Chapter 1 - Market Risk Introduction

Michel Crouhy, Dan Galai and Robert Mark

Market risk is the risk that changes in financial market prices and rates that will reduce the value of a security or portfolio. Price risk can be decomposed into a general market risk component (the risk that the market as a whole will fall in value) and a specific risk component, unique to the particular financial transaction under consideration. To quantify market risk, we need pricing models, algorithms, and statistical models. Pricing models and algorithms relate market prices to risk factors, and statistical models link volatility of market prices to volatility of market risk factors.

Some risk factors are intuitive because they are directly observable, such as stock prices and exchange rates. Other risk factors are less intuitive and more difficult to measure such as interest rates, volatility of interest rates, or correlations between factors, which are not directly observable and should be derived from a combination of market prices and a pricing model. For example, we observe a bond price, but must derive interest rates from the bond price, considering the term of the bond.

Market risk is usually divided into four major types: equity price risk, interest rate risk, foreign exchange risk, and commodity price risk.

Typology of Market Risk Exposures

Interest rate risk is risk that the value of a fixed-income security, say a bond, swap, or swaption, will increase or decrease because of changes to market interest rates. First, an underlying interest rate curve should be specified: say a government treasury curve, a swap curve, or a curve representing a bond rating. There is a need to derive a term structure of interest rates for each underlying class of fixed-income instruments. Second, risk points on the curves should be specified, and interpolation should be applied that ensures no arbitrage opportunities are created. Many different kinds of exposures can arise from differences in maturities and reset dates of instruments and cash flows that are asset-like (i.e., “longs”) and those that are liability-like (i.e., “shorts”).

Equity price risk is associated with the volatility of stock prices and indices. The general market risk of equity refers to the sensitivity of an equity instrument or portfolio value to a change in the level of broad stock market indices such as the S&P 500.The specific or idiosyncratic risk of equity refers to that portion of a stock’s price volatility determined by characteristics specific to the firm, such as its line of business, the quality of management or a breakdown in its production process.

Data for equities are readily available over an extended period. For example, monthly rates of returns for NYSE securities are available since 1926. For the last two decades, we can get tick-by-tick data, at least for U.S. markets. The availability of data led to an abundance of equity research.

Foreign exchange risk arises from open or imperfectly hedged positions, in particular, foreign currency denominated assets and liabilities leading to fluctuations in profits or values as measured in local currency. These positions may arise as a natural consequence of business operations, rather than from any conscious desire to take trading positions in a particular currency. This is a major risk exposure of corporations involved in international trade. Foreign exchange risk can eradicate the expected return from expensive, cross-border investments, and simultaneously place a company at a competitive disadvantage in relation to foreign competitors. It might also generate huge operating losses.

Although it is important to acknowledge exchange rates as a risk factor, risk assessment of foreign-exchange transactions requires knowledge of the dynamics of domestic and foreign interest rates, and of spot exchange rates.

For major currencies (i.e., EUR, USD, GBP, CHF, AUD, and JPY), there is a very liquid and deep OTC market for cash and forward transactions and derivative transactions such as currency swaps and currency options. Only a small fraction of foreign-exchange transactions is conducted on exchanges - primarily currency futures and options traded on the IMM (International Monetary Market) of the CME (Chicago Mercantile Exchange) since 1972 after the collapse of the Bretton Woods Agreement in August 1971. Options on spot exchange rates are traded on the CBOE (Chicago Board Options Exchange).

Commodity price risk is the most difficult risk to assess since most commodities are traded in markets in which the concentration of supply is in the hands of a few suppliers that have the market power to influence prices. Commodity markets are also exposed to breakdowns in delivery (e.g., maintenance of oil refineries and gas pipelines). Companies may store commodities to hedge market price fluctuations. The ease and cost of storage, which varies considerably across commodities, impacts price volatility.

Commodities can be classified according to their characteristics as follows: hard commodities, or non-perishable commodities, the markets for which are further divided into precious metals (e.g. gold, silver and platinum), which have a high price/weight value, and base metals (e.g. copper, zinc and tin); soft commodities, or commodities with a short shelf life that are hard to store, mainly agricultural products (e.g. grains, coffee and sugar); and energy commodities, which consist of oil, gas, electricity ,and other energy products.

There are two levels of market risk management: micro and macro. The micro risk management of a position requires calculation of price sensitivities of positions with respect to many risk factors. For example, traders are in charge of hedging exposures (i.e., sensitivities to individual

risk factors) within risk limits set by the risk management group. This granular approach is useful at the trading desk level. However, senior managers of a financial institution, or a portfolio manager, might be interested in a macro approach that relies on an aggregate measure of risk, based on the overall distribution of portfolio values at a given horizon in the future. Value-at-Risk (VaR) is the archetype of an aggregate risk measure.

Asset-liability Management

Asset-liability management (ALM) is structured decision-making for matching, and deliberately mismatching, the mix of assets (e.g., loans) and liabilities (e.g., deposits) on a firm's balance sheet. ALM is particularly critical to financial institutions such as commercial banks, savings and loans, insurance companies, and pension funds. Banks, for example, are involved in collecting deposits and lending to retail and corporate clients. This financial intermediation generates two types of imbalances: first, between the amount of funds collected and lent, and second, between the maturities and interest rate sensitivities of the sources of funding and lending to clients.

These imbalances drive the net worth of the bank, and its risk profile. For example, deposits generally have a shorter maturity than loans, so the net worth of many banks benefits from a fall in interest rates; the bank pays less interest to its depositors but continues for a period to receive the higher rate from its borrowers. Conversely, the net worth of the same bank deteriorates if interest rates rise. If this downside risk is not managed, it can lead to insolvency in individual institutions or even entire banking industries. Likewise, the mismatching of assets and liabilities almost inevitably leads to a degree of liquidity risk; the greater the mismatch, the more difficult it is to ensure the institution has cash on hand to fulfill its commitments immediately in any conceivable circumstance (e.g., the return of on-demand deposits).

Banks' earnings are particularly sensitive to changes in interest rates and the cost of funds but many ALM principles apply equally to corporations outside the financial sector whose assets and liabilities are sensitive to market risk factors. The three goals of ALM are to:

- Stabilize net interest income (NII), that is, the difference between the amount the bank pays out in interest for funding and the amount it receives from holding assets such as loans (as measured by accounting earnings);
- Maximize shareholder value or net worth (NW), reflected by long-term economic earnings;
- Ensure the bank does not assume too much risk from the mismatching of maturities and amounts between assets and liabilities and from funding liquidity risk (the danger banks will be unable to raise funds quickly and cheaply enough to fulfill its obligations and remain solvent).

Funds Transfer Pricing

The rationale for funds transfer pricing (FTP) is that there are economies of scale and scope when centralizing management of interest-rate risk. Business units have no control over the dynamics of yield curves and other market indices such as the prime rate, so the objective of funds transfer pricing is to remove the non-controllable interest-rate risk from business results. The funds transfer pricing system charges each business unit that requires funds, the cost of funding its activity (e.g., the funding cost to make loans) and hedging its interest-rate risk. The funds transfer pricing system is also used to credit each business unit for its activity to supply funds (e.g., branch that raises deposits).

Each business unit is then able to secure its profit margin at the time of origination of its products (e.g., mortgages), and can focus on developing and managing the business side of its activity and the credit quality of its portfolio (i.e., credit risk remains with the business unit). The transfer pricing system needs to spell out clearly if interest rate risks such as basis risk (e.g., the spread between the prime rate and LIBOR for variable-rate loans indexed on the prime rate) and options risk (e.g., commitment risk for mortgages) remains with the business unit.

The issue remains: what is the appropriate cost of funds to charge to the business units? A matched-maturity FTP is the typical approach banks use. A consistent framework for constructing transfer prices should be based on consistent rules and principles. For example, the framework should include principles that a bank can adopt to determine the interest rate sensitivity of the assets and liabilities, and decide into which periods (or buckets) the assets and liabilities should be placed so they can be matched to facilitate match funding. To conduct ALM and accurate FTP, banks need to place assets, liabilities, and off-balance-sheet items into buckets in terms of their known interest rate sensitivities (for match funding purposes) and maturities (for funding liquidity purposes). The activities are distinct.

Industry Best Practices

A comprehensive market risk management framework calls for policies (e.g., risk appetite and risk disclosure policies), methodologies (e.g., VaR, stress testing, and RAROC methodologies) and infrastructure (e.g., risk staff and risk systems infrastructure) that enable management of all components of market risk in the trading book, as well as of gap risk in the treasury book. A well-designed system infrastructure enables the production of a variety of critical reports that make market risk transparent. These reports are provided to the necessary committees and personnel throughout the governance structure. The governance structure sets policies on risk appetite, and risk management monitors risk appetite within limits determined by policy.

Best-practice institutions set business-line risk limits in terms of risk in both normal and stress markets. These limits ensure that individuals do not take more risk than they are allowed. Risk limits should be aggregated up through the firm, from the business line at the trading desk level

to the top of the corporation. The drill-down capability of a market risk measurement system allows market risk managers to identify the unit taking the most market risk, and the type of market risk to which the entire bank is most exposed.

The limitations of many measures of market risk have been understood for years, but they played a role in obfuscating risks during financial crises. However, each time there is turmoil in the world's markets, the limitations of sophisticated measures of market risk are revealed. For example, VaR models are typically based on the assumption that parameters such as volatilities and correlations are stationary (i.e., they do not change in value during the period risk is measured). This assumption is often proven wrong during extreme market conditions, making VaR an unreliable measure of risk at the moment robust risk analytics are required most.

Each crisis reemphasizes the importance of using multiple risk-measurement tools, including stress tests and scenario analyses, and of achieving the right blend of quantitative rigor and qualitative assessment. Using a range of risk measures helps because each approach has limitations and strengths. VaR cannot easily capture the impact of significant disruptions in liquidity, prices, volatility, and correlations. VaR also struggles to capture strong non-linearities in market risk of the type observed in complex structured products.

Stress testing, which consists of shocking risk factors, is a bottom-up, static process. It assesses the immediate impact on risk positions of these shocks. Scenario analysis is dynamic in nature and a top-down process. Macroeconomic scenarios, which unfold over several periods, say two to three years, are specified. These scenarios are translated into fully consistent trajectories of risk drivers that affect various portfolios. This approach helps uncover extreme situations that might lead to insolvency and liquidity crises; it provides a powerful framework to fine-tune the risk appetite of the firm and design contingency plans that meet regulatory requirements under extreme conditions. This is precisely the spirit of the Dodd Frank Stress Test (DFAST). DFAST requires banks with total consolidated assets of more than $10 billion to conduct annual stress tests. The purpose of DFAST is to assess quantitatively how bank capital levels would fare in a stressful situation.

Content of Market Risk Section

Crouhy, Galai, and Mark discuss market risk management governance and management in Chapter 2. The authors describe post-crises risk governance concerns in financial institutions following the 2007 to 2009 financial crisis. They point out that many of the risk governance concerns, in general, also hold for market risk. The authors describe in detail the history of prudential market risk regulation, and show where the capital standards for market risk were weak in the run up to the financial crisis. They describe the roles and responsibilities in practice of board committees and senior management committees. They also review in detail the roles and responsibilities of the risk management and audit functions. The authors provide insights into model risk governance, steps necessary to mitigate model risk, and valuation in a mark-to-market low trading liquidity world.

Max Wong and Changwei Xiong describes the current and evolving market risk measurement tools in Chapter 3. The basic VaR approach is described along with more advanced VaR models such as extreme value theory. He extends the VaR measure to fat-tailed distributions characterized by extreme events. Wong provides the detailed analytical formula for the standard VaR as well as for advanced univariate and multivariate VaR models.

An in-depth discussion of equity, interest rate, foreign exchange, and commodity products is given in Chapter 4 by Douglas Bongartz-Renuad, and in Chapter 5 by Jan-Peter-Onstwedder. Chapter 4 summarizes key developments in the financial markets over the past four decades. Bongartz-Renuad offers product descriptions for both cash and derivatives markets for equity, interest rates, and foreign exchange products. Chapter 5 discusses categories of commodity products (e.g., energy, metals, and agricultural) and the many players in commodities markets (e.g., producers, processors, consumers, and traders). Onstwedder introduces product descriptions for both cash and derivatives commodity products. David Rowe discusses stress testing in Chapter 6, including the multiple challenges in describing risks for both normal and stress markets. Rowe provides a three-pronged approach to stress testing using selected historical events, reverse stress testing, and structural imagination to assess socioeconomic and geopolitical conditions to define dangerous scenarios. He emphasizes that current risk management resources often fall short of what is required to comprehensively calculate risk in stress markets.

Justin McCarthy provides insights into the historical and current evolution of ALM and funding liquidity management in Chapter 7. In Chapter 8, Douglas Bongartz-Renaud discusses the role FTP plays in ALM. FTP is an essential component of ALM, and is a controversial issue in most organizations since it affects the measured profitability of various business lines. McCarthy talks about the overall causes of the 2007 to 2009 financial crisis, and describes how the ALM function should respond to a financial crisis. He reviews how organizations should manage interest rate risk and funding liquidity risk from a balance-sheet, net-income, and cash-flow perspective. McCarthy provides details on the measurement of ALM related interest rate and liquidity risks. He also discusses Basel Committee principles for the management and supervision of interest rate risk and liquidity risk. McCarthy reveals how to manage a funding liquidity crisis, arguing for a dynamic contingency funding planning approach. Bongartz-Renuad describes a hierarchy and evolution of FTP methods, detailing how to construct an FTP system. He also talks about the rationale for having an FTP system. Bongartz-Renuad describes economies of scale associated with centralizing the management of interest rate risk. He demonstrates the link between components of bank net interest margin and the FTP system. Bongartz-Renuad also discusses how an FTP system can be used to identify how a bank makes money.

