# 1 The Rise of Value at Risk

We can think of financial risk as the risk associated with financial outcomes of one sort or another, but the term 'risk' itself is very difficult to pin down precisely. It evokes notions of uncertainty, randomness, and probability. The random outcomes to which it alludes might be good (e.g., we might win a lottery) or bad (e.g., we might suffer a financial loss), and we may (or may not) prefer to focus on the risks associated with 'bad' events, presumably with a view to trying to protect ourselves against them. There is also the question of quantifiability – some scholars distinguish between 'risk' as something quantifiable and 'uncertainty' as its non-quantifiable counterpart. The notion of 'risk' in its broadest sense therefore has many facets, and there is no single definition of risk that can be completely satisfactory in every situation. However, for our purposes here, a reasonable definition is to consider financial risk as the prospect of financial loss – or maybe gain – due to unforeseen or random changes in underlying risk factors.

In this book we are concerned with the measurement of one particular form of financial risk – namely, market risk, or the risk of loss (or gain) arising from unexpected changes in market prices (e.g., such as security prices) or market rates (e.g., such as interest or exchange rates). Market risks, in turn, can be classified into interest-rate risks, equity risks, exchange rate risks, commodity price risks, and so on, depending on whether the risk factor is an interest rate, a stock price, or another random variable. Market risks can also be distinguished from other forms of financial risk, particularly credit risk (or the risk of loss arising from the failure of a counterparty to make a promised payment) and operational risk (or the risk of loss arising from the failures of internal systems or the people who operate in them).

The theory and the practice of risk management have developed enormously since the pioneering work of Harry Markowitz in the 1950s. The theory has developed to the point where risk management is now regarded as a distinct subfield of the theory of finance, and one that is increasingly taught as a separate subject in the more advanced master's and MBA programmes in finance. The subject has attracted a huge amount of intellectual energy, not just from finance specialists but also from specialists in other disciplines who are attracted to it – as is illustrated by the large number of ivy league theoretical physics PhDs who now go into finance research, attracted not just by high salaries but also by the challenging intellectual problems it poses. The subject has benefited enormously from contributions made by quantitative disciplines such as statistics, mathematics, and computer science (and others, such as engineering and physics). However, the subject is not purely, or even mainly, a quantitative one. At the heart of the subject is the notion of good risk management practice, and above anything else this requires an awareness of the qualitative and organisational aspects of risk management: a good sense of judgement, an awareness of the 'things that can go wrong', an appreciation of market history, and so on. This also means some of the most important principles of risk management actually come from disciplines outside finance, most especially the disciplines of accounting (which tells us about subjects such as management control, valuation and audit), economics (which tells us about how markets behave and about welfare maximisation, among other things), organisational theory (which tells us about how organisations behave), and law (which pervades almost everything in risk management). So, while risk management involves quantitative methods, the subject itself rests on a foundation that is qualitative. In many ways, the subject is much like engineering: it uses sophisticated tools, but context and judgement are everything. And this, perhaps, is the most important thing for any budding risk manager to appreciate – especially one from a quants background.

**Box 1.1** Why Manage Corporate Financial Risks?

At one level, the benefits of risk management are obvious: we reduce the danger of harmful events occurring. However, this response does not fully explain why *firms* might practice financial risk management. Even if individual investors are risk averse and manage the investment portfolio risks, it still does not follow that firms should manage *their* overall corporate risks. If investors have access to perfect capital markets (with all the economic textbook fictions that that entails), they can achieve the degrees of diversification they want through their own actions, and corporate financial risk management would be irrelevant. This is the message of the famous Modigliani–Miller theorem, which says that in an ideal theoretical world with no informational asymmetries, principal–agent problems, taxes, transactions costs or bankruptcy costs, and with 'perfect' frictionless markets, the financial structure of the firm (and, by implication, any risk management) would be irrelevant. Hence, any *explanation* of the benefits of corporate financial risk management must start by identifying which of the Modigliani–Miller assumptions do *not* apply to the real world; relaxing the relevant assumption then enables us to see why firms might benefit from financial risk management. These benefits arise from the following:

- Risk management helps to increase the value of the firm in the presence of bankruptcy costs, because it makes bankruptcy less likely.
- The presence of informational asymmetries means that external finance is more costly than internal finance, and good investment opportunities can be lost. Risk management helps alleviate these problems by reducing the variability of the corporate cash flow.
- Risk management helps investors achieve a better allocation of risks, because the firm would typically have better access to capital markets.
- In the presence of taxes, risk management can help reduce the firm's tax bill, because the amount of tax paid is a convex function of its profits: this means that the less variable its profits, the lower its average tax bill.

## 1.1 THE EMERGENCE OF FINANCIAL RISK MANAGEMENT

The emergence of financial risk management as a discipline is due to a number of factors. One factor is the phenomenal growth in trading activity since the late 1960s, illustrated by the facts that the average number of shares traded per day in the New York Stock Exchange has grown from a little over \$4 million in 1961 to around \$1.6 trillion in early 2005, and that turnover in foreign exchange markets has grown from about a billion dollars a day in 1965 to \$1.9 trillion in April 2004.[1] There have also been massive increases in the range of instruments traded over the past two or three decades, and trading volumes in these new instruments have also grown very

[1] The latter figure is from BIS (2004a), p. 1.

rapidly. New instruments have been developed in offshore markets and, more recently, in the newly emerging financial markets of eastern Europe, India, East Asia, Latin America, Russia, and elsewhere. New instruments have also arisen for assets that were previously illiquid, such as consumer loans, commercial and industrial bank loans, mortgages, mortgage-based securities, and similar assets, and these markets have grown very considerably since the early 1980s.

