadratic utility assumptions can only be justified with contorted reasoning, how then how do you defend the mean-variance approach? The many supporters of the approach argue that the decisions based upon decisions based upon the mean and the variance come reasonably close to the optimum with utility functions other than the quadratic. They also rationalize the use of the normal distribution by pointing out that returns may be log-normally distributed (in which case the log of the returns should be normally distributed) and that the returns on portfolios (rather than individual stocks), especially over shorter time periods, are more symmetric and thus closer to normality. Ultimately, their main argument is that what is lost in precision (in terms of using a more realistic model that looks at more than expected returns and variances) is gained in simplicity.[13]

Implications for Risk Assessment

If we accept the mean-variance framework, the implications for risk measurement are significant.

- The argument for diversification becomes irrefutable. A portfolio of assets will almost always generate a higher return, for any given level of variance, than any single asset. Investors should diversity even if they have special access to information and there are transactions costs, though the extent of diversification may be limited.[14]
- In general, the risk of an asset can be measured by the risk it adds on to the portfolio that it becomes part of and in particular, by how much it increases the variance of the portfolio to which it is added. Thus, the key component determining asset risk will not be its volatility per se, but how the asset price co-moves with the portfolio. An asset that is extremely volatile but moves independently of the rest of the assets in a portfolio will add little or even no risk to the portfolio. Mathematically, the

---

[13] Markowitz, defending the quadratic utility assumptions, notes that focusing on just the mean and the variance makes sense for changes

[14] The only exception is if the information is perfect, i.e., investors have complete certainty about what will happen to a stock or investment. In that case, they can invest their wealth in that individual asset and it will be riskfree. In the real world, inside information gives you an edge over other investors but does not bestow its possessor with guaranteed profits. Investors with such information would be better served spreading their wealth over multiple stocks on which they have privileged information rather than just one.

covariance between the asset and the other assets in the portfolio becomes the dominant risk measure, rather than its variance.

- The other parameters of an investment, such as the potential for large payoffs and the likelihood of price jumps, become irrelevant once they have been factored into the variance computation.

Whether one accepts the premise of the mean-variance framework or not, its introduction changed the way we think about risk from one where the risk of individual assets was assessed independently to one where asset risk is assessed relative to a portfolio of which the asset is a part.

Introducing the Riskless Asset – The Capital Asset Pricing Model (CAPM) arrives

The revolution initiated by Harry Markowitz was carried to its logical conclusion by John Lintner, Jack Treynor and Bill Sharpe, with their development of the capital asset pricing model (CAPM).[15] Sharpe and Linter added a riskless asset to the mix and concluded that there existed a superior alternative to investors at every risk level, created by combining the riskless asset with one specific portfolio on the efficient frontier. Combinations of the riskless asset and the one super-efficient portfolio generate higher expected returns for every given level of risk than holding just a portfolio of risky assets. (Appendix 2 contains a more complete proof of this conclusion) For those investors who desire less risk than that embedded in the market portfolio, this translates into investing a portion of their wealth in the super-efficient portfolio and the rest in the riskless assets. Investors who want to take more risk are assumed to borrow at the riskless rate and invest that money in the super-efficient portfolio. If investors follow this dictum, all investors should hold the one super-efficient portfolio, which should be supremely diversified, i.e., it should include every traded asset in the market, held in proportion to its market value. Thus, it is termed the market portfolio.

To reach this result, the original version of the model did assume that there were no transactions costs or taxes and that investors had identical information about assets

[15] Sharpe, William F., 1961,. Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance, 19 (3), 425-442; Lintner, J., 1965 The valuation of risk assets and the selection of risky

(and thus shared the same estimates for the expected returns, standard deviations and correlation across assets). In addition, the model assumed that all investors shared a single period time horizon and that they could borrow and invest at the riskfree rate. Intuitively, the model eliminates any rationale for holding back on diversification. After all, without transactions costs and differential information, why settle for any portfolio which is less than fully diversified? Consequently, any investor who holds a portfolio other than the market portfolio is not fully diversified and bears the related cost with no offsetting benefit.

If we accept the assumptions (unrealistic though they may seem) of the capital asset pricing model, the risk of an individual asset becomes the risk added on to the market portfolio and can be measured statistically as follows:

$$\text{Risk of an asset} = \frac{\text{Covariance of asset with the market portfolio}}{\text{Variance of the maraket portfolio}} = \text{Asset Beta}$$

Thus, the CAPM extends the Markowitz insight about risk added to a portfolio by an individual asset to the special case where all investors hold the same fully diversified market portfolio. Thus, the risk of any asset is a function of how it covaries with the market portfolio. Dividing the covariance of every asset by the market portfolio to the market variance allows for the scaling of betas around one; an average risk investment has a beta around one, whereas investments with above average risk and below average risk have betas greater than and less than one respectively.

In closing, though, accepting the CAPM requires us to accept the assumptions that the model makes about transactions costs and information but also the underlying assumptions of the mean-variance framework. Notwithstanding its many critics, whose views we will examine in the next two sections, the widespread acceptance of the model and its survival as the default model for risk to this day is testimony to its intuitive appeal and simplicity.

---

investments in stock portfolios and capital budgets, Review of Economics and Statistics, 47: 13-37; Treynor, Jack (1961). Towards a theory of market value of risky assets, unpublished manuscript.

Mean Variance Challenged

From its very beginnings, the mean variance framework has been controversial. While there have been many who have challenged its applicability, we will consider these challenges in three groups. The first group argues that stock prices, in particular, and investment returns, in general, exhibit too many large values to be drawn from a normal distribution. They argue that the "fat tails" on stock price distributions lend themselves better to a class of distributions, called power law distributions, which exhibit infinite variance and long periods of price dependence. The second group takes issue with the symmetry of the normal distribution and argues for measures that incorporate the asymmetry observed in actual return distributions into risk measures. The third group posits that distributions that allow for price jumps are more realistic and that risk measures should consider the likelihood and magnitude of price jumps.