Chapter 2 – Market Risk Governance and Management

Michel Crouhy, Dan Galai and Robert Mark

Introduction[1]

Risk governance[2] is a continuing concern worldwide. This chapter focuses on the **market risk governance and management framework for the trading part** of a financial institution. Volume I of the Handbook, Risk Frameworks & Operational Risk, establishes the requirements for **sound overall risk governance**.

In *Volume III Book 1*, Chapter 2: Risk Governance, Prof. Mainelli discusses the board level role in supporting overall risk governance. The author points out that a Board Risk Committee must cooperate with risk management committees (e.g., through cascading of risk limits). A business risk committee, comprised of senior experts from all key control functions, should be an early reviewer of all material business proposals.

In *Volume III Book 1*, Chapter 3: The Risk Management Framework, Coleman discusses key elements of a best-practice risk framework. These elements include risk capacity, risk appetite, risk policies, risk pricing, culture, and incentives. The author reviews processes an institution should implement to determine its risk capacity and risk appetite. Application of such an assessment to a division within the institution is comprised of a number of Board decisions, including, but not limited to, an agreed business strategy, risk appetite, and capital allocation. These decisions should enable the risk function to determine the policy and limit the framework that will support the strategy. Coleman emphasizes that a policy framework must be implemented verifiably, meaning policy statements should be supported by controls whose effectiveness can be tested periodically by the Bank's audit function.

The first decade of the millennium saw major waves of corporate failures in financial and non-financial sectors, attributed in part to failures of market risk governance. In many cases, boards were provided with misleading information or there was a breakdown in the process by which information was transmitted to the board and shareholders. The breakdowns often involved financial engineering and nondisclosure of economic risks, and outright fraud.

---

[1] Material is substantially drawn from Crouhy, Galai and Mark, 2nd Edition of the Essentials of Risk Management, McGraw Hill, 2013

[2] "Corporate governance involves a set of relationships between a company's management, its board, its shareholders and other stakeholders. Corporate governance also provides the structure through which the objectives of the company are set, and the means of attaining those objectives and monitoring performance are determined." Preamble, OECD Principles of Corporate Governance, 2004, p.11

This first wave of Scandals ranging from Enron to the 2007 to 2009 financial crisis, led a wave of reforms, including legislation in the United States and reform of corporate codes in Europe, which were designed to mend perceived failures in corporate governance practices and especially improve financial controls and reporting. A striking feature of these reforms is that they penalize inattention and incompetence as much as deliberate malfeasance. In the United States, an important mechanism of reform was the changes to stock exchange rules, described in Box 1.

BOX 1: U.S. EXCHANGES TIGHTEN THE RULES

In January 2003, the U.S. Securities and Exchange Commission issued a rule, directed by the Sarbanes-Oxley Act, that requires U.S. national securities exchanges (e.g., NYSE, Amex, and Nasdaq) to ensure that their securities listing standards conform to existing and evolving SEC rules. These standards cover a number of areas critical to corporate governance and risk management, including:

- Composition of the board of directors (e.g., the majority must be independent directors;
- Establishment of a corporate governance committee with duties such as development of broad corporate governance principles and oversight of evaluation of the board and management;
- Duties of the compensation committee (e.g., CEO compensation aligns with corporate objectives;
- Activities of the audit committee (e.g., review external auditors' reports describing the quality of internal control procedures, and adopt and disclose corporate governance guidelines and codes of business conduct).

A note of frustration characterized the debate about market risk corporate governance following the 2007 to 2009 crisis. Would it do any good to reform market risk governance with detailed legislation and new rules? Others argue that a principles-based approach might work better, given that regulators in the banking industry have already set out some of the key principles of improved risk governance in Pillar II of Basel II. Key components of prudential market risk regulation are described in the Appendix.

Table 1 sets out some of the key risk governance concerns in financial institutions following the 2007 to 2009 financial crisis. Many of the risk governance concerns, in general, also hold true for market risk. We return to many of these themes throughout this chapter.

Table 1: Key post-crisis, risk governance concerns — The banking industry

Stakeholder priority	Enquiries into the 2007 to 2009 financial crisis found that there was little focus in some firms on controlling market risks in the tail and considering truly worst-case outcomes. This led to debates about the uniquely complicated set of stakeholders in banking institutions and how this should affect corporate governance structures. In addition to equity, banks have very large amounts of deposits, debt, and implicit guarantees from governments. Depositors, debt holders, and taxpayers have a much stronger interest in minimizing the risk of bank failure than most shareholders, who often press for short-term results.
Board composition	The crisis reignited a long-term debate about how to ensure that bank boards will contain the right balance of independence, engagement, and financial-industry expertise. However, analyses of failed banks do not show a clear correlation between a predominance of expert insiders or independents and either failure or success.
Board risk oversight	One post-crisis trend has been a realization that boards need to become much more involved in market risk oversight. This means educating boards on market risk and ensuring they maintain a direct link to the market risk management infrastructure (e.g., giving CROs direct reporting responsibilities to the board and more generally, re-empowering market risk managers).
Risk appetite	Regulators push banks to set out a formal board-approved, market-risk appetite that defines the firm's willingness to take market risk and tolerate threats to solvency. This can then be translated into a set of market risk limits. Engaging the board in the market risk limit setting process helps to ensure that the board thinks clearly about the firm's risk-taking and what this means for daily risk decisions. However, defining market risk appetites and translating them into market-risk limit frameworks remains a work in progress.
Compensation	One of the key levers of the board in determining bank behavior on market risk is its control over compensation schemes. Some banks have begun to institute reforms such as making bonuses a smaller part of the compensation package, introducing bonus clawbacks, and deferring payments to capture longer-term risks and similar measures. Boards have a particular duty to examine how pay structures exacerbate market risk-taking and whether risk-adjustment mechanisms capture all the key, long-term market risks.

We use the example of an archetypal bank to answer three questions:

- How does best-practice corporate governance relate to best-practice, market-risk management?
- How do boards and senior executives organize the delegation of market-risk management authority through key committees and risk executives?
- How can agreed-on market risk limits be transmitted down the line to business managers in a way that can be monitored and that makes sense in terms of daily business decisions?

The purpose is to demonstrate how market risk management should be articulated from the top of an organization to its bottom. We focus on banks since this topic is particularly critical in banking, but the concepts apply equally to other financial institutions and non-bank corporations.

The Post-Crisis, Risk-Regulatory Framework

With the benefit of hindsight, capital standards were too weak for the type of risks that were building up in the world's financial system in the run-up to the financial crisis of 2007 to 2009. In particular, the scale and nature of market risk and market liquidity risks were not anticipated or properly managed in financial institutions and both financial institutions and their regulators largely ignored the build-up of systemic risk concentrations. Deficiencies leading to the crisis were numerous and are attributable to bankers, investors, rating agencies, and regulators, among others. We summarize the key deficiencies as background for a discussion of prudential regulation, described in the Appendix.

Capital: The level and quality of capital held by banks proved to be inadequate. In particular, tier 1 hybrid capital did not play its intended loss-absorbing role.[3]

Leverage: There was too much leverage in the banking system, particularly when combined with weak credit underwriting. Losses were worsened by procyclical deleveraging.

Conflict of interest: This was a particular problem regarding the *issuer pays* model of ratings for securitized products. This led to inflated ratings from rating agencies. The rated, securitized products carried yields that were higher than usual for bonds in that rating category, indicating the bond market understood that these bonds were riskier than the ratings suggested.

Capital rules governing the trading book: Banks had built massive illiquid credit exposures in their portfolios. The VaR-based capital regime, with its 10-day liquidity horizon, was not designed to measure this kind of risk. Banks abused this regime and warehoused highly illiquid,

[3] Hybrid capital is junior debt, usually subordinated long-term, issued by commercial banks. In terms of seniority, the payment of principal and interest is behind other liabilities and before equity capital.

structured credit assets on the trading book while holding far too little capital against these assets. The assets proved impossible to value when liquidity disappeared.

Poor funding liquidity risk management and insufficient liquidity buffers: Many banks relied excessively on wholesale, short-term funding to finance long-term, illiquid assets and securitized products. Banks particularly used Asset-backed Commercial Paper (ABCP) conduits to increase their reported returns on equity. They did this by moving loans, mortgages, and securitized products off balance sheets into a conduit or Special Investment Vehicle (SIV), for which only a capital requirement for the backup liquidity line was required (as opposed to the higher capital required if the assets had been held in the bank portfolios). In reality, the banks retained exposure to the off-balance-sheet vehicle. During the second half of 2007, when it became difficult roll short-term funding, banks felt obliged to bring the assets held in conduits and SIVs back onto their balance sheets.

Poor incentives and governance: There were many shortcomings in compensation practices, risk management, and the quality of supervision from boards and management, along with a lack of transparency that made it nearly impossible to understand a bank's exposures or the quality of the capital backing these exposures.

Insufficient understanding of systemic risk: Systemic risk arises when firms or markets have the potential to propagate shocks or credit events, and damage the financial system and the broader economy. During the crisis, systemically important and interconnected institutions that were too big to fail incurred major losses.

Setting Stage for Market Risk Governance

From a risk-governance perspective, a primary responsibility of the board is to look after the interests of shareholders, and possibly other stakeholders. For example, concerning market risk, does it make sense for a corporation to assume a particular market risk given the projected returns of the business activity and the potential threat to the corporation if the risk is realized? Boards also need to be sensitive to the concerns of other stakeholders such as debt holders. Debt holders are mostly concerned about the downside of market risk—how likely is it that a market risk will damage a corporation so badly that it will become insolvent?

A board needs to be on alert for any conflicts that might arise between the interests of management in boosting returns while assuming market risks, and the interests of the company's longer-term stakeholders. This type of conflict of interest is called *agency risk* in academic literature.

Conflicts of interest can occur easily if, for example, executives are rewarded with options that they can cash in if the share price of the company rises above a certain level. Such an arrangement gives management an incentive to push the share price up, but not necessarily sustainably. For example, managers might encourage business lines to earn short-term rewards

in exchange for assuming long-term market risks. By the time "the chickens come home to roost," the managers, including CEOs, might have picked up their bonuses or even changed jobs.

The tension between the interests of the CEO and the interests of longer-term stakeholders explains why boards of directors must maintain their independence from executive teams and why there is a global push to separate the role of the CEO and the chairman of the board. The October 2011 bankruptcy of MF Global, a brokerage firm and one of the ten largest U.S. bankruptcies, offers an example of poor governance. Many commentators have pointed out the danger of the board of a company falling under the spell of a charismatic CEO.[4]

This explains why it is becoming difficult to draw a line between market risk governance and market risk management, and we can see some clear effects of this at the organizational level. For example, consider the changing role of chief risk officer (CRO). A key duty of the CRO is often to act as a senior member of the management committee and attend board meetings regularly. The board and the management committee increasingly look to the CRO to integrate corporate governance responsibilities with the risk function's existing market risk responsibilities. Following the financial crisis of 2007 to 2009, many CROs were given a direct reporting line to the board or its risk committees, in addition to reporting to the executive team and CEO.[5]

True Market Risk Governance

The primary responsibility of the board is to ensure that it develops a clear understanding of the bank's business strategy and the fundamental risks and rewards that this implies. The board also ensures that market risks are transparent to managers and stakeholders through adequate internal and external disclosure.

Although the board is not there to manage the business, it is responsible for overseeing management, and holding it accountable. It must also contribute to development of the overall strategic plan for the firm, considering how changes might affect business opportunities and the strategy of the firm. This necessarily includes the extent and types of risks that are

[4] Jon Corzine, CEO of MF Global, took huge bets on European sovereign debt, eventually leading to an increase in required capital, increased margin calls as positions soured, a ratings downgrade, and loss of confidence in the firm. MF Global was left without the cash to support its operations, and faced a classic run on the bank. Bankruptcy followed.

[5] The Basel Committee stated that a bank CRO should "report and have direct access to the board and its risk committee without impediment. Interaction between the CRO and the board should occur regularly. Non-executive board members should have the right to meet regularly—in the absence of senior management—with the CRO." The Basel Committee's revised Principles on Corporate Governance at Banks, Oct 2014 (http://www.bis.org/publ/bcbs294.htm)

acceptable for the firm (i.e., the board must characterize an appropriate market risk appetite for the firm, generally, and market risk appetite, particularly).

A firm's market risk appetite should connect clearly to its overall business strategy and capital plan. Some business activities might simply be wrong for a firm, given the market risks they entail and the size of the activity in relation to the firm's balance sheet. There is a growing sense that business planning, which tends to be driven by earnings goals in a competitive environment, needs to involve risk management from the beginning to test how targets fit with the firm's market risk appetite, and to assess potential downsides. Equally important is clear communication throughout the firm regarding risk appetite and risk position.

To fulfill its risk governance responsibilities, a board must ensure that the bank has put in place an effective risk management program that is consistent with these fundamental strategic and risk appetite choices, and it must ensure that there are effective procedures in place for identifying, assessing, and managing all risks. For every business disaster in which a firm knowingly took on too much market risk, there is another in which it failed to identify a risk such as an underlying liquidity risk, or ignored the risk because it was thought so unlikely that it did not deserve active risk management.

A board might be challenged by the complexity of market risk management, but the principles at a strategic level are simple. There are only four choices in risk management:

- Avoid market risk by choosing not to undertake some activities;
- Transfer market risk to third parties;
- Mitigate market risk by hedging risky activities;
- Accept market risk, recognizing that, if managed correctly, undertaking some risky activities should generate shareholder value.