There has also been a huge growth of financial derivatives activity. Until 1972 the only derivatives traded were certain commodity futures and various forwards and some over-the-counter (OTC) options. The Chicago Mercantile Exchange then started trading foreign currency futures contracts in 1972, and in 1973 the Chicago Board Options Exchange started trading equity call options. Interest-rate futures were introduced in 1975, and a large number of other financial derivatives contracts were introduced in the following years: swaps and exotics (e.g., swaptions, futures on interest rate swaps, etc.) then took off in the 1980s, and catastrophe, credit, electricity and weather derivatives in the 1990s and mortality derivatives in the 2000s. From negligible amounts in the early 1970s, the total notional amounts held in outstanding OTC derivatives contracts grew to $220 trillion by the first half of 2004.[2]

This growth in trading activity has taken place against an environment that was often very volatile. A volatile environment exposes firms to greater levels of financial risk, and provides incentives for firms to find new and better ways of managing this risk. The volatility of the economic environment is reflected in various ways:

- *Stock market volatility*: Stock markets have always been volatile, but sometimes extremely so: for example, on October 19, 1987, the Dow Jones fell 23% and in the process knocked off over $1 trillion in equity capital; and from July 21 through August 31, 1998, the Dow Jones lost 18% of its value. Other western stock markets have experienced similar falls, and some Asian ones have experienced much worse ones (e.g., the South Korean stock market lost over half of its value over 1997).
- *Exchange rate volatility*: Exchange rates have been volatile ever since the breakdown of the Bretton Woods system of fixed exchange rates in the early 1970s. Occasional exchange rate crises have also led to sudden and significant exchange rate changes, including – among many others – the ERM devaluations of September 1992, the problems of the peso in 1994, the east Asian currency problems of 1997–98, the rouble crisis of 1998, Brazil in 1999 and Argentina in 2001.
- *Interest-rate volatility*: There have also been major fluctuations in interest rates, with their attendant effects on funding costs, corporate cash flows and asset values. For example, the Fed Funds rate, a good indicator of short-term market rates in the US, approximately doubled over 1994.
- *Commodity market volatility*: Commodity markets are notoriously volatile, and commodity prices often go through long periods of apparent stability and then suddenly jump by enormous amounts. Some commodity prices (e.g., electricity prices) also show extremely pronounced day-to-day and even hour-to-hour volatility.

The development of risk management has also been spurred on by concerns with the dangers of improper derivatives use, and by a sorry catalogue of risk management disasters since the early 1990s. These dangers were sounded loud and clear by E. Gerald Corrigan, the then

[2] BIS (2004b), p. 9. However, this figure is misleading, because notional values give relatively little indication of what derivatives contracts are really worth. The true size of derivatives trading is therefore better represented by the market value of outstanding derivatives contracts. The same survey estimated this to be $6.4 trillion – which is a little under 3% of the notional amount, but still an astronomical number in its own right.

President of the Federal Reserve Bank of New York, in an address to the New York Bankers Association in early 1992 in which he told them very bluntly:

> You had all better take a very, very hard look at off-balance sheet activities. The growth and complexity of [these] activities and the nature of the credit settlement risk they entail should give us cause for concern . . . . I hope this sounds like a warning, because it is.

Corrigan's concerns proved well grounded. Time and again in the next few years, apparently respectable institutions stunned the world by announcing massive losses, often through unauthorised trades, and always because of inadequate risk management: Showa Shell in 1993, Metallgesellschaft, Kashima Oil, Procter and Gamble and Orange County in 1994, Barings and Daiwa in 1995, NatWest in 1997, Yakult Honsha and LTCM in 1998, Enron in 2001, Allied Irish Bank and WorldCom in 2002, and Parmalat in 2003. Most of these disasters involved losses of over a billion US$, and this list ignores most of the smaller disasters. These disasters have made the financial community keenly aware of the importance of good risk management practice – as the saying goes, if you think risk management is expensive, just look at the alternative. The disasters of the early 1990s and associated concerns with the dangers of derivatives also encouraged a lot of soul searching in the financial risk community, and led to a series of reports which articulated best practice principles. The most widely cited of these was the G-30 Report (1993), but other, similar, reports were also issued by the US General Accounting Office (1994 and 1996), and the Derivatives Policy Group (1995), among others. They all preached much the same message – the need for good management controls including the need to separate front and back offices, the use of VaR models to estimate firmwide risks, the need for adequate systems to manage derivatives risks, the need to monitor counterparty relationships, the importance of good disclosure, and the like.

A fourth factor contributing to the development of risk management was the rapid advance in the state of information technology. Improvements in IT have made possible huge increases in both computational power and in the speed with which calculations can be carried out. Improvements in computing power mean that new techniques can be used (such as computer-intensive simulation techniques) to enable us to tackle more difficult calculation problems. Improvements in calculation speed then help make these techniques useful in real time, where it is often essential to get answers quickly. This technological progress has led to IT costs falling by about 25–30% a year over the past 40 years or so, and the costs of transmitting data have fallen even faster. Improvements in computing power, increases in computing speed, and reductions in computing costs have thus come together to transform the technology available for risk management. Managers are no longer tied down to the simple 'back of the envelope' techniques that they had to use earlier when they lacked the means to carry out complex calculations. Instead, they can now use sophisticated algorithms programmed into computers to carry out real-time calculations that were not possible before, and the ability to carry out such calculations in turn creates a whole new range of risk management possibilities – and problems. [3]

## 1.2 MARKET RISK MEASUREMENT

This book is concerned not so much with the broader field of financial risk management or even with the somewhat narrower field of financial risk measurement, but with the smaller

[3] The contributory factors discussed here are by no means exhaustive. Besides the factors mentioned in the text there are also the issues of globalisation and increasing organisational complexity, which have made firms more opaque and harder to manage, and in some ways more prone to fail.

(but still very extensive!) subfield of market risk measurement. However, the basic principles of market risk measurement apply with suitable extensions or modifications to other types of risk, so much of what the book discusses could be useful to risk managers concerned with other financial (or even non-financial) risks. Moreover, although we focus on measurement issues, risk measurement and risk management are intimately related, and we cannot address one without also saying a lot about the other. In any case, whatever the particular kind of risk we are dealing with, the first thing to know about measuring any risk is to appreciate the broader risk management context within which the measurement itself takes place: context is all important.