Fat Tails and Power Law Distributions

Benoit Mandelbrot, a mathematician who also did pioneering work on the behavior of stock prices, was one of those who took issue with the use of normal and lognormal distributions.[16] He argued, based on his observation of stock and real asset prices, that a power-law distribution characterized them better.[17] In a power-law distribution, the relationship between two variables, Y and X can be written as follows:

$$Y = \alpha^k$$

In this equation, α is a constant (constant of proportionality) and k is the power law exponent. Mandelbrots key point was that the normal and log normal distributions were best suited for series that exhibited mild and well behaved randomness, whereas power law distributions were more suited for series which exhibited large movements and what

---

[16] Mandelbrot, B., 1961, The Variation of Certain Speculative Prices, Journal of Business, v34, 394-419.

[17] H.E. Hurst, a British civil servant, is credited with bringing the power law distribution into popular usage. Faced with the task of protecting Egypt against floods on the Nile rive, he did an exhaustive analysis of the frequency of high and low water marks at dozens of other rivers around the world. He found that the range widened far more than would be predicted by the normal distribution. In fact, he devised a measure, called the Hurst exponent, to capture the widening of the range; the Hurst exponent which has a value of 0.5 for the normal distribution had a value of 0.73 for the rivers that he studied. In intuitive terms, his findings suggested that there were extended periods of rainfall that were better-than-expected and worse-than-expected that caused the widening of the ranges. Mandelbrot's awareness of this research allowed him to bring the same thinking into his analysis of cotton prices on the Commodity Exchange.

he termed "wild randomness". Wild randomness occurs when a single observation can affect the population in a disproportionate way. Stock and commodity prices, with their long periods of relatively small movements, punctuated by wild swings in both directions, seem to fit better into the "wild randomness" group.

What are the consequences for risk measures? If asset prices follow power law distributions, the standard deviation or volatility ceases to be a good risk measure and a good basis for computing probabilities. Assume, for instance, that the standard deviation in annual stock returns is 15% and that the average return is 10%. Using the normal distribution as the base for probability predictions, this will imply that the stock returns will exceed 40% (average plus two standard deviations) only once every 44 years and 55% only (average plus three standard deviations) once every 740 years. In fact, stock returns will be greater than 85% (average plus five standard deviations) only once every 3.5 million years. In reality, stock returns exceed these values far more frequently, a finding consistent with power law distributions, where the probability of larger values decline linearly as a function of the power law exponent. As the value gets doubled, the probability of its occurrence drops by the square of the exponent. Thus, if the exponent in the distribution is 2, the likelihood of returns of 25%, 50% and 100% can be computed as follows:

Returns will exceed 25%: Once every 6 years

Returns will exceed 50%: Once every 24 years

Returns will exceed 100%: Once every 96 years

Note that as the returns get doubled, the likelihood increases four-fold (the square of the exponent). As the exponent decreases, the likelihood of larger values increases; an exponent between 0 and 2 will yield extreme values more often than a normal distribution. An exponent between 1 and 2 yields power law distributions called stable Paretian distributions, which have infinite variance. In an early study, Fama[18] estimated the exponent for stocks to be between 1.7 and 1.9, but subsequent studies have found that the exponent is higher in both equity and currency markets.[19]

---

[18] Fama, E.F., 1965, The Behavior of Stock Market Prices, Journal of Business, v38, 34-105.

[19] In a paper in "Nature", researchers looked at stock prices on 500 stocks between 1929 and 1987and concluded that the exponent for stock returns is roughly 3. Gabaix, X., Gopikrishnan, P., Plerou, V. &

In practical terms, the power law proponents argue that using measures such as volatility (and its derivatives such as beta) under estimate the risk of large movements. The power law exponents for assets, in their view, provide investors with more realistic risk measures for these assets. Assets with higher exponents are less risky (since extreme values become less common) than asset with lower exponents.

Mandelbrot's challenge to the normal distribution was more than a procedural one. Mandelbrot's world, in contrast to the Gaussian mean-variance one, is one where prices move jaggedly over time and look like they have no pattern at a distance, but where patterns repeat themselves, when observed closely. In the 1970s, Mandelbrot created a branch of mathematics called "fractal geometry" where processes are not described by conventional statistical or mathematical measures but by fractals; a fractal is a geometric shape that when broken down into smaller parts replicates that shape. To illustrate the concept, he uses the example of the coastline that, from a distance, looks irregular but up close looks roughly the same – fractal patterns repeat themselves. In fractal geometry, higher fractal dimensions translate into more jagged shapes; the rugged Cornish Coastline has a fractal dimension of 1.25 whereas the much smoother South African coastline has a fractal dimension of 1.02. Using the same reasoning, stock prices that look random, when observed at longer time intervals, start revealing self-repeating patterns, when observed over shorter time periods. More volatile stocks score higher on measures of fractal dimension, thus making it a measure of risk. With fractal geometry, Mandelbrot was able to explain not only the higher frequency of price jumps (relative to the normal distribution) but also long periods where prices move in the same direction and the resulting price bubbles.[20]

Asymmetric Distributions

Intuitively, it should be downside risk that concerns us and not upside risk. In other words, it is not investments that go up significantly that create heartburn and unease but investments that go down significantly. The mean-variance framework, by weighting