In particular, a board should ensure that business and risk management strategies are directed at economic performance while considering accounting performance. For example, trading strategies can be designed to achieve economic performance targets subject to constraints being placed on achieving minimum levels of accounting performance such as ROA > 0. This includes ensuring all the appropriate policies, methodologies, and infrastructures are in place in the Institution.[6] Infrastructure includes both operating elements (e.g., sophisticated software, hardware, data, and operational processes) and personnel.

This might sound like an onerous task, but the board can pull various levers. For example, one way to gauge how seriously a company takes its market risk management is to examine the human capital employed:

- What kind of a career path does the market risk management function offer?;

[6] The OECD's paper on *Corporate Governance and the Financial Crisis: Conclusions and Emerging Good Practices to Enhance Implementation of the Principles*, February 2010 p. 4 states that "an important conclusion is that the board's responsibility for defining strategy and risk appetite needs to be extended to establishing and overseeing enterprise-wide risk management systems."

- To whom do market risk managers report?;
- What salaries are paid to market risk managers in comparison to reward-oriented personnel such as traders?;
- Is there a strong risk culture?

An effective board also establishes strong ethical standards and works to ensure that it understands the degree to which managers follows them. Some banks have set up ethics committees within their business divisions to ensure soft risks such as unethical business practices (e.g., unethical sales practices) do not slip through the mesh of their hard risk reporting framework.

Another important lever available to the board is the firm's performance metrics and compensation strategy. The board has a critical responsibility to ensure that the way staff members are rewarded and compensated is based on risk-adjusted performance, and aligns with shareholders' interests. The increase in misreporting after the millennial stock market boom paralleled the rise of equity-based compensation for CEOs, which arguably provided a perverse incentive to executives to manipulate financial results to boost the share price in the short-term.

A related responsibility is to ensure that any major transactions the bank enters are consistent with the risk authorized and associated strategies of the bank. The board should ensure that the information it obtains about market risk management is accurate and reliable. Directors should demonstrate healthy skepticism (e.g., the quality of the bank stress-testing program) and require information from a cross-section of knowledgeable and reliable sources such as the CEO, market risk officer senior managers, and internal and external auditors. Directors should be prepared to ask tough questions, and they should make themselves able to understand the answers.

However, the duty of the board is not to undertake risk management daily, but to ensure that all the mechanisms used to delegate and drive risk management decisions are functioning properly. Discussed above, the 2007 to 2009 financial crisis highlighted the need to strengthen the role of the board, and therefore[7], board members need to be educated on market risk issues and be given the means to explore and determine the risk appetite of the organization. They should be able to assess the risk of loss that the firm is willing to accept over a specified period, considering its business mix and strategy, earnings goals, and competitive position. This involves understanding the firm's current market risk profile and its business culture regarding the firm's market risk appetite and monitoring the firm's ongoing performance against its market risk appetite.

[7] In October 2010, the Basel Committee issued principles for enhancing corporate governance that addressed such issues as the role of the board of directors, the qualification of board members, and the importance of an independent risk management function (Basel Committee, *Principles for Enhancing Corporate Governance*, October 2010). In the United States, the Dodd-Frank Act requires a dedicated risk committee of the board of directors for publicly traded, bank holding companies with total assets of $10 billion or more, and for systemically important, publicly traded, non-bank financial companies.

Board members of the risk committee need some technical sophistication concerning the risk disciplines, and solid business experience so they can build clear perspectives on risk issues. The risk committee of the board should remain separate from the audit committee since disparate skills are required for each fiduciary responsibility.

Committees: Market Risk Appetite & Market Risk Limits

We set out some of the goals of best-practice risk governance. Now we examine the mechanisms that financial institutions use to translate these goals into reality. We focus on market risk governance in the banking industry, but many of the same principles and structures apply to other types of risks, and other industries.

At most banks, the board charges its primary committees (e.g., audit and risk management committees) with ratifying policies and associated procedures of the bank's risk management. The committees also ensure that implementation of these policies is effective**.**

The committees translate the overall risk appetite of the bank, approved by the board, into a set of limits that flow down through the bank's executive officers and business divisions. All banks, for example, should have in place a risk management committee to monitor market risk reporting, and a system of market risk limits.

The name for each committee varies significantly across the industry, as do the duties of each committee. For our purposes, we imagine an archetypal bank with a senior risk committee that oversees risk management practices and detailed reporting. Junior risk committees that manage specific types of risk such as a market risk committee would often report to this senior risk committee. We examine two mechanisms of risk governance before examining how risk committees use risk metrics and limit frameworks to delegate risk authority down through the bank.

The Special Role of the Audit Committee of The Board

The role of the audit committee of the board is critical to the board's oversight of the bank. The audit committee is responsible for not only reviewing the accuracy of the bank's financial and regulatory reporting, but also ensuring the bank complies with minimum or best-practice standards in other activities such as regulatory, legal, compliance, and risk management. Audit committee members are now required to be financially literate so they can carry out their duties.

We can think of auditing as providing independent verification for the board regarding whether the bank is doing what it says it is doing. Although some of the audit committee's functions sound close to market risk management, this verification function separates the audit committee's work from the work of other risk committees.

The audit committee's duties involve not only checking for infringements, but also overseeing the quality of the processes that underpin financial reporting, regulatory compliance, internal controls, and risk management. In a later section, we examine how the audit function, which often has a direct reporting relationship with the audit committee, acts as an independent check on the bank's risk management.

An audit committee needs members with the right mix of knowledge, judgment, independence, integrity, inquisitiveness, and commitment to function well. In most banks, a nonexecutive director leads the audit committee, and most members are nonexecutives. The audit committee also needs to establish an appropriate interaction with management, which is independent but productive, with all the necessary lines of communication open.

The audit committee needs to ask itself several questions regarding each of its principal duties. For example, with respect to financial statements, the audit committee needs to be satisfied not only that the financial statements are correct, but also that the company adequately addresses the risk that the financial statements may be materially misstated, intentionally or unintentionally.

The audit committee also needs to be clear about the reporting and risk management elements of governance it oversees on behalf of the board. These might include financial reporting, operational effectiveness, and efficiency, and compliance with laws and regulations. Again, the recent financial crisis revealed the weaknesses of the audit committees in many banks and financial institutions (e.g., audit committees did not uncover excess risk assumed by traders or risk of building up large portfolios of structured credit products).

The Special Role of the Risk Management Committee of the Board

At a bank, the risk management committee of the board is responsible for independently reviewing the identification, measurement, monitoring, and controlling of credit, market, and liquidity risks, including the adequacy of policy guidelines and systems. If the committee identifies issues concerning operational risk, it typically refers these to the audit committee for review.

The board of directors also typically delegates to the risk management committee the responsibility of approving market risk limits. These aspects are usually set out in a formal document (e.g., the investment and lending delegation of authority resolution) approved by the board.

The risk management committee reports to the board on a variety of items such as market risk exposure over a specified dollar limit. The risk management committee also monitors major trends in market risk levels, portfolio composition, and industry breakdowns. The committee also typically provides opportunities for separate, direct, and private communication with the chief inspector (i.e., head of internal audit), external auditors, and the management committee.

The Special Role of the Compensation Committee of the Board

One lesson from the 2007 to 2009 financial crisis was that compensation schemes in financial institutions encouraged disproportionate risk-taking, with insufficient regard to long-term market risks. Over the previous two decades, bankers and traders had increasingly been rewarded with bonuses tied to short-term profits or business volume, incentivizing them to front-load fees and income, and backload the risks. Compensation schemes were structured like a call option since compensation increased with the upside, but there were no real penalties in the case of losses. With the help of excessive leverage, this sometimes led bank personnel to bet the entire bank on astonishingly reckless investment strategies.

In many countries, public companies are now required by securities authorities to set up a special board compensation committee to determine the compensation of top executives. This was driven by concerns over market risk governance, particularly the ability of CEOs to convince board members to compensate the CEO and other officers at the expense of shareholders, who had nearly no say in such decisions.

It is now widely recognized that incentive compensation should align with the long-term interests of shareholders and other stakeholders, and with risk-adjusted return on capital. To the extent that this is not the case, it is important for banks to address potential distortions. Incorporating risk management into performance goals and compensation decisions has become a leading practice and compensation planning is viewed as a tool for enterprise-wide risk management.

It will always be tempting for firms to offer attractive compensation packages to revenue-generating talent. International cooperation might be necessary to prevent financial firms from arbitraging the market for human capital through their choice of jurisdiction. In September 2009, the G-20 endorsed the notion that excessive compensation in the financial sector encouraged excessive risk-taking and contributed to the financial crisis. Among the G-20 recommendations were the removal of guaranteed bonuses, the idea that executives should be exposed to downside risk through compensation deferral, and that bonus schemes should include clawbacks in the event that a strategy incurs losses in the longer-term.[8] Moreover, E.U. regulators adopted a rule, which took effect in 2014, that caps bankers' bonuses to the equivalent of their salary, or twice their salary if shareholders explicitly agree with a two-thirds majority. In 2013, the European Parliament voted to cap bonuses in the asset-management industry; bonuses should not exceed base salaries for managers of mutual funds regulated by the European Union.

Stock-based compensation aligns the interests of executives with those of shareholders, but it is not a panacea. Before Lehman's bankruptcy, the employees owned about a third of the firm,

[8] The Financial Stability Board's implementation standards list propositions and periods for deferral such as 40% to 60% lockup of compensation for three years. The Board also recommends that firms prohibit employees from hedging to undermine the intended risk incentive alignment. The Board also suggests that at least 50% of pay be based on shares, along with a share retention policy, as opposed to the use of guaranteed bonuses.

and many employees lost a large chunk of their life savings. Stock ownership can also encourage risk-taking since shareholders' gains are not limited on the upside, but their losses are capped on the downside.

One solution is to make employees creditors of the company by including restricted notes or bonds as part of their compensation package. For example, such a solution has been adopted by UBS, which will pay part of the bonuses of its most highly compensated employees with bonus bonds (i.e., bonds that will be forfeited if the bank's regulatory capital ratio falls below 7.5%). UBS's use of contingent debt is structured to complement this compensation strategy. The contingent debt converts into equity if the capital ratio falls below 5%, a trigger set deliberately lower than the trigger for forfeiture of deferred compensation. The reason is that bond investors are expected to pay more for contingent debt if they expect management to recapitalize the distressed firm before it crosses the threshold for conversion of debt to equity. Compensation policies such as these should improve social welfare more generally by reducing both the likelihood and expected costs of future taxpayer bailouts. [9]

Roles and Responsibilities in Practice

We described the basic structures and mechanisms for risk governance at the board level, but how do these structures and mechanisms cooperate to ensure daily activities of the bank conform to the board-agreed, general market risk appetite and limits set by the board and management committees?

The senior risk committee of the bank recommends to the risk committee of the board an amount of market risk that it is prudent for the risk committee of the board to approve. In particular, the senior risk committee of the bank determines the amount of market risk to be assumed by the bank as a whole, in line with the bank's business strategies. At the top of the hierarchy, the risk committee of the board approves the bank's market risk appetite each year, based on a well-defined and broad set of risk measures such as the amount of overall interest rate risk. The risk committee of the board delegates authority to the bank's senior risk committee, which is chaired by the CEO of the firm, and of which membership includes, among others, the chief risk officer (CRO), head of compliance, heads of the business units, CFO, and treasurer.

---

[9] Compensation schemes similar to this have been advocated by *The Squam Lake Report* (French et al., 2010), which recommends, "Systemically important financial institutions should be required to hold back a substantial share—perhaps 20%—of the compensation of employees who can have a meaningful impact on the survival of the firm. This holdback should be forfeited if the firm's capital ratio falls below a specified threshold. The deferral period—perhaps 5 years—should be long enough to allow much of the uncertainty about managers' activities to be resolved before the bonds mature. Except for forfeiture, the payoff on the bonds should not depend on the firm's performance, nor should managers be permitted to hedge the risk of forfeiture. The threshold for forfeiture should be crossed well before a firm violates its regulatory capital requirements and well before its contingent convertible securities convert to equity."

The senior risk committee is also responsible for establishing, documenting, and enforcing all policies that involve market risk, and delegating specific business-level market risk limits to the CRO of the bank. The CRO is typically a member of the management committee and is responsible for, among other things, designing the bank's market risk management strategy.

Specifically, the CRO, working with the head of market risk, is responsible for the market risk policies, market risk methodologies, risk infrastructure, and market risk governance. The senior risk committee also approves the trading risk mandates. The process for developing and renewing trading mandates should be explicit. For example, trading unit mandates should expire one year after approval. A balance must be struck between ensuring that a business has the limits set high enough to allow it to meet its business goals and maintenance of overall risk standards, including ensuring that limits can be monitored.

Infrastructure and corporate governance groups must be consulted when preparing a trading unit's mandate. The format for obtaining approval of a trading unit mandate should be standardized. For example:

- The trading unit seeking approval should provide an overview and restate the key decisions that need to be taken;
- The trading unit manager should bring everyone up-to-date on the business (e.g., key achievements, risk profile, etc.);
- All trading products should be approved, along with a description of new products (or activities) that might affect the risk profile;
- The trading unit manager should outline future initiatives;
- The proposed risk appetite should be put forward. A report should note the historical degree of use of current limits, and current and proposed limits. It should analyze the impact of full use of limits on liquidity and capital;
- The mandate should describe the operational risks to which the trading unit is exposed, including the impact of finance, legal, compliance, and tax issues.

The senior risk committee of the bank delegates the authority to make daily decisions on its behalf to the CRO, including the authority to approve market risks that exceed the limits assigned to the bank's various businesses, as long as these limits do not breach the overall market risk limits approved by the board.