Having delineated our subject matter, it is useful to look at how the techniques of market risk measurement have developed over the years. We therefore turn now to consider the earlier – in retrospect, antediluvian – risk measurement tools that were used before risk managers had computers capable of carrying out more complex calculations.

## 1.3 RISK MEASUREMENT BEFORE VaR

### 1.3.1 Gap Analysis

One common approach is gap analysis, which was initially developed by financial institutions to give a crude idea of interest-rate risk exposure. Gap analysis starts with the choice of an appropriate horizon period (e.g., 1 year ahead). We then determine how much of our asset or liability portfolio will re-price within this period, and the amounts involved give us our rate-sensitive assets and rate-sensitive liabilities for the chosen horizon. The gap is the difference between these, and our interest-rate exposure is taken to be the change in net interest income that occurs in response to a change in interest rates. This is assumed to be equal to the gap times the interest-rate change:

$$\Delta NII = (GAP)\Delta r \tag{1.1}$$

where $\Delta NII$ is the change in net interest income and $\Delta r$ is the change in interest rates.

Gap analysis is fairly simple to carry out, but has its limitations: it only applies to on-balance sheet interest-rate risk, and even then only crudely; it looks at the impact of interest rates on income, rather than on asset or liability values; and results can be sensitive to the choice of horizon period.

### 1.3.2 Duration Analysis

A second traditional method used by financial institutions for measuring interest-rate risks is duration analysis. The (Macaulay) duration $D$ of a bond (or any other fixed-income security) can be defined as the weighted average term to maturity of the bond's cash flows, where the weights are the present value of each cash flow relative to the present value of all cash flows:

$$D = \frac{\sum_{i=1}^{n} i \times PVCF_i}{\sum_{i=1}^{n} PVCF_i} \tag{1.2}$$

where $PVCF_i$ is the present value of the period $i$ cash flow, discounted at the appropriate spot period yield. The duration measure is useful because it gives an approximate indication of the sensitivity of a bond price to a change in yield:

$$\% \textit{ change in bond price} \approx -\frac{D\Delta y}{(1+y)} \tag{1.3}$$

where $y$ is the yield and $\Delta y$ the change in yield. The bigger the duration, the more the bond price changes in response to a change in yield. The duration approach is very convenient because duration measures are easy to calculate and the duration of a bond portfolio is a simple weighted average of the durations of the individual bonds in the portfolio. It is also better than gap analysis insofar as it looks at changes in asset (or liability) values, rather than just changes in net income.

However, duration approaches also have similar limitations: they ignore risks other than interest-rate risk; they are crude,[4] and even with various refinements to improve accuracy,[5] duration-based approaches are still inaccurate relative to more sophisticated approaches to interest-rate term structure analysis (such as Heath–Jarrow–Morton) or market-based models (such as Brace–Gatarak–Musiela). Moreover, the main reason for using duration approaches in the past – their (comparative) ease of calculation – is no longer of much significance, since more sophisticated models can now be programmed into personal computers to give their users more accurate answers very rapidly.

### 1.3.3 Scenario Analysis

Another approach is scenario analysis (or 'what if' analysis), in which we set out different scenarios and investigate what we stand to gain or lose under them. To carry out scenario analysis, we select a set of scenarios – or paths describing how relevant variables, stock prices, interest rates, exchange rates, etc., might evolve over a horizon period. We then postulate the cash flows and/or accounting values of assets and liabilities as they would develop under each scenario, and use the results to come to a view about our exposure. Scenario analyses can be more or less sophisticated, and early scenario analyses were inevitably crude given the limited computing power available. However, they did at least give some idea of what firms stood to lose under specific circumstances.

Scenario analysis is not easy to carry out. A lot hinges on our ability to identify the 'right' scenarios, and there are relatively few rules to guide us when selecting them. We need to ensure that the scenarios are reasonable and do not involve contradictory or excessively implausible assumptions, and we need to think through the interrelationships between the variables involved. We also want to make sure, as best we can, that we have all the main scenarios covered. Scenario analysis also tells us nothing about the likelihood of different scenarios, so we need to use our judgement when assessing the practical significance of different scenarios. In the final analysis, the results of scenario analyses are highly subjective and depend to a very large extent on the skill (or otherwise) of the analyst.

[4] They are crude because they only take a first-order approximation to the change in the bond price, and because they implicitly presuppose that any changes in the yield curve are parallel ones (i.e., all yields across the maturity spectrum change by the same amount). Duration-based hedges are therefore inaccurate against yield changes that involve shifts in the slope of the yield curve.

[5] There are two standard refinements. (1) We can take a second-order rather than first-order approximation to the bond price change. The second-order term – known as convexity – is related to the change in duration as yield changes, and this duration–convexity approach gives us a better approximation to the bond price change as the yield changes. However, duration–convexity usually gives only modest improvements in accuracy over the basic duration approach. (2) An alternative refinement is to use key rate durations (Ho (1992)): if we are concerned about shifts in the yield curve, we can construct separate duration measures for yields of specified maturities (e.g., short-term and long-term yields); these would give us estimates of our exposure to changes in these specific yields and allow us to accommodate non-parallel shifts in the yield curve.

### 1.3.4 Portfolio Theory

A somewhat different approach to risk measurement is provided by portfolio theory. Portfolio theory starts from the premise that investors choose between portfolios on the basis of their expected return, on the one hand, and the standard deviation (or variance) of their return, on the other.[6] The standard deviation of the portfolio return can be regarded as a measure of the portfolio's risk. Other things being equal, an investor wants a portfolio whose return has a high expected value and a low standard deviation. These objectives imply that the investor should choose a portfolio that maximises expected return for any given portfolio standard deviation. A portfolio that meets these conditions is efficient, and a rational investor will always choose an efficient portfolio. When faced with an investment decision, the investor must therefore determine the set of efficient portfolios and rule out the rest. Some efficient portfolios will have more risk than others, but the more risky ones will also have higher expected returns. Faced with the set of efficient portfolios, the investor chooses one particular portfolio on the basis of his or her own preferred trade-off between risk and expected return. An investor who is very averse to risk will choose a safe portfolio with a low standard deviation and a low expected return, and an investor who is less risk averse will choose a riskier portfolio with a higher expected return.