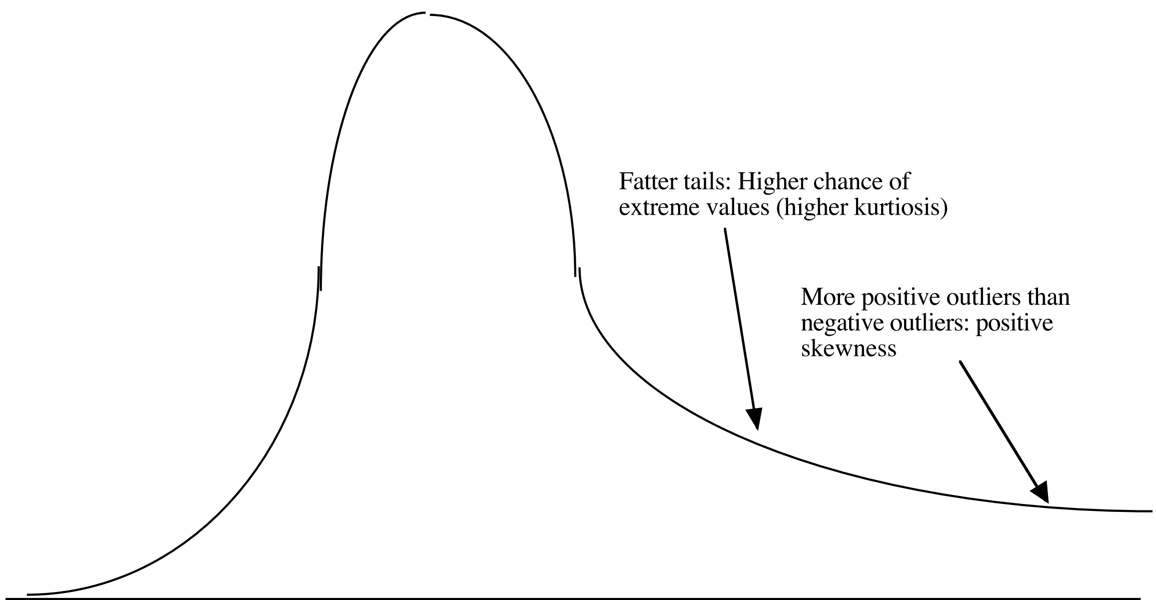
---

Stanley, H.E., 2003, A theory of power law distributions in financial market fluctuations. Nature 423, 267-70.

both upside volatility and downside movements equally, does not distinguish between the two. With a normal or any other symmetric distribution, the distinction between upside and downside risk is irrelevant because the risks are equivalent. With asymmetric distributions, though, there can be a difference between upside and downside risk. As we noted in chapter 3, studies of risk aversion in humans conclude that (a) they are loss averse, i.e., they weigh the pain of a loss more than the joy of an equivalent gain and (b) they value very large positive payoffs – long shots – far more than they should given the likelihood of these payoffs.

In practice, return distributions for stocks and most other assets are not symmetric. Instead, as shown in figure 4.4, asset returns exhibit fat tails and are more likely to have extreme positive values than extreme negative values (simply because returns are constrained to be no less than -100%).

Figure 4.4: Return distributions on Stocks



Note that the distribution of stock returns has a higher incidence of extreme returns (fat tails or kurtosis) and a tilt towards very large positive returns (positive skewness). Critics of the mean variance approach argue that it takes too narrow a view of both rewards and risk. In their view, a fuller return measure should consider not just the magnitude of

[20] Mandelbrot has expanded on his thesis in a book on the topic: Mandelbrot, B. and R.L. Hudson, 2004, The (Mis)behavior of Markets: A Fractal View of Risk, Ruin and Reward, Basic Books.

expected returns but also the likelihood of very large positive returns or skewness[21] and more complete risk measure should incorporate both variance and possibility of big jumps (co-kurtosis).[22] Note that even as these approaches deviate from the mean-variance approach in terms of how they define risk, they stay true to the portfolio measure of risk. In other words, it is not the possibility of large positive payoffs (skewness) or big jumps (kurtosis) that they argue should be considered, but only that portion of the skewness (co-skewness) and kurtosis (co-kurtosis) that is market related and not diversifiable.

Jump Process Models

The normal, power law and asymmetric distributions that form the basis for the models we have discussed in this section are all continuous distributions. Observing the reality that stock prices do jump, there are some who have argued for the use of jump process distributions to derive risk measures.

Press, in one of the earliest papers that attempted to model stock price jumps, argued that stock prices follow a combination of a continuous price distribution and a Poisson distribution, where prices jump at irregular intervals. The key parameters of the Poisson distribution are the expected size of the price jump (μ), the variance in this value (δ^2) and the likelihood of a price jump in any specified time period (λ) and Press estimated these values for ten stocks. In subsequent papers, Beckers and Ball and Torous suggest ways of refining these estimates.[23] In an attempt to bridge the gap between the CAPM and jump process models, Jarrow and Rosenfeld derive a version of the capital

---

[21] The earliest paper on this topic was by Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, Journal of Finance 31, 1085-1100. They generated a three-moment CAPM, with a measure of co-skewness (of the asset with the market) added to capture preferences for skewness, and argued that it helped better explain differences across stock returns. In a more recent paper, Harvey, C. and Siddique, A. (2000). Conditional skewness in asset pricing tests, Journal of Finance, 55, 1263-1295, use co-skewness to explain why small companies and low price to book companies earn higher returns

[22] Fang, H. and Lai T-Y. (1997). Co-kurtosis and capital asset pricing, *The Financial Review*, 32, 293-307. In this paper, the authors introduce a measure of co-kurtosis (stock price jumps that are correlated with market jumps) and argue that it adds to the risk of a stock.

[23] Beckers, S., 1981, A Note on Estimating the Parameters of the Diffusion- Jump Process Model of Stock Returns, Journal of Financial and Quantitative Analysis, v16, 127-140; Ball, C.A. and W.N. Torous, 1983, A Simplified Jump Process for Common Stock Returns, Journal of Financial and Quantitative Analysis, v18, 53-65.

asset pricing model that includes a jump component that captures the likelihood of market jumps and an individual asset's correlation with these jumps. [24]

While jump process models have gained some traction in option pricing, they have had limited success in equity markets, largely because the parameters of jump process models are difficult to estimate with any degree of precision. Thus, while everyone agrees that stock prices jump, there is little consensus on the best way to measure how often this happens and whether these jumps are diversifiable and how best to incorporate their effect into risk measures.

Data Power: Arbitrage Pricing and Multi-Factor Models

There have been two developments in the last three decades that have changed the way we think about risk measurement. The first was access to richer data on stock and commodity market information; researchers could not only get information on weekly, daily or even intraday prices but also on trading volume and bid-ask spreads. The other was the increase in both personal and mainframe computing power, allowing researchers to bring powerful statistical tools to bear on the data. As a consequence of these two trends, we have seen the advent of risk measures that are based almost entirely on observed market prices and financial data.

Arbitrage Pricing Model

The first direct challenge to the capital asset pricing model came in the mid-seventies, when Steve Ross developed the arbitrage pricing model, using the fundamental proposition that two assets with the same exposure to risk had to be priced the same by the market to prevent investors from generating risk-free or arbitrage profits.[25] In a market where arbitrage opportunities did not exist, he argued that you can back out measures of risk from observed market returns. Appendix 3 provides a short summary of the derivation of the arbitrage pricing model.

---

[24] Jarrow, R.A. and E.R. Rosenfeld, 1984, Jump Risks and the Intertemporal Capital Asset Pricing Model, Journal of Business, v 57, 337-351.
[25] Ross, Stephen A., 1976, *The Arbitrage Theory Of Capital Asset Pricing*, Journal of Economic Theory, v13(3), 341-360.