At many banks, the CRO plays a pivotal role in informing the board and the senior risk committee of the bank about the appetite for market risk across the bank. The CRO also communicates the views of the board and senior management down through the organization. Each business unit, for example, might be given a mandate to assume market risk on behalf of the bank up to a market risk limit. The senior risk committee of the bank must also satisfy itself that the bank's infrastructure can support the bank's market risk management objectives. The senior risk committee of the bank provides a detailed review and approval (say, annually) of

each business unit mandate in terms of the respective risk limits, and delegates the monitoring of these limits to the CRO.

In large banks, developing and renewing this authority is explicit. For example, business unit risk authority typically expires one year after the senior risk committee of the bank approves it. The CRO might approve an extension of a market risk authority beyond one year to accommodate the senior risk committee's schedule.

A balance needs to be struck between ensuring that a business can meet its business goals and maintenance of overall market risk standards (including ensuring that limits can be monitored). Key infrastructure and market risk governance groups are normally consulted when preparing a business unit's mandate. The CRO is responsible for independently monitoring market risk limits throughout the year. The CRO might order business units to reduce their positions or close them out because of concerns about market risks.

The CRO also delegates some responsibilities to the heads of the various business units. For example, at an investment bank, the head of global trading is likely responsible for the risk management and performance of all trading activities, and he or she in turn delegates the management of limits to the business managers. The business managers are responsible for the risk management and performance of the business, and they in turn delegate limits to the bank's traders. This delegation is summarized in Figure 1, with reference to market risk authorities.

Figure 1: Market risk delegation

Below the board committee level, executives and business managers are necessarily dependent on each other when they manage and report on risk in a bank. Business managers also ensure timely, accurate, and complete deal capture, and approve the official profit and loss (P&L) statement.

The bank's operations function is particularly critical to market risk oversight. In the case of an investment bank, for example, the function independently books trades, settles trades, and reconciles front- and back-office positions, which should provide the core record of all the bank's dealings. Product Control staff members prepare the P&L report and independent valuations (e.g., mark-to-market of the bank's positions), and support the operational needs of the various businesses.

The financial crisis highlighted the need to re-empower risk officers in financial institutions, particularly at a senior level. The lessons are that CROs:

- Should not be just after-the-fact risk managers, but also risk strategists; they should play a role in determining the risks that the bank assumes, and help manage those risks. To ensure there is a strategic focus on risk management at a high level, the CRO in a bank or other financial institution should report to the CEO and have a seat on the risk management committee of the board;
- The CRO should engage directly and regularly with the risk committee of the board. The CRO should also report regularly to the full board to review risk issues and exposures. A strong, independent voice means the CRO has a mandate to bring to the attention of both line and senior management, or the board, situations that could materially violate risk-appetite guidelines;
- The CRO should be independent of line business management and have a strong enough voice to impact decisions meaningfully;
- The CRO must evaluate all new financial products to verify that the expected return is consistent with the risks undertaken, and the risks are consistent with the business strategy of the institution.

Market Risk Limits and Limit Policies

A corporation must be able to tie its board-approved market risk appetite and market risk tolerances to business strategies, to achieve best-practice market risk governance. This means, in turn, that an appropriate set of limits and authorities must be developed for each portfolio of business, each type of market risk across the relevant risk factors (within each portfolio of business), and the entire portfolio.

Market risk limits the control risk that arises from changes in absolute price (or rate) of an asset. The bank will also want to set tight policies regarding exposure to both asset/liability management risk and market liquidity risk, especially in the case of illiquid products. The nature of each market risk limit varies widely, depending on the bank's activities, size, and sophistication. It is best practice for institutions to set down the process by which they establish

market risk limits, review market risk exposures, approve limit exceptions, and develop the analytic methodologies used to calculate the bank's market risk exposures.

At many banks, best-practice market risk governance calls for development and implementation of sophisticated market risk metrics such as value-at-risk (VaR) measures for market risk or potential exposure limits by risk grade for credit risk.

Risk-sensitive measures such as VaR are useful for expressing risk under normal market conditions and for most kinds of portfolios but are worse under extreme circumstances or for specialized portfolios (e.g., some kinds of option portfolios). So, limits should also relate to scenario and stress-testing measures to ensure the bank can survive worst-case scenarios (e.g., extreme volatility in the markets).

Most institutions employ two types of limits; we call them limit types A and B. Type A (also called tier 1) limits might include a single overall limit for each asset class (e.g., a single limit for interest-rate products), and a single, overall stress-test limit and cumulative loss from peak limit. Type B (also called tier 2) limits are more general and cover authorized business and concentration limits (e.g., by credit class, industry, maturity, region, etc.).

The setting of the market risk limit level in terms of a particular metric should be consistent with underlying standards for market risk limits (proposed by the risk management function and approved by the senior risk committee).

It is unrealistic, in practice, to set market risk limits so that they are likely to be used fully in the normal course of events without leading to limit transgressions. Instead, limit setting needs to consider an assessment of the business unit's historical use of limits. For example, type A limits for market risk might be set at a level such that the business, in the normal course of its activities and in normal markets, has exposures of about 40% to 60% of its limit. Peak use of limits, in normal markets, should generate exposures of perhaps 85% of the limit. If a business is consistently generating exposures at a sufficiently high usage of limits (say above 85% of the limit) in the normal course of its activities then the bank may call for a re-appraisal of the business.

A consistent limit structure helps a bank consolidate its approach to risk across many businesses and activities. Additionally, if the limits are expressed in a common language of risk such as economic capital, type B limits can be made fungible across business lines. Nevertheless, such transfers require joint approval of the head of a business and the CRO.

Risk Management Systems

Many risk management systems are developed to perform unique functions, but in some cases, the functions overlap. Typically, firms face the problem of fragmentation in their existing risk management systems. The systems cannot easily communicate with each other, sometimes

known as the *islands of automation* problem. This causes redundancy, expensive processing, and increased costs since each system must be supported separately.

An effective risk management system needs to generate the necessary risk management information on all risks, perform specific analytical functions, and permit multitasking. The risk management system should allow for easy integration of new applications and platforms, but balance this flexibility with the need for management control. An **application architecture** establishes the technical, functional, and operational characteristics of application systems (i.e., their construction and use).

The risk management system needs to be supported by an **information technology (IT) architecture** that is employed in all of the company's information processing. The IT architecture is essentially a set of standards and guidelines (input from business principles) that should be adhered to by staff members when they make technological decisions. The design of the IT infrastructure should optimize the exchange of information between each entity within the firm, and all should be operating within a unified IT framework.

Risk Management Data

Ensuring the integrity of risk data provides an important competitive advantage since data is translated into risk management information for both transaction-makers and policy-makers. A **data architecture** deals with the establishment of an environment in which all information can be accessed and understood by any associate of the firm. The **organization architecture** deals with the responsibilities and interrelationships necessary to ensure a comprehensive information interchange between parties.

One lesson learned from the 2007 to 2009 global financial crises was that banks' IT architectures and data architectures were inadequate. Basel published a series of principles for effective data aggregation and risk reporting in January 2013.[10] Effective implementation of these principles enhances risk management at banks. The principles cover four areas: i) overarching governance and infrastructure, ii) risk data aggregation capabilities, iii) risk reporting practices, and iv) supervisory review, tools, and cooperation.

IT design needs to consider the means by which key risk management information is gathered from the various internal and external systems into a Risk Data Warehouse. One task is to organize the necessary risk management data (transmitted to the risk management system from multiple legacy systems) into a common format (i.e., data dictionary). It needs to consider how risk management information might change over time. The information might be static

[10] Basel Committee, *Principles for Effective Data Aggregation and Risk Reporting* (January 2013) http://www.bis.org/publ/bcbs239.pdf

(e.g., contractual details of a transaction such as the coupon of a corporate bond) or dynamic (e.g., market information such as daily closing prices).

Monitoring Market Risk

Once a bank sets its market risk limits in a way that is meaningful to its business lines, how should it monitor those limits to ensure they are followed? Market risk is perhaps the most time-sensitive of limits.

All market-risk positions should be valued daily. Units independent of traders should prepare daily profit and loss statements and provide them to (non-trading) senior management. All assumptions used in the models to price transactions and value positions should be verified independently.

There should be timely and meaningful reports to measure the compliance of the trading team with risk policy and risk limits. There should be a timely escalation procedure for any limit exceptions or transgressions (i.e., it should be clear what a manager must do if his/her subordinates breach the limits).

Variance between the volatility of the value of a portfolio and of that predicted using the bank's risk methodology should be evaluated. Stress simulations should be executed to determine the impact of major market or credit risk changes to P&L.

The bank must distinguish data used for monitoring type A limits (where data must be independent of risk takers) and data used to supply other kinds of management information. For other types of analyses, where timeliness is a requirement, risk managers might be forced to use front-office systems as the most appropriate sources. For example, real-time market risk measurement such as that used to monitor intraday trading exposures might simply have to be derived from front-office systems, but data used in limit monitoring must be:

- Independent of the front office;
- Reconciled with the official books of the bank to ensure their integrity;
- Derived from consolidated data feeds;
- In a data format that allows risk to be measured (e.g., it might employ the market-risk VaR or credit-risk VaR methodology).

Business units should be under strict orders to advise the risk management function that they might exceed a limit well before the limit excess occurs. For example, there might be an alert when an exposure is at, say, 85% of type A or type B limit. The CRO, jointly with the head of the business line, might then petition the senior risk committee of the bank for a temporary increase in limits. If risk management is advised of a planned excess, it should be more likely that the excess will be approved. This gives the business unit a necessary incentive to provide early warnings.

What happens if the limit is breached? The risk management function, illustrated in Figure 2, should immediately place any excess on a daily "limit type A or limit type B exception report," with an appropriate explanation and a plan of action to cope with the excess. The head of risk management might authorize the use of a reserve.

Figure 2: Limit excess escalation procedure

Limit type A excesses must be cleared or corrected immediately. Limit type B excesses should be cleared or approved within a short timeframe, say a week. The market risk manager should then report all limit excesses across the bank on an exception report, which might be discussed at a daily risk meeting, and should distinguish limit type A and type B excesses. No manager should have the power to exclude excesses from the daily excess report.

When market risk limits become effective, they impose a hidden cost; the bank cannot assume additional market risk and thus might have to give up profitable opportunities. As a limit is approached, the opportunity cost of the limit should be evaluated so the bank can decide whether the limit should be relaxed.

What is the Role of the Audit Function?

We set out, generally, market risk management that should support best-practice market risk governance, but how does a board know that executives and business managers are living up to the board's stated intentions (and to minimum legal and regulatory requirements)? The answer lies in the bank's audit function and the periodic investigations it carries out across the bank. One role of the audit function is to provide an independent assessment of the design and implementation of the bank's risk management. For example, regulatory guidelines typically require internal audit groups to review overall risk management annually. This means addressing the adequacy of documentation, effectiveness of the process, integrity of the risk management system, organization of the risk control unit, integration of risk measures into daily risk management, etc.

Regulatory guidelines typically require auditors to address the approval process for validation derivatives valuation models and systems used by front- and back-office personnel, validation of any significant changes to market risk measurement, and scope of market risks captured by the market risk measurement model. Regulators also require that internal auditors examine the integrity of the management information system and the independence, accuracy, and completeness of position data.

Beyond local regulatory requirements, an audit objective should be to evaluate the design and conceptual soundness of market risk measures, including methodologies associated with stress testing. Internal auditors should verify the accuracy of models through examination of back-testing.

Audits should also evaluate the soundness of elements of the risk management information system (i.e., the risk MIS) such as coding and implementation of internal models. This should include examination of controls over market position data capture, and controls over parameter estimation (e.g., volatility and correlation assumptions).

Audit responsibilities often include providing assurance concerning the design and conceptual soundness of the financial rates database used to generate parameters entered in the market VaR and credit VaR analytic engines. Audits also review the adequacy and effectiveness of monitoring risk, progress of plans to upgrade risk management systems, adequacy and effectiveness of application controls within the risk MIS, and reliability of validation.

Audits should also examine documentation related to compliance with qualitative/quantitative criteria outlined by regulatory guidelines. They should comment on the reliability of VaR reporting frameworks. Box 2 sets out, generally, what a statement of audit's findings on the market risk management function might look like. It also clarifies dangers that might arise from confusion between the role of risk management and audit.

BOX 2: EXAMPLE STATEMENT OF AUDIT FINDINGS

If all is well from a market risk management perspective, the audit should state that adequate processes exist for providing reliable risk control and ensuring compliance with regulatory criteria. For example, in short form, the audit group's conclusion regarding market risk control in a bank trading business might be that:

1. The market risk control unit is independent of the business units;
2. The internal market risk models are used by business managers;
3. The bank's market risk measurement model captures all material risks.

Furthermore, if all is well, the audit group should state that adequate and effective processes exist for:

1. Risk-pricing models and valuation systems used by front- and back-office personnel;
2. Documenting the market risk management systems and processes;
3. Validating significant changes to market risk measurement;
4. Ensuring the integrity of the market risk management information system;
5. Positioning data capture (and that any positions not captured do not materially affect risk reporting);
6. Verifying the consistency, timeliness, and reliability of data sources used to run internal models, and that data sources are independent;
7. Ensuring the accuracy and appropriateness of volatility and correlation assumptions;
8. Ensuring the accuracy of the valuation and risk transformation calculations;
9. Verifying the model's accuracy through frequent back-testing.

Internal auditors have devised international standards to provide objective assurances regarding control, governance, and risk management. The Institute of Internal Auditors (IIA) provides guidance that has been organized into an International Professional Practices Framework (IPPF), offering both mandatory and strongly recommended guidance. The IPPF has performance standards that encompass a variety of activities.[11]

Model Risk Governance

For simple instruments such as stocks and straight bonds, model risk governance is less important. However, model risk governance becomes a compelling issue for institutions that trade over-the-counter (OTC), exotic derivatives products and execute complex arbitrage strategies.