One of the key insights of portfolio theory is that the risk of any individual asset is not the standard deviation of the return to that asset, but the extent to which that asset contributes to overall portfolio risk. An asset might be very risky (i.e., have a high standard deviation) when considered on its own, and yet have a return that correlates with the returns to other assets in our portfolio in such a way that acquiring the new asset does not increase the overall portfolio standard deviation. Acquiring the new asset would then be riskless, even though the asset held on its own would still be risky. The moral of the story is that the extent to which a new asset contributes to portfolio risk depends on the correlation or covariance of its return with the returns to the other assets in our portfolio – or, if one prefers, the beta, which is equal to the covariance between the return to asset $i$ and the return to the portfolio divided by the variance of the portfolio return. The lower the correlation, other things being equal, the less the asset contributes to overall risk. Indeed, if the correlation is negative, it will offset existing risks and lower the portfolio standard deviation.

Portfolio theory provides a useful framework for handling multiple risks taking account of how those risks interact with each other. It is therefore of obvious use to – and is widely used by – portfolio managers, mutual fund managers and other investors. However, it tends to run into estimation and data problems. The risk-free return is not too difficult to estimate, but estimating the expected market return and the betas is often problematic. The expected market return is highly subjective and each beta is specific not only to the individual asset to which it refers, but also to our current portfolio. To estimate a beta coefficient accurately, we need data on the returns to the new asset and the returns to all our existing assets, and we need a sufficiently long data set to make our statistical estimation techniques reliable. The beta also depends on our existing portfolio and we should, in theory, re-estimate all our betas every time our portfolio changes.

For some time after it was first advanced in the 1950s, the data and calculation requirements of portfolio theory led many to see it as quite impractical. To get around some of these problems,

[6] This framework is often known as the mean–variance framework, because it presupposes that knowledge of the mean and variance (or standard deviation) of portfolio returns are sufficient to guide investors' decisions. This also implies that the variance/standard deviation of portfolio returns can be regarded as a measure of the portfolio risk. We shall have more to say on these issues in Chapter 2.

William Sharpe and others in the 1960s suggested a short-cut, the famous (or infamous) Capital Asset Pricing Model (CAPM), the essence of which was to work with betas estimated against a hypothetical market portfolio. The CAPM made portfolio theory much more practical given the data and computational power available at the time. However, it *is* a short-cut, and is potentially misleading, not least because the widespread adoption of the CAPM often leads people to talk about *the* beta for an asset, as if the asset had only a single beta. The CAPM gives us good answers if the CAPM beta estimated against the hypothetical market portfolio is close to the 'true' beta evaluated against the portfolio we actually hold, and in practice we seldom know whether it is.[7] If the two portfolios are sufficiently different, the 'true' beta might be very different from the CAPM beta, and the CAPM could be very misleading.[8] However, even in its more general form, portfolio theory can also be unreliable, not least because we might have poor parameter estimates and because the underlying assumptions on which it is based – that risks are normal, or near normal – might not be appropriate. We shall have more to say on these issues presently.

### 1.3.5 Derivatives Risk Measures

When dealing with derivatives positions, we can also estimate their risks by their Greek parameters: the delta, which gives us the change in the derivatives price in response to a small change in the underlying price; the gamma, which gives us the change in the delta in response to a small change in the underlying price (or, if we prefer, the second derivative of the derivative's price with respect to a change in the underlying price); the rho, which gives us the change in derivatives price for a small change in the interest rate; the vega, which gives us the change in derivatives price with respect to a change in volatility; the theta, which gives us the change in derivatives price with respect to time; and so on. A seasoned derivatives practitioner can make good use of estimates of these parameters to assess and manage the risks of a derivatives position, but doing so requires considerable skill. The practitioner needs to be able to deal with a number of different risk 'signals' at the same time, under real-time constraints, and the Greeks themselves can be volatile: for instance, the gamma of an at-the-money vanilla option becomes increasingly unstable as the option approaches expiry, and the volatility of vega is legendary.

However, the use of these measures for financial risk management makes sense only within the confines of a dynamic hedging strategy: the measures, and resulting hedge positions, only work against small changes in risk factors, and only then if they are revised sufficiently frequently. There is always a worry that these measures and their associated hedging strategies

[7] There are also other problems. (1) If we wish to use this short-cut, we have relatively little firm guidance on what the hypothetical portfolio should be. In practice, investors usually use some 'obvious' portfolio such as the basket of shares behind a stock index, but we never really know whether this is a good proxy for the Capital Asset Pricing Model (CAPM) market portfolio or not. It is probably not. (2) Even if we pick a good proxy for the CAPM market portfolio, it is still doubtful that *any* such portfolio will give us good results. If we wish to use proxy risk estimates, there is a good argument that we should abandon single-factor models in favour of multi-factor models that can mop up more systematic risks. This leads us to the Arbitrage Pricing Theory (APT) of Ross (1976). However, the APT has its own problems: we can't easily identify the risk factors, and even if we did identify them, we still don't know whether the APT will give us a good proxy for the systematic risk we are trying to proxy.

[8] We can also estimate risks using statistical reduced-form approaches. The idea is that we postulate a measurable relationship between the exposure-variable we are interested in (e.g., the loss/gain on our bond or FX portfolio or whatever) and the factors that we think influence that loss or gain. We then estimate the parameters of this relationship by an appropriate econometric technique, and the parameter estimates give us an idea of our risk exposures. This approach is limited by the availability of data (i.e., we need enough data to estimate the relevant parameters) and (usually) by linearity assumptions, and it is also usually limited to market risks, because we normally have data only on the prices of marketable securities. There can also be problems caused by misspecification and instability in estimated statistical relationships.

might fail to cover us against major market moves such as stock market or bond market crashes, or a major devaluation. We may have hedged against a small price change, but a large adverse price move in the wrong direction could still be very damaging: our underlying position might take a large loss that is not adequately compensated for by the gain on our hedge instrument. There is also the danger that we may be dealing with a market whose liquidity dries up just as we most need to sell. When the stock market crashed in October 1987, the wave of sell orders prompted by the stock market fall meant that such orders could take hours to execute, and sellers got even lower prices than they had anticipated. The combination of a large market move and the sudden drying up of market liquidity can mean that positions take large losses even though they are supposedly protected by dynamic hedging strategies. It was this sort of problem that undid portfolio insurance and other dynamic hedging strategies in the stock market crash, when many people suffered large losses on positions that they thought they had hedged.