The statistical technique that Ross used to extract these risk measures was factor analysis. He examined (or rather got a computer to analyze) returns on individual stocks over a very long time period and asked a fundamental question: Are there common factors that seem to cause large numbers of stock to move together in particular time periods? The factor analysis suggested that there were multiple factors affecting overall stock prices; these factors were termed market risk factors since they affected many stocks at the same time. As a bonus, the factor analysis measured each stock's exposure to each of the multiple factors; these measures were titled factor betas.

In the parlance of the capital asset pricing model, the arbitrage pricing model replaces the single market risk factor in the CAPM (captured by the market portfolio) with multiple market risk factors, and the single market beta in the CAPM (which measures risk added by an individual asset to the market portfolio) with multiple factor betas (measuring an asset's exposure to each of the individual market risk factors). More importantly, the arbitrage pricing model does not make restrictive assumptions about investor utility functions or the return distributions of assets. The tradeoff, though, is that the arbitrage pricing model does depend heavily on historical price data for its estimates of both the number of factors and factor betas and is at its core more of a statistical than an economic model.

Multi-factor and Proxy Models

While arbitrage pricing models restrict themselves to historical price data, multi-factor models expand the data used to include macro-economic data in some versions and firm-specific data (such as market capitalization and pricing ratios) in others. Fundamentally, multi-factor models begin with the assumption that market prices usually go up or down for good reason, and that stocks that earn high returns over long periods must be riskier than stocks that earn low returns over the same periods. With that assumption in place, these models then look for external data that can explain the differences in returns across stocks.

One class of multi factor models restrict the external data that they use to macroeconomic data, arguing that the risk that is priced into stocks should be market risk and not firm-specific risk. For instance, Chen, Roll, and Ross suggest that the following macroeconomic variables are highly correlated with the factors that come out of factor analysis: the level of industrial production, changes in the default spread (between corporate and treasury bonds), shifts in the yield curve (captured by the difference between long and short term rates), unanticipated inflation, and changes in the real rate of return.[26] These variables can then be correlated with returns to come up with a model of expected returns, with firm-specific betas calculated relative to each variable. In summary, Chen, Roll and Ross found that stock returns were more negative in periods when industrial production fell and the default spread, unanticipated inflation and the real rate of return increased. Stocks did much better in periods when the yield curve was more upward sloping – long term rates were higher than short term rates – and worse in periods when the yield curve was flat or downward sloping. With this approach, the measure of risk for a stock or asset becomes its exposure to each of these macroeconomic factors (captured by the beta relative to each factor).

While multi-factor models may stretch the notion of market risk, they remain true to its essence by restricting the search to only macro economic variables. A second class of models weakens this restriction by widening the search for variables that explain differences in stock returns to include firm-specific factors. The most widely cited study using this approach was by Fama and French where they presented strong evidence that differences in returns across stocks between 1962 and 1990 were best explained not by CAPM betas but by two firm-specific measures: the market capitalization of a company and its book to price ratio.[27] Smaller market cap companies and companies with higher book to price ratios generated higher annual returns over this period than larger market cap companies with lower book to price ratios. If markets are reasonably efficient in the long term, they argued that this must indicate that market capitalization and book to price ratios were good stand-ins or proxies for risk measures. In the years since, other factors

[26] Chen, N., R. Roll and S.A. Ross, 1986, *Economic Forces and the Stock Market*, Journal of Business, 1986, v59, 383-404.

[27] Fama, E.F. and K.R. French, 1992, *The Cross-Section of Expected Returns*, Journal of Finance, v47, 427-466. There were numerous other studies prior to this one that had the same conclusions as this one but their focus was different. These earlier studies uses their findings that low PE, low PBV and small companies earned higher returns than expected (based on the CAPM) to conclude that either markets were not efficient or that the CAPM did not work.

have added to the list of risk proxies – price momentum, price level per share and liquidity are a few that come to mind.[28]

Multi-factor and proxy models will do better than conventional asset pricing models in explaining differences in returns because the variables chosen in these models are those that have the highest correlation with returns. Put another way, researchers can search through hundreds of potential proxies and pick the ones that work best. It is therefore unfair to argue for these models based purely upon their better explanatory power.

The Evolution of Risk Measures

The way in which we measure risk has evolved over time, reflecting in part the developments in statistics and economics on the one hand and the availability of data on the other. In figure 4.5, we summarize the key developments in the measurement of risk and the evolution of risk measures over time:

[28] Stocks that have gone up strongly in the recent past (his momentum), trade at low prices per share and are less liquid earn higher returns than stocks without these characteristics.

Figure 4.5: Key Developments in Risk Analysis and Evolution of Risk Measures

Key Event		*Risk Measure used*
Risk was considered to be either fated and thus impossible to change or divine providence in which case it could be altered only through prayer or sacrifice.	Pre-1494	None or gut feeling
Luca Pacioli posits his puzzle with two gamblers in a coin tossing game	1494	
Pascal and Fermal solve the Pacioli puzzle and lay foundations for probability estimation and theory	1654	Computed Probabilities
Graunt generates life table using data on births and deaths in London	1662	
Bernoulli states the "law of large numbers", providing the basis for sampling from large populations.	1711	Sample-based probabilities
de Moivre derives the normal distribution as an approximatiion to the binomial and Gauss & Laplace refine it.	1738	
Bayes published his treatise on how to update prior beliefs as new information is acquired.	1763	
Insurance business develops and with it come actuarial measures of risk, basedupon historical data.	1800s	Expected loss
Bachelier examines stock and option prices on Paris exchanges and defends his thesis that prices follow a random walk.	1900	Price variance
Standard Statistics Bureau, Moody's and Fitch start rating corporate bonds using accounting information.	1909-1915	Bond & Stock Ratings
Markowitz lays statistical basis for diversification and generates efficient portfolios for different risk levels.	1952	Variance added to portfolio
Sharpe and Lintner introduce a riskless asset and show that combinations of it and a market portfolio (including all traded assets) are optimal for all investors. The CAPM is born.	1964	Market beta
Risk and return models based upon alternatives to normal distribution - Power law, asymmetric and jump process distributions	1960-	
Using the "no arbitrage" argument, Ross derives the arbitrage pricing model; multiple market risk factors are derived from the historical data.	1976	Factor betas
Macroeconomic variables examined as potenntial market risk factors, leading the multi-factor model.	1986	Macro economic betas
Fama and French, examining the link between stock returns and firm-speciic factors conclude that market cap and book to price at better proxies for risk than beta or betas.	1992	Proxies

It is worth noting that as new risk measures have evolved, the old ones have not been entirely abandoned. Thus, while much of academic research may have jumped on the portfolio theory bandwagon and its subsequent refinements, there are still many investors who are more comfortable with subjective judgments about risk or overall risk categories (stocks are risky and bonds are not).