The danger of dependence on models has been clear from early in the history of the derivatives markets. However, the importance of model risk governance became apparent during and after

[11] See the *Professional Guidance* section of the IIA's website. These standards include Managing the Internal Audit Activity, Nature of Work, Engagement Planning, Performing the Engagement, Communicating Results, Monitoring Progress, and Resolution of Senior Management's Acceptance of Risk. The Governance and Risk Management Standards are a subset of the Nature of Work standard, and the Risk Management standards cover topics such as evaluating an organization's risk exposure, evaluating fraud risks, reviewing risk during consulting, and risk knowledge gained during consultancy.

the 2007 to 2009 financial crisis when severe losses on an entirely unexpected scale were incurred on trading positions. Consequently, the Basel Committee required financial institutions to assess the model risk associated with their trading activities (i.e., the risk of losses due to using the wrong or a misspecified model for pricing and hedging securities).[12]

Part of the model risk governance challenge resides in the complexity of models. There has been a relentless increase in the complexity of valuation theories used to support financial innovations such as caps, floors, swaptions, spread options, credit derivatives, and other more exotic derivative instruments, and a parallel rise in the threat from model risk. Since 2004, there has been a constant stream of new financial products based on market volatility, most notably options based on the VIX, volatility index, which we discuss below. However, product innovation raced ahead of the ability to price the new instruments accurately or hedge the associated risks.

Technology also played a role. Computers are now so powerful that there is temptation to develop ever more complex models that are increasingly less understood by managers. The technology available has substantially increased the chances of creating losses (and profits). The result is that not a single market crisis passes without incurring several large trading losses that can be ameliorated with strong model risk governance. In 1997, the Bank of England conducted a survey highlighting the variation in models that existed among 40 major derivative trading firms based in London. Vanilla foreign exchange instruments had low variation in both value and sensitivities, but some exotic derivatives had large variations not only in value, but also in sensitivity measures: 10% to 20% for swaptions, and up to 60% for exotic foreign exchange instruments. Another study in the same year showed that the models available to calculate VaR sometimes gave very different answers when applied to the same portfolio.[13]

It is, therefore, unsurprising that trading firms of all kinds can experience substantial losses in stormy market environments, and sometimes even when things are calm. Since models are used for valuation, inferior model risk governance can make a strategy appear very profitable on paper, even though the bank is incurring economic losses or unwise risk exposures, perhaps

---

[12] According to the Basel Committee, "banks must explicitly assess the need for valuation adjustments to reflect two forms of model risk: the model risk associated with using a possibly incorrect valuation methodology, and the risk associated with using unobservable (and possibly incorrect) calibration parameters in the valuation model". Basel Committee on Banking Supervision, *Revisions to the Basel II Market Risk Framework, Bank for International Settlements*, February 2011.

[13] Researchers presented an identical asset portfolio to a number of commercial vendors of software for VaR calculations. Each was asked to use the same volatility inputs, obtained from JP Morgan's Risk Metrics, and report the aggregate VaR for the entire portfolio and the VaR figure for each type of instrument (such as swaps, caps and floors, and swaptions). Variation among vendors was striking given they were analyzing the same position (in relatively simple instruments), using the same methodology and market parameter values. For the whole portfolio, the estimates ranged from $3.8 million to $6.1 million, and for the portion containing options, the VaR estimates varied from $747,000 to $2,100,000. C. Marshall and M. Siegel, "Value-at-Risk: Implementing a Risk Measurement Standard," *Journal of Derivatives* 4(3), 1997, pp. 91-111.

over several years. By the time the fault is corrected, a large hole might have appeared underneath the bank's accounts.

Mitigating Model Risk

One important way to mitigate model risk is to invest in research to improve models and develop better statistical tools, either internally at the bank or externally at a university (or at an analytically oriented consulting organization). An even more vital way of reducing model risk is to establish a process for the independent validation of how models are both selected and constructed. This should be complemented by independent oversight of the P&L calculation.

The role of validation is to offer assurance to the firm's managers that any model for the valuation of a given security proposed by, say, a trading desk is reasonable. In other words, it provides assurance that the model offers a reasonable representation of how the market values the instrument, and the model has been implemented correctly. Validation should consist of:

1. *Documentation.* The validation team should ask for full documentation of the model, including both the assumptions underlying the model and its mathematical expression;
2. *Soundness of model.* An independent model validator must verify that the mathematical model is a reasonable representation of the instrument being valued;
3. *Independent access to financial rates.* The model validator should check that the bank's middle office has independent access to an independent market-risk management financial rates database (to facilitate independent parameter estimation);
4. *Benchmark modeling*. The model validator should develop a benchmark model based on assumptions, and on the specifications of the deal;
5. *Health-check and stress test the model.* The model should be checked to confirm it possesses the basic properties that all derivatives models should possess such as put-call parity and other no-arbitrage conditions. The model validator should also stress test the model;
6. *Build a formal treatment of model risk into the overall risk management procedures, and periodically re-evaluate models*.

Valuation in a Marked-to-Market World During Low Liquidity

Financial instruments are held in the:

- Trading book, where they are measured at fair value through profit and loss, or
- Banking book as assets available for sale (AFS), where they are subject to amortized cost accounting (also called accrual accounting).

Any change to the fair value of a trading book instrument has a direct impact on a firm's income statement in the period in which said change occurs. Changes in the fair value of financial assets classified as AFS are recorded directly in equity, without affecting profit and loss until the

financial assets are sold, at which point the cumulative change in fair value is charged or credited to the income statement.

In contrast, unless held for sale, loans are typically measured at amortized cost using the effective interest method, with less allowance or provision for impairment losses. Loans held for sale might be reported in trading or AFS portfolios, or in the United States, in held-for-sale portfolios at the lower of cost or fair value.

Instruments subject to fair value accounting are valued with reference to prices obtained from active markets when these are available for identical or similar instruments. When market liquidity dries up (e.g., during a market crisis), price discovery based on market prices becomes much more difficult. Other valuation techniques such as applying a model to estimate a value might become necessary. When liquid market prices are unavailable, other approaches inevitably carry with them a range of uncertainties, and can give a false impression of precision.

The accounting standard for fair value (FAS 157) creates a hierarchy of inputs into fair value measurements, from most to least reliable:

- Level 1 inputs are unadjusted quoted market prices in active liquid markets for identical products;
- Level 2 inputs are other directly or indirectly observable market dates. There are two broad subclasses of these inputs. The first and generally preferable subclass is the quoted market prices in active markets for similar instruments. The second subclass is other observable market inputs such as yield curves, exchange rates, empirical correlations, etc. These inputs yield mark-to-model measurements that are disciplined by market information, but can only be as reliable as the models and inputs employed;
- Level 3 inputs are unobservable, firm-supplied estimates such as forecasts of home price depreciation and the resulting severity of credit losses on mortgage-related positions;
- Fair value/mark-to-market accounting has generally proven to be highly valuable in promoting transparency and market discipline and is an effective and reliable accounting method for securities in liquid markets. However, it can create serious, self-reinforcing challenges that make valuation more difficult and increase uncertainties around those valuations, when there is no or severely limited liquidity in secondary markets. Three criticisms of fair value accounting have been expressed:[14]

[14] Looking at the pros and cons of fair value accounting, it still appears better than the alternative of accrual accounting. Accrual accounting suppresses the reporting of losses and reduces incentives for voluntary disclosure. This means it can discourage actions necessary to resolve a crisis. The Savings and Loans crisis in the United States provides the best illustration. The crisis began when interest rates rose during the first oil crisis/recession from 1973 to 1975, causing thrifts' fixed mortgage assets to experience large economic losses that were unrecognized under amortized cost accounting. This non-recognition of economic losses allowed bank regulators and policy-makers to permit the crisis to continue for 15 years, effectively encouraging thrifts to invest in risky assets, exploit deposit insurance, and in some cases commit fraud, activities that worsened the ultimate cost of the crisis.

- unrealized losses recognized under fair value accounting might reverse over time. Market prices might deviate from fundamental values because of market illiquidity or prices are bubble prices;
- market illiquidity might render fair values difficult to measure, yielding overstated and unreliable reported losses;
- firms reporting unrealized losses under fair value accounting might trigger unhelpful feedback effects (i.e., trigger further deterioration of market prices through a destabilizing downward spiral of forced liquidations, write-downs, and higher risk and liquidity premiums).

Conclusion: Steps to Success

In complex, risk-taking organizations, it is impossible to separate best-practice market risk management from best-practice corporate governance. Boards cannot monitor and control the financial condition of a risk-taking institution without excellent market risk management and market risk metrics. Meanwhile, the risk management function depends on sponsors at the senior executive and board levels to gain the investment it requires, and the influence it needs to balance powerful business leaders.

It is worth stressing an important lesson from business history. Many fatal market risks in a corporation are associated with business strategies that at first look like runaway successes, and only later do the overlooked or discounted risks become apparent.

At a best-practice institution, everything flows from a clear and agreed-on risk management policy at the top. For example, senior managers and the board must approve a clear notion of the institution's market risk appetite, and set out how this is to be linked to an enforceable system of market risk limits and market risk metrics.

Without this kind of platform, it is difficult for risk managers further down the management chain to make decisions on how they approach and measure market risk. For example, without a clearly communicated concept of an institution's risk appetite, how would risk managers define a worst-case market risk during any extreme risk scenario analysis? How would they decide whether the institution could live with the small chance of a worst-case outcome, or alternatively, avoid risk to solvency by severely limiting business volumes or even closing a business line down (in the face of attractive profits)?

The risk committees of the institution also need to be involved, to some degree, in setting the basic risk measurement methodologies employed by the institution. Most banks know they must define their market risk in terms of practical market, but banks have also extended their risk measurement framework to include more sophisticated approaches to market liquidity risk and a new class of firm-wide stress tests. It is important that risk committees understand the strengths and weaknesses of new metrics to make sense of risk reports.

There are also unavoidable strategic, political, and investment reasons why the board and top executive managers must be involved closely in determining an institution's risk management strategy. Without their involvement, how can the managers of the institution agree on a credible organizational infrastructure that avoids both gaps and duplications in risk oversight? The key to designing an efficient organization is to ensure that the roles and responsibilities of each risk mechanism and unit are carefully spelled out and remain complementary. Meanwhile, data for market risk analysis, including enterprise-wide macroeconomic stress testing, must be drawn from many business lines and bank functions. An enterprise-wide perspective is increasingly essential.

We should not think of board and top management time spent on market risk management as time spent on the defensive risk-control aspects of the business. A best-practice market risk system can be applied to gain offensive advantages. A board with a sound understanding of the market risk profile of its key existing or anticipated business lines can support aggressive strategic decisions with much more confidence. Risk measures such as VaR, stress testing, and economic capital offer a way of setting risk limits, but they are also vital in helping the institution decide which business lines are profitable, once risk is considered.

Ideally, businesses use the risk infrastructure as a tactical management tool during deal analysis and pricing, and consider its results in incentive compensation schemes to ensure market risk management and business decisions align. A joint approach to governance and risk management has become a critical component of a globally integrated, best-practice institution, from board level to business line.

Appendix

Prudential Market Risk Regulation

Background: The Explosion of Bank Market Risk

When devising the 1988 Accord, regulators focused primarily on the credit risks to which banks were exposed, and ignored market risk and other risks, but this hardly reflected the reality of many bank's risk exposures, even during the 1980s. Contemporary banks engage in a range of activities that extend beyond lending and the credit risk it generates. They trade and finance all types of securities, and derivatives such as swaps, forward contracts, and options, either for their own account to act as a market maker or to facilitate customer transactions.

This kind of bank trading activity grew exponentially in the 1980s and 1990s, so that by the time the Basel Committee published its important 1996 Market Risk Amendment, the Federal Reserve Bank estimated that U.S. banks possessed over $37 trillion of off-balance-sheet assets and liabilities, in comparison to approximately $1 trillion 10 years earlier. According to a more recent BIS publication, as of November 2011, banks worldwide had total exposure to derivatives of approximately $708 trillion.

The rise in importance of risk management instruments over the last few decades was driven by a rise in volatility in many of the principal financial markets, which led banks to become both users and providers of risk management instruments. The prime example of this change is the foreign currency market. From 1944, with the signing of the Bretton Woods Agreement, international foreign exchange rates were fixed artificially. Central banks intervened in their foreign currency markets whenever necessary to maintain stability. Exchange rates were changed infrequently, with the permission of the World Bank and the International Monetary Fund (IMF). These bodies usually required a country that devalued its currency to adopt tough economic measures to ensure the future stability of the currency.

The regime of fixed exchange rates broke down during the late 1960s due to global economic forces. These included a vast expansion of international trading and inflationary pressures in the major economies. The shift to flexible foreign exchange rates introduced daily (and intra-day) volatility to exchange rates. As the volatility surfaced in traded foreign currencies, the financial market began to offer special tools to insure against these new risks.

The first contracts were various kinds of futures and forwards, though foreign currency options soon followed. In 1972, the Mercantile Exchange in Chicago (CME) created the International Monetary Market (IMM), which specialized in foreign currency futures and options on futures on the major currencies. In 1982, the Chicago Board Option Exchange (CBOE) and the Philadelphia Stock Exchange introduced options on spot exchange rates. Banks joined the trend by offering OTC forward contracts and options on exchange rates.

Development of interest rate volatility and derivatives instruments followed a similar story from the early 1970s. The equity and commodity markets also came to support significant derivatives markets, often developed by banking institutions. As the result of bank activity in these new derivative markets, banks naturally became increasingly exposed to volatile derivative instruments, meaning, these exposures had to be risk managed carefully.