## 1.4 VALUE AT RISK

### 1.4.1 The Origin and Development of VaR

In the late 1970s and 1980s, a number of major financial institutions started work on internal models to measure and aggregate risks across the institution as a whole. They started work on these models for their own risk management purposes – as firms became more complex, it was becoming increasingly difficult, and yet also increasingly important, to be able to aggregate their risks taking account of how they interact with each other, and firms lacked the means to do so. These firms were also running into problems managing risks across increasingly diverse positions. They would impose limits on traders and asset managers, but with the information and management systems available at the time, the limits were enforced on a piecemeal basis, and all sorts of inconsistencies and other undesirable effects would result: ‘good’ trades or investments would be passed over because they ran into arbitrary limits, risks were being taken with inadequate awareness of their overall effects on the firm, reducing risk in one area seldom allowed greater risk-taking elsewhere, and so on. Good capital allocation was also undermined in much the same way. But perhaps the biggest problem was the absence of integrated risk management. There was little consistency between the limits imposed (and capital allowed) and the risks being taken, and the committees entrusted with setting and managing limits and with capital allocation lacked the tools to do much better.

As firms wrestled with these problems, a consensus gradually evolved that what was needed was some sense of the probability of losses at the firmwide level. This gave rise to the notion of value at risk (or VaR), which enabled firms to get a better sense of their overall risks, and to achieve a more rational allocation of limits and capital down the various business lines.

The best known of these systems is the RiskMetrics system developed by JP Morgan. According to industry legend, this system originated when the chairman of JP Morgan, Dennis Weatherstone, asked his staff to give him a daily one-page report indicating risk and potential losses over the next 24 hours, across the bank’s entire trading portfolio. This report – the famous ‘4:15 report’ – was to be given to him at 4:15 each day, after the close of trading. To achieve this objective, the Morgan staff had to develop a system to measure risks across different trading positions, across the whole institution, and also aggregate these risks into a single risk measure. The measure used was value at risk (or VaR), or the maximum likely loss over the next trading day, and the VaR was estimated from a system based on standard portfolio theory, using estimates of the standard deviations and correlations between the returns to different traded instruments. While the theory was straightforward, making this system operational involved a huge amount of work: measurement conventions had to be chosen, data sets constructed, statistical assumptions agreed, procedures chosen to estimate volatilities and correlations, computing systems established to carry out estimations, and many other practical problems had to be resolved. Developing this methodology took a long time, but the main elements – the data systems, the risk measurement methodology, and the basic mechanics – were all in place and working by around 1990. At that point it was decided to start using the '4.15 report', and it was soon found that the new risk management system had a major positive effect. In particular, it 'sensitised senior management to risk–return trade-offs and led over time to a much more efficient allocation of risks across the trading businesses'.[9] The new risk system was highlighted in JP Morgan's 1993 research conference and aroused a great deal of interest from potential clients who wished to buy or lease it for their own purposes.

The publication of the G30 and other reports around the same time highlighted the potential usefulness of VaR systems in a much more prominent way, and the notion of VaR itself, almost unknown in 1990, rapidly became the most talked about subject in the risk management field: very soon every self-respecting financial institution wanted its own VaR system.

Meanwhile, other financial institutions had been working on their own internal models, and VaR software systems were also being developed by specialist companies that concentrated on software but were not in a position to provide data. The resulting systems differed quite considerably from each other. Even where they were based on broadly similar theoretical ideas, there were still considerable differences in terms of subsidiary assumptions, use of data, parameter estimation procedures, and many other 'details'. Besides, not all VaR systems were based on portfolio theory: some systems were built using historical simulation approaches that estimate VaR from histograms of past profit and loss data, and other systems were developed using Monte Carlo (or random number) simulation techniques.

These firms were keen to develop and strengthen their management consultancy businesses, but at the same time they were conscious of the limitations of their own models and were wary about giving too many secrets away. While most firms kept their models secret, JP Morgan decided to make its data and basic methodology available so that outside parties could use them to write their own risk management software. Early in 1994, Morgan set up the RiskMetrics unit to do this and the RiskMetrics model – a simplified version of the firm's own internal model – was completed in eight months. In October that year, Morgan made its RiskMetrics system and the necessary data freely available on the internet: outside users could now access the RiskMetrics model and plug their own position data into it.

This bold move attracted a lot of attention, and the resulting public debate about the merits of RiskMetrics was useful in further raising awareness of VaR and of the issues involved in building and operating VaR systems.[10] Making the RiskMetrics data available also gave a major boost to the spread of VaR systems by giving software providers and their clients access to data sets that they were often unable to construct themselves. It also encouraged many of

[9] Guldimann (2000), p. 57.

[10] A good example is the exchange between Longerstaey and Zangari (1995), on the one hand, and Lawrence and Robinson (1995a), on the other, on the safety or otherwise of RiskMetrics. The issues covered in this debate – the validity of underlying statistical assumptions, the estimation of volatilities and correlations, and similar issues – go right to the heart of risk measurement, and will be dealt with in more detail in later chapters.

the smaller software providers to adopt the RiskMetrics approach or make their own systems compatible with it.