Conclusion

To manage risk, we first have to measure it. In this chapter, we look at the evolution of risk measures over time. For much of recorded time, human beings attributed negative events to fate or divine providence and therefore made little effort to measure it quantitatively. After all, if the gods have decided to punish you, no risk measurement device or risk management product can protect you from retribution.

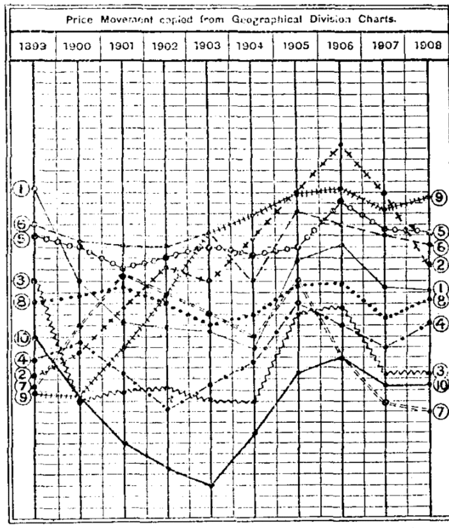
The first break in this karmic view of risk occurred in the middle ages when mathematicians, more in the interests of success at the card tables than in risk measurement, came up with the first measures of probability. Subsequent advances in statistics – sampling distributions, the law of large numbers and Bayes' rule, to provide three examples – extended the reach of probability into the uncertainties that individuals and businesses faced day to day. As a consequence, the insurance business was born, where companies offered to protect individuals and businesses from expected losses by charging premiums. The key, though, was that risk was still perceived almost entirely in terms of potential downside and losses.

The growth of markets for financial assets created a need for risk measures that captured both the downside risk inherent in these investments as well as the potential for upside or profits. The growth of services that provided estimates of these risk measures parallels the growth in access to pricing and financial data on investments. The bond rating agencies in the early part of the twentieth century provided risk measures for corporate bonds. Measures of equity risk appeared at about the same time but were primarily centered on price volatility and financial ratios.

While the virtues of diversifying across investments had been well publicized at the time of his arrival, Markowitz laid the foundation for modern portfolio theory by making explicit the benefits of diversification. In the aftermath of his derivation of efficient portfolios, i.e. portfolios that maximized expected returns for given variances, three classes of models that allowed for more detailed risk measures developed. One class included models like the CAPM that stayed true to the mean variance framework and measured risk for any asset as the variance added on to a diversified portfolio. The second set of models relaxed the normal distribution assumption inherent in the CAPM and allowed for more general distributions (like the power law and asymmetric distributions) and the risk measures emanating from these distributions. The third set of models trusted the market to get it right, at least on average, and derived risk measures by looking at past history. Implicitly, these models assumed that investments that have earned high returns in the past must have done so because they were riskier and looked for factors that best explain these returns. These factors remained unnamed and were statistical in the arbitrage pricing model, were macro economic variables in multi factor models and firm-specific measures (like market cap and price to book ratios) in proxy models.

Appendix 1: Measuring Risk in Portfolios – Financial Review of Reviews – 1909

5½% INVESTMENT SCHEME.



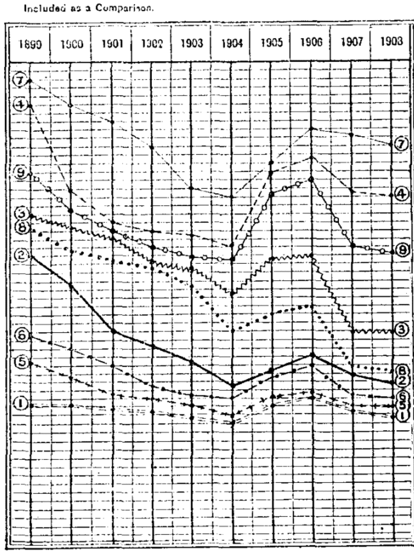
NOTE.—The lines in this Chart are identical with the lines given in each separate Geographical Division Chart.

Geographical Division.	Name and Quantity of Stock.	Denoting Number.	Annual Values at average prices of year: 1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	Annual Income: 1899	1900	1901	1902	1903	1904	1905	1906	1907	1908
BRITISH.	50 Shares Alliance and Dublin Consumers' Gas	1	1112	987	967	965	962	960	1075	1062	1025	981	52½	52½	52½	52¾	52½	52½	52¾	52½	51½	60
BRITISH COLONIES.	£1,300 Toronto Mortgage Company Common Stock	2	975	1007	1111	1199	1176	1248	1355	1436	1324	1371	⊕50	50	50	50	50	60	60	60	60	60
EUROPE, NORTH.	12 Shares Hamburg Electric Supply Ordinary	3	984	870	879	882	873	870	954	960	909	906	54	48	42	42	45	45	48	48	48	48
EUROPE, SOUTH.	50 Shares Vienna Bank Verein	4	1040	1080	1000	920	984	1040	1136	1096	1072	1100	64	60	52	56	56	60	60	60	60	60
ASIA.	£1,000 Nagasaki Harbour Improvement Loan 6% Bonds	5	900	880	840	870	900	870	900	980	940	930	60	60	60	60	60	60	60	60	60	60
AFRICA.	£1,000 Upper Congo Belgian Trading Company 6% Preference Stock	6	1060	1050	1040	1040	1060	980	1080	1060	1050	1050	60	60	60	60	60	60	60	60	60	60
NORTH AMERICA.	£1,000 Laclede Gas Light 5% Cumulative Preferred Stock	7	930	980	1030	1010	990	975	1030	950	900	900	50	50	50	50	50	50	50	50	50	50
CENTRAL AMERICA.	£1,000 New Trinidad Lake Asphalte 6% Debentures	8	980	985	1000	980	950	960	1020	1030	1000	1020	60	60	60	60	60	60	60	60	60	60
SOUTH AMERICA.	£1,500 Brazil Great Southern Railway 6% 1st Debentures	9	1095	1065	1200	1305	1440	1485	1530	1545	1470	1485	90	90	90	90	90	90	90	90	90	90
INTER-NATIONAL.	£1,500 Anglo-American Telegraph Ordinary Stock	10	975	892	787	735	727	787	900	945	900	870	55⅛	46¾	45¾	45⅜	45½	41¼	48¾	58⅛	52½	52⅛
Totals of Annual Values and Annual Income		£	10051	9796	9854	9906	10062	10175	10380	11064	10590	10613	595⅜	577¾	562¼	565⅞	569¼	578¾	589¼	598⅜	591¾	590⅓