Group of Thirty (G-30) Policy Recommendations

The 1996 Amendment to the Basel Accord had a notable precursor. In 1993, the Group of Thirty (G-30) published a report that described best-practice, price risk management recommendations for dealers and end-users of derivatives (and for legislators, regulators, and supervisors). The report was based in part on a detailed survey of industry practices among dealers and end-users worldwide.

G-30 focused on providing practical guidance in terms of managing derivatives businesses, offering an important benchmark against which participants could measure their own price risk management practices. Its recommendations covered sound market risk policies (e.g., establishment of a market risk function independent of trading decisions), credit risk policies, enforceability policies, infrastructure policies, accounting and disclosure policies, etc. They continue as cornerstones of modern bank risk management frameworks.

The 1996 Market Risk Amendment (The 1996 Amendment)

Recommendations in the G-30 report established qualitative standards for bank risk management of derivatives risk, but the explosion in bank trading of derivatives and more mundane securities clearly had implications for how regulators calculated the amount of regulatory capital a bank required for market risk. The 1996 Amendment to the 1988 Accord, implemented in 1998, extended the initial Accord to include risk-based capital requirements for market risks banks incur in their trading accounts. The fundamental innovation of the 1996 Amendment was to allow sophisticated banks to use their own internal VaR model to calculate regulatory capital for market risk in their trading book.

Market risk is not the only risk arising from instruments such as derivatives; they also give rise to credit risk. Under the 1996 Amendment, off-balance-sheet derivatives such as swaps and options are subject to both the market risk requirement and the credit risk capital requirements stipulated in the original 1988 Accord. In contrast, on-balance-sheet assets (e.g., bonds) in the trading portfolio are subject to the market risk capital requirement only, a feature that offsets the aggregate effect of the new rules on the amount of capital that banks must set aside.

Banks that adopted the internal models approach realized substantial capital savings, in the order of 20% to 50% depending on the size of their trading operations and type of instruments traded, because internal models can be designed to capture diversification effects by modeling correlations between positions. In addition to market risk capital adequacy requirements, the Basel Committee set limits on concentration risks. Under the Amendment, risks that exceed 10% of the bank's capital must be reported, and banks are forbidden to take positions that are greater than 25% of the bank's capital. As a historical footnote, had these rules been in effect in 1994, Barings Bank would have been prohibited from building up such huge exchange-traded futures positions, and the world's most famous example of rogue trading might have been avoided. At the time the bank collapsed in February 1995, Barings' exposures on the SIMEX and OSE were 40% and 73% of its capital, respectively.

1996 Amendment Qualitative Requirements

Before an institution became eligible to use its own internal VaR model to assess regulatory capital related to market risk, the regulators insisted it must put sound risk management practices in place, largely in accord with the G-30 recommendations we described earlier. The institution needed to demonstrate that it had a strong risk management group, independent of the business units that the group monitored and that it reporting directly to the senior executive managers of the institution.

Implementing a VaR model is an initiative that require effort and co-operation. An important part of setting up any VAR model for regulatory purposes is ensuring that the risk factor model inputs are reliable and accurate. A formal validation system is needed to approve the models, modifications to them, their assumptions, and their calibration.

Enhancements to the Basel II Framework

During the 2007 to 2009 crisis, a number of banks experienced large losses in their trading books, the risk of which had not been captured in the banks' VaR models. This pointed to a number of deficiencies in the VaR-based capital methodology, which is typically based on a 99%, one-day VaR scaled up to 10 days. The following additional capital requirements were therefore imposed:[15]

- *Stressed VaR.* Banks using internal models in the trading book must calculate a stressed VaR based on a 12-month period of significant financial stress. The calculation should be portfolio specific. This additional capital requirement recognizes that traditional VaR calculations capture the risk of normal markets, and are not calibrated to periods of stress;
- *Incremental Capital Charge* (IRC). Many of the losses during the credit crisis were not caused by defaults, but rather the loss of liquidity and declines in value due to credit migration and widening of credit spreads. The IRC represents an estimate of the default and migration risks of unsecuritized credit products over a one-year horizon at the 99.9% confidence level, considering the liquidity horizons of individual positions or sets of positions.[16]

The IRC encompasses all positions subject to a capital requirement for specific interest rate risk according to the internal models approach, except securitization positions that have a different treatment, as discussed below.

The IRC model should also capture the impact of rebalancing positions at the end of their liquidity horizons to achieve constant risk over a one-year horizon. Existing exposures are rebalanced at the end of the liquidity horizon or rolled over when they mature to maintain the initial risk level as indicated by a risk metric such as VaR or the profile of exposures by credit rating and concentration. The IRC requirement includes the impact of clustering of default and migration events during stressed markets. The impact of diversification between default or migration events and other market factors is not considered. Therefore, the IRC capital requirement is simply added to the VaR-based capital requirement for market risk.

For securitized products in the trading book, the capital requirements of the banking book apply with the exception of correlation trading portfolios. Resecuritizations (e.g., collateralized debt obligations of asset-backed securities such as CDOs of ABSs) also receive a specific rating-based requirement, reflecting their prominent role during the credit crisis.

---

[15] Banks were originally expected to comply with the revised requirements by December 31, 2010. However, by 2012, only Australian, European, and several Asian banks had implemented Basel 2.5.

[16] The liquidity horizon represents the time required to sell the position or hedge all material risks covered by the IRC model in a stressed market. The liquidity horizon has a minimum term of three months.

Correlation trading books[17] are exempt from the full treatment for securitization positions and qualify for a revised standardized requirement or a capital requirement based on a comprehensive risk measure (CRM) that captures not only incremental default and migration risks, but also all price risks, including basis risk. Capital requirements for these portfolios remain subject to a floor of 8% of the standardized requirement. Banks using the internal models approach for market risk should have in place a rigorous and comprehensive stress testing program. Banks' stress scenarios should cover a range of factors that can create extraordinary losses or gains in trading portfolios. These factors include low-probability events in all major types of risks (e.g., market, credit, operational risks, and liquidity risk).

Scenarios should include past periods of significant disturbances such as the 1987 equity market crash, the exchange-rate mechanism crises of 1992 and 1993, the fall in bond markets in the first quarter of 1994, the 1998 Russian financial crisis and subsequent LTCM failure, the bursting of the technology stock bubble at the turn of the millennium, and the 2007 to 2009 financial crisis. The scenarios should include both large price movements and a sharp reduction in liquidity associated with these events. The bank should also develop a second type of scenario to evaluate the sensitivity of the bank's market risk exposure to shocks in risk factors such as volatilities and correlations. Regulators say banks must develop bank-specific scenarios, selecting the most challenging scenarios given the unique characteristics of each bank's portfolios. In combination, the result of the Basel 2.5 reforms is that each bank must meet a capital requirement daily, expressed as:

Capital = max {VaR, k*(average VaR over 60 days)} + max {Stress VaR, k*(average Stress VaR over 60 days)} +IRC

where:

- $k \geq 3$ is a regulatory parameter linked to the number of back-testing outliers reported by the Institution in the last year;
- VaR is measured at the 99% confidence level over a 10-day period, and combines both general market risk and specific risk;
- Stress VaR is computed using data from a stressful period such as 2007 to 2009;
- IRC is a credit VaR, calculated over a one-year period at the 99.9% confidence level that should capture both default risk and migration risk, and should be calibrated to the bank's own through-the-cycle, historical loss experience. All positions that generate a potential credit risk should be included in the IRC. All sovereign bonds are subject to the

[17] Correlation trading portfolios might include simple securitization exposures and n-to-default credit derivatives that meet the following criteria: i) the position is not a resecuritization, or an option on a securitization tranche, or a synthetically leveraged super-senior tranche, and ii) all reference entities are single-name products, including single-name credit derivatives and CDS Index tranches, and tailored tranches for which a two-way market exists. Even so, these desks are exposed to basis risk (e.g., between the tailored and index tranches). At sophisticated firms, these risks are measured through VaR, which typically includes base correlation and specific VaRs.

IRC, which poses the thorny question of the probability of default of a country (e.g., the United States).[18]

Prudential Market Risk Regulation: Discussion of Basel 2.5

Trading book capital was driven predominantly by one risk measure, VaR. Under Basel 2.5, it is driven by VaR, stressed VaR, IRC, CRM, and the standardized requirements for securitization, plus a standardized floor for the CRM. The problem with this additional complexity is that it has many internal inconsistencies. Basel 2.5 is a patchwork of over-conservative, overlapping rules that when combined, generate a punitive level of capital for the trading book. For some trades, the amount of capital might exceed the face value of the position (i.e., be more than the bank can lose).[19] According to the Basel Committee's own calculations, as a result of the Basel 2.5 revisions, market risk capital requirements increased by an estimated average of three to four times for large, internationally active banks.

Fundamental Review of the Trading Book

Basel 2.5 was an emergency response to the undercapitalization of banks' trading books revealed during the subprime crisis. However, it suffers from a number of recognized shortcomings. Regulators are preparing for a more fundamental review that is likely to address the following areas:

- *Lack of coherence:* the current framework is characterized by a layer of overlapping capital requirements that can lead, as we discussed above, to a capital requirement that is higher than the maximum loss;
- *Boundary between the trading book and the banking book:* large differences in the capital requirements for similar types of risk in the trading book and the banking book (e.g., treatment of interest rate risk) might lead to regulatory arbitrage;
- *Market liquidity risk:* the industry needs to develop a comprehensive framework that captures the risk of market illiquidity during stressed periods;

---

[18] Even if some sovereign bonds are subject to a 0% risk weight under the Standardized Approach, they attract a capital charge under the IRC.

[19] Assume for illustrative purposes that i) volatility under stressed market conditions is three times the volatility of a normal market environment, and ii) returns are normally distributed so that StressVaR is three times NormalVaR, neglecting IRC for the purpose of the exercise. Now suppose that the portfolio has an annualized volatility in normal market conditions of 10%. Over 10 days, the standard deviation is 2%. The 10-day standard deviation in stress conditions is thus 6%, according to our (reasonable) assumptions. The sum of these (i.e., 8%) must be multiplied by the 99% standard normal critical value of 2.33, and then by a multiplier of at least 3. Assuming a green-zone model (i.e., a multiplier of three), regulatory capital under the new rules (ignoring the IRC) is 2.33 × 3 × 8% = 56% of the portfolio exposure. Given a well-diversified, partially hedged portfolio, with an annualized volatility of 5% and old regulatory capital of 7% of exposure, the new charge is 28%. However, with a partially diversified, lightly hedged portfolio, with normal volatility of 15% and stress volatility of 60%, the new rules lead to a capital charge of 105% of the size of the portfolio, which, if the positions are long, is higher than the maximum loss that could be incurred on this portfolio. Under these simple assumptions, the new regulatory capital charge is always four times the capital charge without the stressed component.

- *Internal models-based approach and risk measure:* A number of weaknesses have been identified regarding the use of VaR for determining regulatory requirements, including its inability to capture tail risk. The Basel Committee is contemplating adopting an alternative: the expected shortfall (ES) approach, measuring the expected loss beyond a given confidence level. In addition, risk models would be calibrated to a period of significant financial stress;
- *Standardized approach:* The current standardized approach to market risk would be revamped to improve risk sensitivity to reduce the risk sensitivity gap between the standardized and models-based approaches. The revised standardized approach would also become a more credible fallback in case a bank's internal risk model is deemed inadequate;
- Credit Value Adjustments (CVA): The relationship between counterparty credit risk and the trading book regime needs to be clarified.

Chapter 3 – Market Risk Measurement

Max Wong and Changwei Xiong

Value at Risk - Overview

Under the Basel regime for banking regulation, Value-at-risk (VaR) is the *de facto* risk model for computation of regulatory capital requirements. However, VaR has been criticized severely, especially after the 2008 credit crisis, because many assumptions behind the model failed during that stressful period; VaR understated risks during times it was needed most.

Due to Basel's endorsement, use of VaR and similar models is widespread in many areas of banking regulation. In the banking book, we have the IRB (internal ratings based) model for credit risk and the IRRBB (interest rate risk in banking book) model for interest rate risk. In the trading book, we have the VaR, stressed VaR, and IRC (incremental risk charge) models for most trading operations, and the specific risk model and CRM (comprehensive risk model) for correlation trading, and in operational risk, we have the OpVaR model. These actuarial-based or VaR-like models always involve sampling an empirical distribution or specifying an assumed distribution. VaR is a statistical measure of risk based on the loss quantile of such distributions; it measures risk based on a total portfolio, considering diversification.

Generally, banks use VaR models for two purposes. First, VaR is a model for calculation of regulatory minimum capital requirements. Since this minimum capital is required to protect against crises that would threaten survival of the bank, it is prudent to set the confidence level extremely high, typically at 99% or higher. Second, VaR is used for daily risk management, setting of risk limits, and risk attribution analysis, whereby lower confidence is acceptable and desirable since it affords a more precise estimate of VaR. This is not only a risk-scaling exercise since asset classes are impacted disparately by those quantile changes. Generally, the further one goes into the (high-quantile) tail, the riskier, for example, AAA-rated securities appear in comparison to B-rated ones.