The subsequent adoption of VaR systems was very rapid, first among securities houses and investment banks, and then among commercial banks, pension funds and other financial institutions, and non-financial corporates. As the models proliferated, the VaR concept also became more familiar, and by the mid-1990s the VaR had already established itself as *the* dominant measure of financial risk – a meteoric rise when one considers that the VaR was almost unknown only a few years earlier. The state of the art also improved rapidly. Developers and users became more experienced; the combination of plummeting IT costs and continuing software development meant that systems became more powerful, much faster, and much more sophisticated; VaR systems were extended to cover more types of instruments; and the VaR methodology itself was extended to deal with other types of risk besides the market risks for which VaR systems were first developed, including credit risks, liquidity (or cash flow) risks and operational risks.

**Box 1.2** Portfolio Theory and VaR

In some respects VaR is a natural progression from earlier portfolio theory (PT). Yet there are also important differences between them:

- PT interprets risk in terms of the standard deviation of the return, while VaR approaches interpret it in terms of the maximum likely loss. The VaR notion of risk – the VAR itself – is more intuitive and (arguably) easier for laypeople to grasp.
- PT presupposes that P/L or returns are normally (or near normally) distributed, whereas VaR approaches can accommodate a very wide range of possible distributions. VaR approaches are therefore much more flexible.
- VaR approaches can be plausibly applied to a much broader range of risk problems: PT theory is (for the most part) limited to market risks, while VaR approaches can be applied much more flexibly to credit, liquidity and other risks as well.
- The variance–covariance approach to VaR has the same theoretical basis as PT – in fact, its theoretical basis *is* portfolio theory – but the other two main approaches to VaR (i.e., the historical simulation and simulation approaches) do not. VaR systems can be based on a wider range of estimation methods.

### 1.4.2 Attractions of VaR

So what is VaR, and why is it important? The basic concept was nicely described by Linsmeier and Pearson:

> Losses greater than the value at risk are suffered only with a specified small probability. Subject to the simplifying assumptions used in its calculation, value at risk aggregates all of the risks in a portfolio into a single number suitable for use in the boardroom, reporting to regulators, or disclosure in an annual report. Once one crosses the hurdle of using a statistical measure, the concept of value at risk is straightforward to understand. It is simply a way to describe the magnitude of the likely losses on the portfolio.[11]

[11] Linsmeier and Pearson (1996), p. 3.

The VaR has a number of significant attractions over traditional risk measures:

- VaR provides a *common* consistent measure of risk across different positions and risk factors. VaR can be applied to any type of portfolio, and it enables us to compare the risks of different portfolios. It enables us to measure the risk associated with a fixed-income position, say, in a way that is comparable to the risk associated with an equity position. In this the VaR is a distinct improvement over traditional methods (e.g., duration and convexity approaches only apply to fixed-income positions, Greek risk measures only apply to derivatives positions, and portfolio theory approaches apply to equity and similar (e.g., commodity) positions).
- VaR enables us to *aggregate* the risks of subpositions into an overall measure of portfolio risk, and in doing so take account of the ways in which different risk factors interact (or correlate) with each other. This is another attraction of the VaR, because most traditional risk measures do not (easily) allow for the 'sensible' aggregation of component risks.
- VaR is *holistic* in that it takes full account of all driving risk factors, whereas many traditional approaches either only look at risk factors one at a time (e.g., Greek measures) or else resort to simplifications to collapse multiple risk factors into one (e.g., duration–convexity approaches collapse the spot-rate curve into a single yield, and CAPM approaches collapse different equity returns into a single 'market' return). VaR is also holistic in another sense: it focuses assessment on a complete portfolio, often at the firmwide level, and not just on individual positions in it.
- VaR is *probabilistic*, and gives a risk manager useful information on the probabilities associated with specified loss amounts. By comparison, many traditional measures (e.g., duration–convexity, Greeks, etc.) only give us the answers to 'what if' questions and don't give an indication of likelihoods.
- VaR is expressed in the simplest and most easily understood *unit of measure*, namely, 'lost money'. Many other measures are expressed in less transparent units of measure (e.g., average period to cash flow, etc.). Hence, the VaR is expressed in terms of a unit that is easier to convey.

These are significant attractions, which do a lot to explain why VaR became so popular.

VaR information can be used in many ways (albeit, not without problems, some of which we will address in due course): (1) Senior management can use it to set their overall risk target, and from that determine risk targets and position limits down the line. If they want the firm to increase its risks, they would increase the overall VaR target, and vice versa. (2) VaR can be used to determine capital requirements, both at the firmwide and business-unit level: the riskier the activity, the higher the VaR and the greater the capital requirement. VaR can also be used to specify the position limits imposed on business units. (3) VaR can be useful for reporting and disclosing purposes, and firms increasingly make a point of reporting VaR information in their annual reports.[12] (4) VaR-based decision rules can guide investment, hedging, trading and portfolio management decisions, and do so taking account of the implications of alternative choices for the portfolio risk as a whole.[13] It can also be used to carry out portfolio-wide (or macro) hedging strategies that are otherwise difficult to implement.[14] (5) VaR information can be used to provide new remuneration rules for traders, managers and other employees that take

[12] For more on the use of VaR for reporting and disclosure purposes, see, e.g., Dowd (2000b), Jorion (2002), Moosa and Knight (2001) or Woods *et al.* (2004).

[13] VaR-based decision rules are covered more fully in, e.g., Dembo (1997), Lucas and Klaassen (1998), Dowd (1999a,c) and Sentana (2001).

[14] Such strategies are explained in more detail in, e.g., Litterman (1996), Kuruc and Lee (1998) and Dowd (1999a).

account of the risks they take, and so discourage the excessive risk-taking that occurs when employees are rewarded on the basis of profits alone, without any reference to the risks they took to get those profits. (6) Systems based on VaR methodologies can be used to measure other risks such as credit, liquidity and operational risks. In short, VaR can help provide for a more consistent and more integrated approach to the management of different financial risks, and so lead to better risk management overall.

### 1.4.3 Criticisms of VaR

Most risk practitioners embraced VaR with varying degrees of enthusiasm, but there were also those who warned that VaR had deeper problems and could be dangerous.