⊕ Estimated.
NOTE.—The Interest allowance for contingencies is £40 in 1908.

BRITISH INVESTMENT SCHEME.

Included as a Comparison.



Included as a Comparison.

Description of Enterprise.	Name and Quantity of Stock.	Denoting Number.	Annual Values at average prices of year: 1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	Annual Income: 1899	1900	1901	1902	1903	1904	1905	1906	1907	1908
CORPORATION.	£1,000 Southampton Corporation 3½% Stock, Redeemable 1945	1	1024	1005	1004	1001	996	987	999	990	975	975	35	35	35	35	35	35	35	35	35	35
RAILWAY.	£800 Furness Railway 4% Consolidated Preference Stock	2	1048	996	920	896	856	820	836	852	824	780	32	32	32	32	32	32	32	32	32	32
TRAMWAY.	200 Shares City of Birmingham Trams 5% Cumulative Preference	3	1090	1080	1075	1025	1020	975	1030	1055	950	925	50	50	50	50	50	50	50	50	50	50
GAS.	50 Shares Alliance & Dublin Consumers' Gas 10% Standard	4	1112	987	967	965	962	960	1075	1062	1025	981	52½	52½	52½	52½	52½	52½	52½	52½	51½	50
BANK.	12 Shares Parr's Bank	5	1074	1068	1026	1014	1008	1005	1023	1056	1020	1020	45	48	48	48	48	48	48	48	48	48
ELECTRICITY.	£1,200 London Electric Supply Corporation 4% 1st Mortgage Debenture Stock	6	1260	1212	1170	1164	1194	1158	1162	1176	1110	1098	48	48	48	48	48	48	48	48	48	48
COTTON.	£1,300 English Sewing Cotton 4% 1st Mortgage Debenture Stock	7	1345	1319	1235	1098	1059	1053	1170	1248	1241	1235	52	52	52	52	52	52	52	52	52	52
IRON, COAL AND STEEL.	£800 Vickers, Sons & Maxim 5% Preferred Stock	8	1060	1028	1012	1012	936	868	920	936	872	848	40	40	40	40	40	40	40	40	40	40
SOAP.	700 Shares Pears (A. & F.) Ordinary	9	1006	964	950	929	922	922	1010	1024	940	940	70	70	70	70	70	70	70	70	70	70
Totals of Annual Values and Annual Income		£	10019	9659	9359	9104	8953	8768	9245	9399	8957	8802	425	427½	427½	427½	427½	427½	427½	427½	426½	425

The nine other Investment Schemes included in the first set of Charts would have shown equal want of stability of capital to the above investment holding, and this example is given in order to show how the others would compare with the geographically distributed investment schemes.

Appendix 2: Mean-Variance Framework and the CAPM

Consider a portfolio of two assets. Asset A has an expected return of μ_A and a variance in returns of σ^2_A, while asset B has an expected return of μ_B and a variance in returns of σ^2_B. The correlation in returns between the two assets, which measures how the assets move together, is ρ_{AB}. The expected returns and variance of a two-asset portfolio can be written as a function of these inputs and the proportion of the portfolio going to each asset.

$$\mu_{portfolio} = w_A \mu_A + (1 - w_A) \mu_B$$

$$\sigma^2_{portfolio} = w_A^2 \sigma^2_A + (1 - w_A)^2 \sigma^2_B + 2 w_A w_B \rho_{AB} \sigma_A \sigma_B$$

where

w_A = Proportion of the portfolio in asset A

The last term in the variance formulation is sometimes written in terms of the covariance in returns between the two assets, which is

$$\sigma_{AB} = \rho_{AB} \sigma_A \sigma_B$$

The savings that accrue from diversification are a function of the correlation coefficient. Other things remaining equal, the higher the correlation in returns between the two assets, the smaller are the potential benefits from diversification. The following example illustrates the savings from diversification.

If there is a diversification benefit of going from one asset to two, as the preceding discussion illustrates, there must be a benefit in going from two assets to three, and from three assets to more. The variance of a portfolio of three assets can be written as a function of the variances of each of the three assets, the portfolio weights on each and the correlations between pairs of the assets. It can be written as follows -

$$\sigma_p^2 = w_A^2 \sigma^2_A + w_B^2 \sigma^2_B + w_C^2 \sigma^2_C + 2 w_A w_B \rho_{AB} \sigma_A \sigma_B + 2 w_A w_C \rho_{AC} \sigma_A \sigma_C + 2 w_B w_C \rho_{BC} \sigma_B \sigma_C$$

where

w_A, w_B, w_C = Portfolio weights on assets

$\sigma^2_A, \sigma^2_B, \sigma^2_C$ = Variances of assets A, B, and C

$\rho_{AB}, \rho_{AC}, \rho_{BC}$ = Correlation in returns between pairs of assets (A&B, A&C, B&C)

Note that the number of covariance terms in the variance formulation has increased from one to three. This formulation can be extended to the more general case of a portfolio of n assets:

$$\sigma_p^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} w_i \; w_j \; \rho_{ij} \; \sigma_i \; \sigma_j$$

The number of terms in this formulation increases exponentially with the number of assets in the portfolio, largely because of the number of covariance terms that have to be considered. In general, the number of covariance terms can be written as a function of the number of assets:

Number of covariance terms = n (n-1) /2

where n is the number of assets in the portfolio. Table 4A.1 lists the number of covariance terms we would need to estimate the variances of portfolios of different sizes.