Definition of VaR

VaR is (an estimate of) a single number representation of risks for the entire distribution. Specifically, it is a quantile. This simplification makes risk management and reporting easier. We define VaR as an estimate of the loss from a fixed set of trading positions over a fixed time horizon (typically 1 day or 10 days) that would be equal or exceeded with a given probability. Mathematically, the VaR is defined by the probability statement:

$$\mathbb{P}(V_T - V_0 \leq \mathrm{VaR}) = 1 - \alpha \qquad (1)$$

where α is the confidence level and V_t the value of a portfolio under review at time t. We use $V_T - V_0$ to denote the portfolio gain or loss (or P&L, short for profit and loss) over the specified time horizon T. For example, if a portfolio has a 10-day, 99% confidence level VaR of -$1 million (the VaR here is assumed to be negative; the smaller the VaR, the higher the risk; people also often quote VaR as a positive number) today, there is a probability $(1 - \alpha)$ of 1% that the portfolio might lose by $1 million *or more* over the next 10 days. Several details of the VaR definition are worth mentioning:

1. VaR is an estimate, not a value defined uniquely. In theory, the value of any VaR estimate depends on the stochastic process that drives the random realizations of market data. For more sophisticated VaR models, data generation must be identified (or modeled), and its parameters calibrated. This requires resorting to historical experience, which raises many practical issues such as the length of the historical sample used and whether more recent events should be weighted more heavily than those further in the past. In practice, stable, long-running, random processes do not generate market data because there are regime changes. For example, the market state during a crisis is different from that of a post- or pre-crisis period. A model that is unable to capture the dynamic nature of the market will be "too little, too late" at capturing risks. Differing methods for dealing with the uncertainty surrounding changes in regimes are at the core of why VaR estimates are seldom unique.
2. The trading positions are assumed fixed over the forecast risk horizon (say, 10 days for regulatory VaR reporting). This can be unrealistic in an investment banking or trading portfolio setting where trades are bought and sold at a high turnover rate daily. In practice, simple scaling rules are used to scale estimates 1-day VaR to the longer risk horizon. Without this simplification, it is necessary to model what happens within the specified time horizon and make behavioral assumptions related to trading strategies during the period.
3. VaR does not address the distribution of potential losses on occasions when the VaR estimate is exceeded (i.e., it is oblivious to the tail losses beyond VaR). Hence, VaR is not the worst-case loss or expected loss; VaR is the *minimum* loss given the probability.
4. Although the VaR figure is the focus for regulatory reporting and limits monitoring, banks also look at other VaR-related measures to assess how additional/component positions affect risks and diversification. We discuss this risk decomposition next.

Marginal VaR

A trading portfolio might contain many component positions of instruments and products. To analyze VaR and its risk contribution from each component, risk managers often break the portfolio VaR down through *VaR decomposition*. For example, we have a portfolio whose value today is V_p. The portfolio consists of N components in which the value of the i-th component asset is V_i. Risk managers often like to know the contribution of risk from components to the portfolio's total VaR. Marginal VaR is one such measures, defined as the partial derivative of portfolio VaR with respect to the component asset value:

$$\mathrm{MVaR}_i \equiv \frac{\partial \mathrm{VaR}_p}{\partial V_i} \tag{2}$$

Marginal VaR describes the change in total VaR resulting from "one" dollar change of the component value. Before we discuss marginal VaR further, we introduce the basic notion of *modern portfolio theory* (MPT) and see its relationship to VaR. Portfolio return R_p for a given time horizon is the weighted sum of component asset returns R_i for $i = 1, \cdots, N$:

$$R_p = \sum_{i=1}^{N} w_i R_i \tag{3}$$

where $w_i = V_i/V_p$ is the weight of the i-th component asset and R_i the i-th component asset return. The summation formula can also be written succinctly in matrix notation.

$$R_p = w^T R \tag{4}$$

where w is a column vector of the weights and R a column vector of the returns. We use superscript T to denote a matrix transpose operation. MPT assumes that the returns of component assets follow a multivariate Gaussian distribution, which implies that the portfolio returns also follow a Gaussian distribution. The following formula gives the variance of the portfolio return σ_p^2:

$$\sigma_p^2 = w^T \Sigma w = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{5}$$

where Σ is the covariance matrix of returns of component assets. Entry σ_{ij} of matrix Σ is the covariance term between the returns of the i-th and j-th component assets, and it becomes a variance term if $i = j$. Assuming portfolio returns follow a Gaussian distribution, VaR can be derived easily from the return variance σ_p^2:

$$\mathrm{VaR}_p = z\sigma_p V_p \tag{6}$$

where z is the $(1 - \alpha)$ percentile of a standard normal (VaR tends to exclude the Expected Loss component. For simplicity, we assume a zero mean for the Gaussian distribution). For example, if the confidence level $\alpha = 99\%$, then $z = -2.326$. You can compute this using the Microsoft Excel function $\mathrm{NORMSINV}(1 - \alpha)$. Given the relationship in (6), we can derive the marginal VaR from (2) as:

$$\mathrm{MVaR}_i \equiv \frac{\partial \mathrm{VaR}_p}{\partial V_i} = \frac{\partial \left(z\sigma_p V_p \right)}{\partial \left(w_i V_p \right)} \tag{7}$$

$$= \frac{z\partial\sigma_p}{\partial w_i}$$

where we use $V_i = w_i V_p$ in the denominator. The partial derivative in (7) can be derived further. We differentiate (5) with respect to vector w. This gives a first derivative in a row vector form, written as:

$$\frac{\partial\sigma_p}{\partial w} = \frac{1}{\sigma_p} w^T \Sigma \tag{8}$$

Its i-th component is calculated as:

$$\frac{\partial\sigma_p}{\partial w_i} = \frac{1}{\sigma_p} \sum_{j=1}^{N} w_j \sigma_{ji} = \frac{\sigma_{ip}}{\sigma_p} \tag{9}$$

where we denote the term:

$$\sigma_{ip} = \sum_{j=1}^{N} w_j \sigma_{ji} \tag{10}$$

the covariance between the returns of the i-th component asset and the portfolio. Since the covariance can be regarded as a product of two volatilities along with the correlation between them, we can write it as:

$$\sigma_{ip} = \sigma_i \sigma_p \rho_{ip} \tag{11}$$

where σ_i is the return volatility of the i-th component asset and ρ_{ip} the correlation of returns between the i-th component asset and portfolio. After substituting (9) and (11) into (7), the marginal VaR reads:

$$\begin{aligned} \mathrm{MVaR}_i &= z\frac{\sigma_{ip}}{\sigma_p} = z\sigma_i\rho_{ip} \\ &= \frac{\mathrm{VaR}_i}{V_i}\rho_{ip} \end{aligned} \tag{12}$$

We follow the *capital asset pricing model* (CAPM), and define a sensitivity term β_{ip} between the portfolio and its i-th component asset as the ratio between the component covariance and the portfolio variance:

$$\beta_{ip} = \frac{\sigma_{ip}}{\sigma_p^2} \tag{13}$$

The marginal VaR in (12) can then be expressed as:

$$\begin{aligned} \mathrm{MVaR}_i &= z\sigma_p\beta_{ip} \\ &= \frac{\mathrm{VaR}_p}{V_p}\beta_{ip} \end{aligned} \qquad (14)$$

The equations above show that the marginal VaR of a component is the product of a) the percentage VaR of the overall portfolio, and b) the component beta against the full portfolio. The component beta defined in (13) can be negative (zero) in case the component has offsetting risks with (i.e., is uncorrelated with) the portfolio. Generally, the beta of a long and a corresponding short position are of equal size but of opposite sign so those component risks cancel in the overall risk position.

When we use (14) to calculate marginal VaR, we assume the asset returns follow a multivariate Gaussian distribution. Although the assumption of Gaussian asset returns is questionable, the assumption that portfolio returns follow a Gaussian distribution is defensible, at least in quiet markets, due to the central limit theorem; the average of a large number of similar-sized, independent, random variables with finite variance converges to a Gaussian distribution. However, during a crisis, two assumptions can be violated: a) the underlying variable might not be of finite variance, and (b) the codependence structures for the strongly non-Gaussian and/or correlations are so large, the CLT does not apply (yet). If the assumption does not hold, the MVaR calculated by (12) or (14) offers only approximate values.

Incremental VaR

Besides the marginal VaR, there is another risk measure called incremental VaR, which measures the change in VaR due to a new position added to the portfolio. Mathematically, it is defined as:

$$\mathrm{IVaR}_a \equiv \mathrm{VaR}_{p+a} - \mathrm{VaR}_p \qquad (15)$$

where VaR_{p+a} is the VaR of the portfolio after including the position a, and VaR_p the VaR of the portfolio excluding position a. In the notation above, all VaR values are negative. Clearly, when $\mathrm{IVaR}_a > 0$, the position a contributes by increasing overall diversified VaR by the amount IVaR_a (making the VaR number less negative). In other words, if $\mathrm{IVaR}_a > 0$, the added position is risk reducing; it hedges some of the portfolio risk by the amount IVaR_a. Conversely, if $\mathrm{IVaR}_a < 0$, the position is risk increasing.

Marginal and Incremental VaR relate closely; the Marginal VaR is the Incremental VaR for the proverbial *one-dollar increment*. This is expressed as a percentage. If the MVaR of a position is, say, 5%, the IVaR for a small increase is also 5%; dollar amounts scale with the size of the increase. IVaR is always smaller (i.e., more negative) than MVaR. The reason is that the IVaR effectively uses the *average beta* of the position over the range of the increment, and beta is a

strictly increasing function of the component size. This is easy to see; the bigger the component, the more important it becomes as part of the portfolio. In the limit of very large sizes, the beta of any component is plus one, even for a position that began at a negative beta.

Component VAR

A third risk measure is the component VaR (CVaR), which has the nice feature of being additive (unlike the above measures); the component VaRs add to the portfolio VaR . This technique of risk decomposition is thus used often for management information and capital allocation because it is more intuitive, convenient, and easier to understand. For example, a treasurer often uses it to allocate trading budgets and measure risk-adjusted performance of individual desks. However, this method should not be used for risk management and hedging since it ignores the reality of risk diversification in portfolios. Component VaR, in this case, partitions the portfolio VaR into parts that add to the total diversified VaR. By definition, the component VaR can be expressed in terms of the marginal VaR we discussed previously:

$$\mathrm{CVaR}_i \equiv \mathrm{MVaR}_i V_i = \frac{\partial \mathrm{VaR}_p}{\partial V_i} V_i \qquad (16)$$

VaR_p is homogenous in the portfolio notionals (it has to be; the scaling factor lambda can be thought of as a simple transformation in numéraire):

$$VaR_p(\lambda V_i) = \lambda VaR\_p(V_i)$$

If we derive this expression with respect to lambda and take it at the point where lambda is unity, we find equation:

$$\sum_{i=1}^{N} \mathrm{CVaR}_i = \mathrm{VaR}_p \qquad (17)$$

Thus, the component VaRs sum to the total VaR of a portfolio. This expression holds independently of the underlying assumptions on the portfolio distribution. We can define the component risk measure for any reasonable risk measure using equation (16) above, and it will always satisfy equation (20).

Basic Concepts and Definitions

The VaR system used in a bank is nothing more than an aggregation engine to find the portfolio level (or joint) P&L distribution, and then compute the quantile loss at the portfolio level. However, even before the risk aggregation step, many upstream preprocesses need to occur, including market data capture and cleaning, risk-factor mapping, positional capture, full revaluation of positions under various scenarios, etc. Before exploring the various VaR methodologies or systems, we introduce these preprocessing concepts.

Mapping Position to Risk Factors

The first step to building a risk management system is the mapping of risk factors, in which a superset of risk drivers are identified and mapped to a *subset* of risk factors. Why do we need to reduce, and in some cases simplify, the number of risk factors that goes into the VaR model? A portfolio P&L is the sum of P&L of all deals in a portfolio. The P&L of each deal can be derived by observing daily changes in market prices of the deals (i.e., marking-to-market). In theory, one can analyze the risk of a portfolio by examining changes in P&L contributed by each deal. In practice, banks analyze the risk of a portfolio by examining risk factors that drive changes in portfolio P&L. In other words, the P&L used for VaR and risk management is computed, theoretically, as a function of risk factors, instead of marking the portfolio to market. Given a set of risk factors, most trade-booking systems used by banks have pricing libraries that allow computation of the present value (PV) and, by extension, the P&L of each deal. Thus, we project our positions onto a small set of risk factors. This process of describing positions in terms of standard risk factors is known as risk-factor mapping. Modeling risk-by-risk factors offers many advantages:

1. **It allows proxying.** We might not have sufficient historical data for some positions. For example, we might have an emerging market security that has a very short data history, or a tailored OTC (over-the-counter) instrument that has no price observation. In such circumstances, it might be necessary to map the security to a comparable index or proxy asset that does have sufficient data for modeling.
2. **It reduces the dimension of the problem.** A typical bank might have hundreds of thousands of deals mapped to a smaller subset of risk factors. This greatly reduces the necessary computer time to perform risk simulations. In effect, reducing a highly complex portfolio to a consolidated set of risk-equivalent positions in basic risk factors simplifies the problem, allowing faster simulations. The reduction in the dimension also improves the precision of the tail measures such as VaR.
3. **It is more natural for the purpose risks analysis to decompose portfolio risks in terms of its risk factors.** Often, these risk factors are driven uniquely by changes in macroeconomic factors. For example, central bank policy actions and expectations have a direct impact on interest rates risk factors. Other risk factors might be less affected here.

As an example, we consider an FX option. According to the Black-Scholes valuation formula, an FX option is primarily a function of the following risk drivers: FX spot rate, interest rates, and volatility. Even if a portfolio contains a thousand option deals (of the same currency), they can all be represented (or mapped) to just these three types of risk factors. This provides a great deal of efficacy to the business of risk management.

Although it is true that every forex option can be correctly priced using its *implied volatility* (i.e., the volatility that when plugged into the Black Scholes formula, backs out the correct price), this leaves us with as many implied vols as we have options. The big step to take here is to model only a small, but sufficient, number of volatilities across the entire options portfolio.