A key issue was the validity or otherwise of the statistical and other assumptions underlying VaR, and both Nassim Taleb (1997a,b)[15] and Richard Hoppe (1998) were critical of the naïve transfer of mathematical and statistical models from the physical sciences where they are well suited to social systems where they were often invalid. Such applications often ignore important features of social systems – the ways in which intelligent agents learn and react to their environment, the non-stationarity and dynamic interdependence of many market processes, and so forth – features that undermine the plausibility of many models and leave VaR estimates wide open to major errors.

A related argument was that VaR estimates are too imprecise to be of much use, and empirical evidence on this issue is worrying, as it suggests that different VaR models can give vastly different VaR estimates (see, e.g., Beder (1995)). To make matters worse, VaR models are also exposed to considerable implementation risk as well – so even theoretically similar models could give quite different VaR estimates because of differences in the ways in which the models were implemented (Marshall and Siegel (1997)).

The danger here is obvious: if VaR estimates are too inaccurate and users take them seriously, they could take on much bigger risks and lose much more than they had bargained for. As Taleb put it, 'You're worse off relying on misleading information than on not having any information at all. If you give a pilot an altimeter that is sometimes defective he will crash the plane. Give him nothing and he will look out the window.'[16] Such criticism is not easy to counter.

A deeper problem is that risk is endogenous: if VaR is used to control or remunerate risk-taking, those being controlled will respond to VaR constraints in the pursuit of their own interest. For example, traders will have an incentive to seek out and trade positions where risk is over- or underestimated (Ju and Pearson (1999)). They will therefore take on more risk than suggested by VaR estimates that fail to take account of how traders or other affected parties will respond – so our VaR estimates will be biased downwards – and the evidence suggests that the magnitude of these underestimates can be very substantial. VaR limits might also encourage traders to respond by taking more low-probability, high-impact risks, their motivation being that such risks are likely to pay off (because they increase earnings in normal times), and the occasional very high loss is allowable because it meets the VaR constraint: VaR limits can encourage traders to write deep-out-of-the-money options against their employers' assets.

[15] Taleb was also critical of the tendency of some VaR proponents to overstate the usefulness of VaR. He was particularly dismissive of Philippe Jorion's (1997) claim that VaR might have prevented disasters such as Orange County. Taleb's response was that these disasters had other causes – especially, excessive leverage. As he put it, a Wall Street clerk would have picked up these excesses with an abacus, and VaR defenders overlook the point that there are simpler and more reliable risk measures than VaR. Taleb is clearly right: any simple duration analysis should have revealed the rough magnitude of Orange County's interest-rate exposure. The root problem with Orange County was not the absence of VaR, as such, but the absence of risk management.

[16] Taleb (1997a), p. 37.

There are also good reasons to think that the use of VaR as a regulatory constraint might discourage good risk management practices.[17]

Others suggested that the use of VaR might destabilise the financial system. Thus, Taleb (1997a) pointed out that VaR players are dynamic hedgers, and need to revise their positions in the face of changes in market prices. If everyone uses VaR, there is a danger that this hedging behaviour will make uncorrelated risks become very correlated – and again firms will bear much greater risk than their VaR models might suggest. Poorly thought through regulatory VaR constraints can also destabilise the financial system by inducing banks to increase their risk-taking: for example, a VaR cap gives risk managers an incentive to protect themselves against mild losses, but not against larger ones in excess of VaR. VaR regulatory constraints can also exacerbate cyclical effects, and so aggravate financial crises, or even bring them about.[18]

Proponents of VaR could respond that many of these criticisms are not specific to VaR as such, but would also apply (in varying extents) to other risk measures as well. They could also argue that some of these problems were due to the misuse of VaR rather than to the VaR itself; for example, they could claim with considerable justification that some of the problems regarding the regulatory use of VaR were due to the failings of the regulatory system. But not all these criticisms can be answered, and new problems were also coming to light that would, in time, not only undermine the VaR's position as the dominant measure of financial risk, but also destroy the VaR's claim to be regarded as a 'proper' measure of financial risk in the true sense of the term. This takes us to the next major development in financial risk management, the theory of coherent financial risk measures, and this is perhaps best deferred to the next chapter.

---

[17] See, e.g., Danielsson and Zigrand (2001).

[18] See Danielsson (2002), Danielsson and Zigrand (2001), Basak and Shapiro (2001) and Danielsson *et al.* (2001).

# Appendix
# Types of Market Risk

As explained in the text, market risk is the risk of loss associated with unexpected movements in market prices or rates, and is to be distinguished from other types of risk such as credit risk (involving possible losses from default events) and operational risk (involving losses from people or systems failures). However, market risk cannot be entirely divorced from these other types of risk, and can sometimes be created by them. For example, credit (i.e., default) events can lead to changes in market prices or rates (e.g., they might affect bond spreads or bond prices, or the prices of credit derivatives) and so trigger market losses. Operational events can also lead to market losses, a good example being the collapse in the value of the shares in Barings Bank to £1.

There are many different categories of market risk, and the most common categories include:

- Equity risks: the market risks associated with positions in equity markets.
- Fixed-income risks: the risks associated with positions in fixed-income instruments (e.g., bonds and structured notes) and interest-sensitive instruments generally (e.g., interest-rate derivatives).
- FX risks: risks associated with foreign and cross-currency positions.
- Commodity risks: risks on agricultural, energy, metals, and similar positions.
- Miscellaneous market risks: risks on weather, temperature, and catastrophe instruments, underwriting risks on insurance portfolios, the risks of property positions, and so on.

Market risks also differ enormously in terms of the types of players who are taking the relevant risks, the reasons they take them, the horizons to which they operate, and the valuation methods used to establish profits or losses. These factors are all interrelated, and are also related to the types of instrument involved and the risk management methods used.

One way to look at these issues is from the perspective of context. To begin with, there is a trading context. Players in a trading context might be securities houses and similar firms who trade equities, futures, options, bonds, and so on. They would operate on shorter horizons, and often on horizons of a trading day, or of one or two weeks at five trading days to a week, and some traders even operate on intra-day horizons. They would operate on both organised exchanges and relatively liquid OTC (over-the-counter) markets. The sophistication of the instruments traded will vary enormously, from very simple vanilla instruments at one end to highly sophisticated instruments at the other. They would also operate with advanced market risk measurement systems, and they would typically value their positions (and so determine gains and losses) using mark-to-market valuation methodologies. In addition, these traders would often aim to beat benchmarks, although the type of benchmark used would depend on the market. Performance evaluation would sometimes involve an adjustment for risks taken, and sometimes not.