Table 4A.1: Number of Covariance Terms

Number of Assets	*Number of Covariance Terms*
2	1
10	45
100	4950
1000	499500
10000	49995000

This formulation can be used to estimate the variance of a portfolio and the effects of diversification on that variance. For purposes of simplicity, assume that the average asset has a standard deviation in returns of $\overline{\sigma}$ and that the average covariance in returns between any pair of assets is $\overline{\sigma_{ij}}$. Furthermore, assume that the portfolio is always equally weighted across the assets in that portfolio. The variance of a portfolio of n assets can then be written as

$$\sigma_p^2 = n \left(\frac{1}{n}\right)^2 \overline{\sigma}^2 + \frac{(n-1)}{n} \; \overline{\sigma}_{ij}$$

The fact that variances can be estimated for portfolios made up of a large number of assets suggests an approach to optimizing portfolio construction, in which investors trade off expected return and variance. If an investor can specify the maximum amount of risk he is willing to take on (in terms of variance), the task of portfolio optimization becomes the maximization of expected returns subject to this level of risk. Alternatively, if an investor specifies her desired level of return, the optimum portfolio is the one that minimizes the variance subject to this level of return. These optimization algorithms can be written as follows.

Return Maximization | *Risk Minimization*

Maximize Expected Return | Minimize return variance

$$E(R_p) = \sum_{i=1}^{i=n} w_i \, \mathrm{E(R_i)} \qquad\qquad \sigma_p^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} w_i w_j \sigma_{\mathrm{ij}}$$

subject to

$$\sigma_p^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} w_i w_j \sigma_{\mathrm{ij}} \le \hat{\sigma}^2 \qquad\qquad E(R_p) = \sum_{i=1}^{i=n} w_i \, \mathrm{E(R_i)} = \mathrm{E(\hat{R})}$$

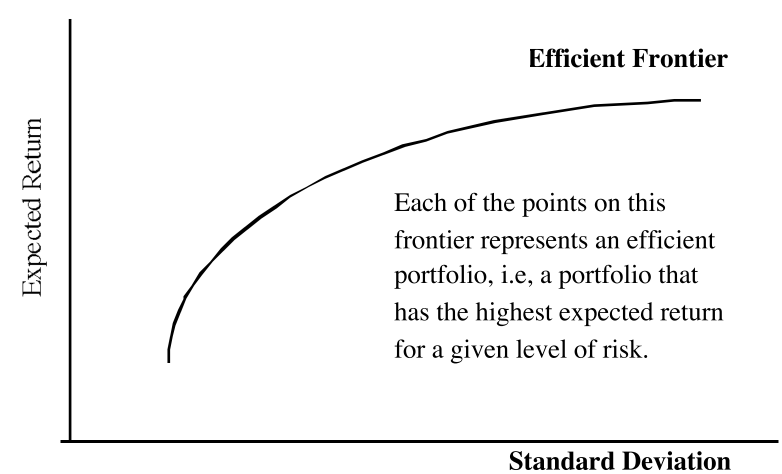
where,

$\hat{\sigma}$ = Investor's desired level of variance

$\mathrm{E(\hat{R})}$ = Investor's desired expected returns

The portfolios that emerge from this process are called **Markowitz portfolios**. They are considered efficient, because they maximize expected returns given the standard deviation, and the entire set of portfolios is referred to as the **Efficient Frontier**. Graphically, these portfolios are shown on the expected return/standard deviation dimensions in figure 4A.1 -

Figure 4A.1: Markowitz Portfolios



The Markowitz approach to portfolio optimization, while intuitively appealing, suffers from two major problems. The first is that it requires a very large number of inputs, since the covariances between pairs of assets are required to estimate the variances of portfolios. While this may be manageable for small numbers of assets, it becomes less so when the entire universe of stocks or all investments is considered. The second problem is that the Markowitz approach ignores a very important asset choice that most investors have -- riskless default free government securities -- in coming up with optimum portfolios.

To get from Markowitz portfolios to the capital asset pricing model, let us considering adding a riskless asset to the mix of risky assets. By itself, the addition of one asset to the investment universe may seem trivial, but the riskless asset has some special characteristics that affect optimal portfolio choice for all investors.

(1) The riskless asset, by definition, has an expected return that will always be equal to the actual return. The expected return is known when the investment is made, and the actual return should be equal to this expected return; the standard deviation in returns on this investment is zero.

(2) While risky assets' returns vary, the absence of variance in the riskless asset's returns make it uncorrelated with returns on any of these risky assets. To examine what happens to the variance of a portfolio that combines a riskless asset with a risky portfolio, assume that the variance of the risky portfolio is σ_r^2 and that w_r is the proportion of the overall portfolio invested to these risky assets. The balance is invested in a riskless asset, which has no variance, and is uncorrelated with the risky asset. The variance of the overall portfolio can be written as:

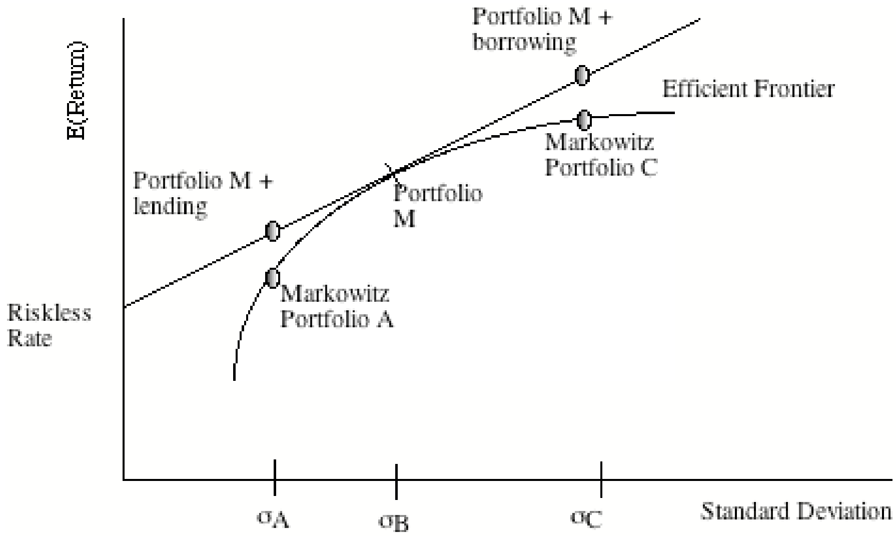
$$\sigma^2_{portfolio} = w_r^2\, \sigma^2_r$$

$$\sigma_{portfolio} = w_r\, \sigma_r$$

Note that the other two terms in the two-asset variance equation drop out, and the standard deviation of the overall portfolio is a linear function of the portfolio invested in the risky portfolio.