There is also an investment context, which would include many of the activities of banks. In this context, firms operate to longer horizons and deal in instruments that often involve significant credit risk (i.e., so market and credit risks are very closely related), and valuation is difficult because secondary markets are thin or non-existent. These operators would include many banks, who would be concerned with medium-term horizons, and insurance and pension funds, with the latter in particular being concerned with very long-term horizons. Risk assessment is often not as sophisticated as in a trading context, and it is less common for performance evaluation to be risk-related. There are also other contexts. These include a treasury (or payments-related) context, in which financial institutions are involved in large-scale electronic funds transfer (EFT) through payments systems such as SWIFT, CHIPS, etc. The amounts involved are massive, and there are major concerns with intra-day credit exposures and the payments at risk. There is also an insurance context, in which insurance companies trade market risk exposures obtained through their underwriting activities, and a property context, in which operators buy and sell exposures to property prices.

There are related differences in terms of valuation methodologies. Essentially, market positions can be valued in one of three main ways. The first and in many ways the best valuation methodology is mark-to-market: positions are revalued periodically at current market prices, and investors realise their profits/losses with each such revaluation. This methodology works well with liquid markets and fairly clear end-of-day market prices based on real market trades. The classic example is where positions on organised markets are marked-to-market at the end of each trading day, and traders' margins are adjusted accordingly so that all gains or losses are realised immediately. The applicability of mark-to-market therefore depends on having a liquid market for the instruments concerned, and mark-to-market is closely associated with short-term trading horizons.

A second valuation methodology is mark-to-model: positions are revalued against hypothetical prices generated by a model. This is often used as a substitute for mark-to-market in situations where mark-to-market is not feasible, so 'current' market prices either do not exist or are unreliable because trading is too thin. Mark-to-model is not as reliable as mark-to-market and depends crucially on the validity of the models used, including the ways in which they are calibrated. The models used therefore have to be carefully chosen and calibrated, and regularly reviewed. Mark-to-model is also open to abuse if the models are poorly chosen or tampered with, and there have been numerous cases of hidden losses and fraud related to the misuse of valuation models (e.g., a common one being to fiddle the values of volatility parameters in option-pricing models to artificially boost mark-to-model valuations and, of course, to boost the bonuses that go with them).

A third approach to valuation is book valuation using standard accounting methods (i.e., GAAP). In theory, book valuation is highly questionable and has many well-known deficiencies when used to value market positions: historical costs often give poor indications of current values; book valuation treats depreciation in a crude and ad hoc manner, and gives scope for losses to be hidden, and earnings are often excessively smoothed. However, it can be applied in circumstances where other approaches are not feasible or are themselves even less reliable, and one of its advantages is precisely that its deficiencies are well understood. Managers using book methods can therefore make their own discretionary adjustments to take account of any biases or other problems they perceive in the book values.

We can also differentiate market risks in terms of where they stand along a liquidity spectrum that recognises both the common elements among different risks as well as their differences from each other.[19] There are four key points along this spectrum:

[19] This was suggested and explained further by Drzik (1996).

- *Smooth markets*: These are textbook liquid markets with large numbers of participants and high turnovers. Standard examples are currency markets or markets for US Treasuries. Positions are easily valued by marking to market and risk is assessed on the basis of VaR supplemented by stress tests.
- *Choppy markets*: These markets are less deep and less liquid, and have fewer participants and lower turnover. Examples are OTC markets in equities. Participants tend to use similar methods as participants in smooth markets, but these methods are less reliable in choppy markets because of liquidity and valuation problems.
- *Icy markets*: These markets are even thinner and less liquid, and secondary markets exist but are very limited. Trades tend to be negotiated rather than screen based, and prices are often not transferable across deals. Pricing is usually mark-to-model supplemented by some marking to market with adjustments for liquidity and other concerns. Risk calculations are mainly carried out by loss modelling, with some VaR analysis adjusted for liquidity risks.
- *Frozen markets*: These markets are extremely thin and there are few, if any, secondary markets. Assets are usually bought to hold to maturity and products are highly tailored. Pricing is highly judgemental, and often based on reserve-adjusted book values supplemented by marking to model and relatively arbitrary allowances for illiquidity. Risk evaluation is highly problematic.

Each point on this continuum is characterised by certain features:

- *Valuation*: Market prices are good valuation guides in smooth markets, somewhat less useful in choppy markets, even less use in icy markets, and of no use in frozen ones. We therefore rely almost entirely on mark-to-market valuation for smooth markets. When markets become choppy we make more use of mark-to-model valuation and to some extent judgemental methods. With icy markets the balance shifts further towards judgemental and book methods, and with frozen markets we rely on little else.
- *Risk control*: Risk control in smooth markets is straightforward, since profits and losses are easy to ascertain, and relies mainly on limits backed up by *ex post* monitoring and performance evaluation. Risk control in choppy markets is similar except for the need to pay more attention to issues of liquidity and unrealised gains/losses. In icy and frozen markets we have to devote even more attention to these issues, and also to systems to mitigate or stop losses from accumulating.
- *Remuneration*: As markets become less liquid, we should place increasing emphasis on deferred compensation. Deferment enables us to reduce the compensation of individuals whose decisions produce losses that only become apparent later. The prospect of such penalties gives them more incentive to act responsibly.

In short, market risks are complex and multidimensional. In attempting to estimate them, it is very important to keep in mind that any estimates we make are very dependent on the valuation methodologies used, and these are closely related to market liquidity and the assumptions embodied in the ways in which valuation methodologies are chosen and implemented, and these assumptions may not be valid. Thus, estimates of market risk are inevitably subject both to *liquidity risk* and to more general *model risk*.