The significance of this result can be illustrated by returning to figure 4A.1 and adding the riskless asset to the choices available to the investor. The effect of this addition is explored in figure 4A.2.

Figure 4A.2: Introducing a Riskless Asset



Consider investor A, whose desired risk level is σ_A. This investor, instead of choosing portfolio A, the Markowitz portfolio containing only risky assets, will choose to invest in a combination of the riskless asset and a much riskier portfolio, since he will be able to make a much higher return for the same level of risk. The expected return increases as the slope of the line drawn from the riskless rate increases, and the slope is maximized when the line is tangential to the efficient frontier; the risky portfolio at the point of tangency is labeled as risky portfolio M. Thus, investor A's expected return is maximized by holding a combination of the riskless asset and risky portfolio M. Investor B, whose desired risk level is σ_B, which happens to be equal to the standard deviation of the risky portfolio M, will choose to invest her entire portfolio in that portfolio. Investor C, whose desired risk level is σ_C, which exceeds the standard deviation of the risky portfolio M, will borrow money at the riskless rate and invest in the portfolio M.

In a world in which investors hold a combination of only two assets -- the riskless asset and the market portfolio -- the risk of any individual asset will be measured relative to the market portfolio. In particular, the risk of any asset will be the risk it adds on to the market portfolio. To arrive at the appropriate measure of this added risk, assume that σ^2_m is the variance of the market portfolio prior to the addition of the new asset, and that the variance of the individual asset being added to this portfolio is σ^2_i. The market value portfolio weight on this asset is w_i, and the covariance in returns between the individual asset and the market portfolio is σ_{im}. The variance of the market portfolio prior to and after the addition of the individual asset can then be written as

Variance prior to asset i being added = $\sigma^2{}_m$

Variance after asset i is added = $\sigma^2{}_{m'} = w_i{}^2\, \sigma^2{}_i + (1 - w_i)^2\, \sigma^2{}_m + 2\, w_i\, (1-w_i)\, \sigma_{im}$

The market value weight on any individual asset in the market portfolio should be small since the market portfolio includes all traded assets in the economy. Consequently, the first term in the equation should approach zero, and the second term should approach $\sigma^2{}_m$, leaving the third term (σ_{im}, the covariance) as the measure of the risk added by asset i. Dividing this term by the variance of the market portfolio yields the beta of an asset:

Beta of asset = $\frac{\sigma_{im}}{\sigma_m^2}$

Appendix 3: Derivation of the Arbitrage Pricing Model

Like the capital asset pricing model, the arbitrage pricing model begins by breaking risk down into firm-specific and market risk components. As in the capital asset pricing model, firm specific risk covers information that affects primarily the firm whereas market risk affects many or all firms. Incorporating both types of risk into a return model, we get:

$$R \quad = \quad E(R) \quad + \quad m \quad + \quad \varepsilon$$

where R is the actual return, E(R) is the expected return, m is the market-wide component of unanticipated risk and ε is the firm-specific component. Thus, the actual return can be different from the expected return, either because of market risk or firm-specific actions. In general, the market component of unanticipated returns can be decomposed into economic factors:

$$\begin{aligned} R &= R + m + \varepsilon \\ &= R + (\beta_1 F_1 + \beta_2 F_2 + ... + \beta_n F_n) + \varepsilon \end{aligned}$$

where

β_j = Sensitivity of investment to unanticipated changes in factor j

F_j = Unanticipated changes in factor j

Note that the measure of an investment's sensitivity to any macro-economic factor takes the form of a beta, called a **factor beta**. In fact, this beta has many of the same properties as the market beta in the CAPM.

The arbitrage pricing model assumes that firm-specific risk component (ε) is can be diversified away and concludes that the return on a portfolio will not have a firm-specific component of unanticipated returns. The return on a portfolio can be written as the sum of two weighted averages -that of the anticipated returns in the portfolio and that of the market factors:

$$\begin{aligned} R_p = {} & (w_1R_1+w_2R_2+...+w_nR_n)+ (w_1\beta_{1,1}+w_2\beta_{1,2}+...+w_n\beta_{1,n}) F_1 + \\ & (w_1\beta_{2,1}+w_2\beta_{2,2}+...+w_n\beta_{2,n}) F_2 \; ..... \end{aligned}$$

where,

w_j = Portfolio weight on asset j

R_j = Expected return on asset j

$\beta_{i,j}$= Beta on factor i for asset j

The final step in this process is estimating an expected return as a function of the betas specified above. To do this, we should first note that the beta of a portfolio is the weighted average of the betas of the assets in the portfolio. This property, in conjunction with the absence of arbitrage, leads to the conclusion that expected returns should be linearly related to betas. To see why, assume that there is only one factor and three portfolios. Portfolio A has a beta of 2.0 and an expected return on 20%; portfolio B has a beta of 1.0 and an expected return of 12%; and portfolio C has a beta of 1.5 and an expected return on 14%. Note that the investor can put half of his wealth in portfolio A and half in portfolio B and end up with a portfolio with a beta of 1.5 and an expected return of 16%. Consequently no investor will choose to hold portfolio C until the prices of assets in that portfolio drop and the expected return increases to 16%. By the same rationale, the expected returns on every portfolio should be a linear function of the beta. If they were not, we could combine two other portfolios, one with a higher beta and one with a lower beta, to earn a higher return than the portfolio in question, creating an opportunity for arbitrage. This argument can be extended to multiple factors with the same results. Therefore, the expected return on an asset can be written as

$$E(R) = R_f + \beta_1 [E(R_1)-R_f] + \beta_2 [E(R_2)-R_f] \ ...+ \beta_n [E(R_n)-R_f]$$

where

R_f = Expected return on a zero-beta portfolio

$E(R_j)$ = Expected return on a portfolio with a factor beta of 1 for factor j, and zero for all other factors.

The terms in the brackets can be considered to be risk premiums for each of the factors in the model.