Scenario Generation

VaR is derived from a distribution of P&L, and is its quantile. In practice, this distribution is comprised of scenarios sampled from history in the so-called historical simulation VaR approach. In contrast, *parametric* approaches exist in which a *theoretical* distribution of P&L is assumed (but this is far less popular in banks; see the next section for the VaR methods). The scenarios are generated from historical time series of risk factors, where each risk factor comes in the form of a daily time series of level data (i.e., prices or rates). Since VaR relies on *returns* scenarios to form the P&L distribution, the series of level data must be transformed into series of returns.

As a typical example, we can choose a 500-business day rolling observation period (or window) representing two calendar years. So, each risk factor has a return series represented by a scenario vector of length 500. Once we derive the return scenarios, we can apply the return series to estimate a parametric distribution for P&L, and compute the VaR analytically. The estimated parameters could be the moments of the distribution, for example. Alternatively, we can use the scenario vector to generate a P&L distribution empirically (non-parametrically), and take its quantile as VaR. For example, we can use a scenario to shift the current levels (or base levels) of risk factors to the shifted levels. Assets in a portfolio are then revalued at the current and shifted levels. The difference between the two valuations is the P&L for that scenario. In summary, the set of scenarios computes and gives a P&L distribution. We often call this the P&L vector. The VaR is then the empirical quantile of the resulting P&L vector.

Three common types of returns can be computed from the level data, the results of which are the scenarios derived from history: a) absolute return, b) relative return, and c) log return. The absolute return takes the difference between two levels:

$$\text{absolute\_return}(i) = \text{level}(i) - \text{level}(i-1) \tag{18}$$

where $i = 1, \cdots, 500$ is the scenario number. The relative return is a percentage change:

$$\text{relative\_return}(i) = \frac{\text{level}(i)}{\text{level}(i-1)} - 1 \tag{19}$$

The log return takes the natural logarithm of the ratio of the two levels:

$$\text{log\_return}(i) = \ln \frac{\text{level}(i)}{\text{level}(i-1)} \tag{20}$$

The relative return asymptotically converges to the log return since the two levels get closer. They are also approximately the same for small perturbations. Relative return or log return is suitable for assets that trade based on price (e.g., exchange rates, price indices, and stock market indices), the big advantage being that in this case, prices never become negative. For

assets that trade on yield (e.g., interest rates and bond yields), absolute return is often a better representation.

For example, the S&P 500 would be modeled using relative or log returns, while bond yields might be modelled using absolute returns (a caveat being that this gives non-zero probability to negative yields, but in practice, this probability is usually very small so this is acceptable. As of 2014, some government bonds are trading on negative yields).

Risk Sensitivities (Greeks)

Risk sensitivity is an important topic for risk management not only because it is used for limits monitoring/control of risk taking, but also because it is used in parametric methods of computing VaR. Risk sensitivities, or Greeks, measure the change in the present value of a position due to a specified change in the risk factor to which the position is exposed. They are often used in risk management and control (e.g., to hedge portfolios), and to set and assess risk limits. They are also often used during VaR calculations since they allow approximation of price changes of a portfolio under the VaR scenarios in a computationally efficient manner. We consider a few basic types of sensitivities, listed in
Table 1.

Table 1. Risk sensitivities

Sensitivity	*Type*	*Definition*	*Application*
Delta (δ)	First Order	P&L due to a small change in price	All derivatives based on assets that trade on price (e.g., equities and FX)
Gamma (γ)	Second Order	Second order P&L correction due to a small change in price	
Vega ($\mathcal{V}$)	First Order	P&L due to change in volatility (typically 1 point change)	All (non-linear) Derivatives
PV01	First Order	P&L due to +1 basis point change in rate	All derivatives based on assets that trade on yield (e.g., swaps and bonds)
Convexity	Second Order	Second-order P&L correction due to +1 basis point change in rate	
CR01	First Order	P&L due to +1 basis point change in credit spread	All derivatives based on credit assets

If we have a scenario, we want to know the P&L generated by each of the positions in the portfolio. We fundamentally have two options: a) we can run a full revaluation of every position based on those new parameters, or b) we approximate the P&L impact by developing the P&L using a Taylor expansion (Greeks relate closely to partial derivatives, mathematically speaking). For example, if we have a 10-year, fixed-coupon bond with semi-annual coupon payments, we might price the bond using the usual cost-of-carry formula:

$$V(y) = p\left(1+\frac{y}{2}\right)^{-10} + \sum_{i=1}^{10} c\left(1+\frac{y}{2}\right)^{-i} \quad (21)$$

where p is the par value paid on maturity, c the fixed-coupon cash flow, and y the bond yield rate. The (symmetrical) first-order risk sensitivity, PV01, of the bond is defined as:

$$\text{PV01} = \frac{V(y+0.5\text{bp}) - V(y-0.5\text{bp})}{1\text{bps}} \quad (22)$$

The perturbations used are not always 1bp wide, but they are generally normalized back to this level. The second-order sensitivity (convexity) 6 of the bond (with a +- 1bp perturbation) can be computed using the following second-order central difference formula:

$$\text{CONVEXITY} = \frac{V(y+1\text{bp}) - 2V(y) + V(y-1\text{bp})}{(1\text{bp})^2} \quad (23)$$

In first order, the bond P&L can be approximated using:

$$\text{P\&}L \approx \text{PV01} \times \Delta y \quad (24)$$

where Δy is the scenario change of yield rate in basis points (bps). However, a linear approach is reliable only when the products' payoff is linear or close to linear. For example, forwards, futures, and swaps have values that are nearly linear depending on the values of the underlying assets (i.e., risk factors). If the positions have considerable optionality or other nonlinear features, such as in the case of options or exotic products, linear approximations can be unreliable. In this case, we try to accommodate nonlinearity by including the second-order (gamma or convexity) term of the Taylor expansion. This is called delta-gamma approximation. For the bond example, we can get better P&L approximation by including the convexity correction term:

$$\text{P\&}L \approx \text{PV01} \times \Delta y + \frac{1}{2} \times \text{CONVEXITY} \times (\Delta y)^2 \quad (25)$$

Using the sensitivity approach in the P&L calculation means we do not perform full-revaluations of a deal using its pricing formula repeatedly, which might incur a heavy computation load. Instead, we approximate the P&L by multiplying the deal's risk factor sensitivity (which needs to be computed only once) with the corresponding risk factor's scenario return. In fact, we compute the Greeks only once at the portfolio level, and from that point onward, we can ignore the number of positions in the portfolio.

Using Greeks is a big computational advantage over the full revaluation approach, and can be used without loss of accuracy for moves that are small enough for the first- or second-order

approximation to be sufficiently precise. It is also possible (albeit complicated, especially if non-trivial cross sensitivities are involved) to use higher-order terms. A problem is when payoffs are digital (e.g., barrier options), where close to the boundary, a Taylor expansion fails.

Distributional Assumption and Volatility Estimation

When computing a tail risk measure such as VaR, it is necessary to make assumptions about the distribution of portfolio returns. A widespread assumption is that the (log) returns of the assets for any given period form a joint Gaussian distribution, and they are independent and identically distributed for non-overlapping periods. A one-dimensional Gaussian distribution can be described uniquely by only two parameters: its mean and variance (for a Gaussian vector, the mean is a vector and the covariance is a symmetrical matrix).

However, looking at real financial time series, we often find that their distributions are heavy tailed and skewed; they are not Gaussian. The true probability of a very large return, especially on the downside, for assets where this notion makes sense, at the tails is greater than the one estimated under a Gaussian distribution with same mean and variance. This finding challenges measurement and use of VaR at high confidence levels. If VaR is calculated under the assumption of a Guassian distribution and yet the markets are heavily tailed, VaR understates the true risk during crises. Since VaR is used for computation of required regulatory capital, available capital of banks might be insufficient to withstand losses when disaster strikes. The naïve solution is to increase the multiplier relating capital requirement to the VaR measure, which, especially for market risk VaR, is an arbitrary number. Without a detailed view of how *taily* a distribution is, it is difficult to impossible to understand whether the multiplier is sized appropriately.

One area in which these assumptions fall short is that in real markets, big moves tend to follow big moves, meaning that at the very least, there is codependence of the variances (volatilities) at two adjacent points in time; high volatility follows high volatility, and vice versa. The quickest way to identify volatility clustering is to plot the return series and check for clusters visually (Figure 1). More formally, one can test for autocorrelation of squared returns. During stress, financial data often exhibit positive autocorrelation in squared returns. Since an increase in volatility heightens the probability of large returns, it makes the empirical distribution of the return appear more heavily tailed. Under this view, fat tails are not an intrinsic feature of the distribution, but are a result of changing (or stochastic) volatility, but at any given point in time, the instantaneous distribution is still normal and independent (other than through its volatility parameters) from the instantaneous distribution at other points in time.

The volatility of risk factors determines the P&L distribution, and therefore has substantial influence on the VaR. To improve the forecasting power of the VaR, we must use historical observations that represent current market variation best. The simplest measure for the volatility is the standard deviation s (we do not discuss whether to normalize with N or N-1; in practice, this is largely irrelevant). Its square gives the variance defined by:

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(r_i - \bar{r})^2 \qquad (26)$$

which assigns equal weights to historical observations. In practice, this measure loses its forecasting power if the return distribution changes (i.e., is not constant) over the observation period.

Volatility clustering indicates that the asset returns are not independent. Although it is impossible to predict the direction of the returns based on historical returns, it is possible to predict their volatility. If we want to capture the volatility-clustering phenomenon during VaR calculations, we can estimate the *conditional* volatility, that is, volatility conditional on the recent past. We discuss two widely used methods for this: a) exponentially weighted moving average (EWMA), and b) generalized autoregressive conditional heteroskedasticity (GARCH) models. Heteroskedasticity refers to non-constant volatility in a return series. The GARCH model is more sophisticated and difficult to implement, but offers potential advantages.

The EWMA model was proposed by JP Morgan's Riskmetrics© in 1994 [1]. It quickly became a popular benchmark after Basel adopted VaR as the *de facto* model for risk capital under its internal-models approach. EWMA estimates the volatility by assigning heavier weights to recent observations than those from the distant past. The EWMA volatility forecast σ_i for day i is given by the recursive equation:

$$\sigma_i^2 = \lambda\sigma_{i-1}^2 + (1-\lambda)r_{i-1}^2 \qquad (27)$$

where λ is the decay factor that determines how rapid the weight decays as an observation goes into the past. Since λ is positive, today's variance autocorrelates positively with yesterday's variance, so EWMA captures the idea of volatility clustering. Parameter λ can also be seen as a persistence parameter; the higher the value of λ, the more persistently high (low) variance will lead to high (low) variance. Riskmetrics proposed $\lambda = 0.94$ in its daily volatility calculation for the stock market. This value gives a volatility forecast closest to the realized ones in history.

One way of estimating the value of λ is through maximum likelihood estimation (MLE). This method assumes a parametric distribution (e.g., normal or Student *t* distribution) for the return series. The idea is to find an optimal λ such that the computed σ series maximizes the probability of the realization of the observed return series. For example, if we assume normal distributions for the returns, (i.e., $r_i \sim N(0, \sigma_i^2)$), the time-dependent conditional variances σ_i^2 are given recursively by EWMA (27). The occurrence of the return series would have a probability that is proportional to the product of the probability density functions (PDF) of the series, which is called likelihood function $\mathcal{L}$. We can thus calibrate the parameter λ to match the observations in history. The likelihood function is:

$$\mathcal{L} = \prod_{i=1}^{N} \varphi(r_i; \sigma_i^2) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{r_i^2}{2\sigma_i^2}\right) \tag{28}$$

where $\varphi(r_i; \sigma_i^2)$ is the PDF of a normal distribution with mean of zero and variance σ_i^2. We assume a zero mean for returns for simplicity. The MLE finds the optimal λ such that the resulting σ_i series maximizes $\mathcal{L}$. This is equivalent to maximizing the natural logarithm of the likelihood function $\ln \mathcal{L}$ because log transformation is monotonous:

$$\ln \mathcal{L} = \sum_{i=1}^{N} \left(-\ln \sigma_i - \frac{1}{2} \ln 2\pi - \frac{r_i^2}{2\sigma_i^2}\right) \tag{29}$$

GARCH models are similar to EWMA in that both address volatility clustering. They are natural extensions of the autoregressive conditional heteroskedasticity (ARCH) models proposed by Engle (1982) [2] by including an autoregressive moving average (ARMA) model for error variance. This generalization makes the GARCH model very flexible, and the numerous free parameters allow the model to calibrate to various behaviors and characteristics of a market. All GARCH models share a common feature: yesterday's risk correlates positively with today's risk (i.e., an autoregressive structure exists in risk). The GARCH model, or more accurately GARCH(p,q) model, has a general form:

$$\sigma_i^2 = \omega + \sum_{k=1}^{p} \alpha_k \sigma_{i-k}^2 + \sum_{k=1}^{q} \beta_k r_{i-k}^2 \tag{30}$$

where parameters $\omega > 0$ and $\alpha_k, \beta_k > 0$ for $k > 0$ ensure strong positivity of the conditional variance, and we also require:

$$\sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{q} \beta_k < 1 \tag{31}$$

to ensure stationarity of the conditional process. Otherwise, the model is intractable (i.e., unsolvable). Lag lengths p and q define the order of the dependence of current volatility on the past information. Hence, the recursive definition in the model allows a non-constant volatility conditional on the volatilities and return realizations in the past. Again, we assume a zero mean for returns for simplicity. When we set $p, q = 1$, it gives the simplest GARCH model known as GARCH(1,1), popular in financial applications:

$$\sigma_i^2 = \omega + \alpha \sigma_{i-1}^2 + \beta r_{i-1}^2 \tag{32}$$

This discussion focuses on this simple model. To estimate parameters ω, α and β in GARCH(1,1) model, we can again use maximum likelihood estimation. For example, in most GARCH models,

returns are assumed to follow a normal distribution specified by mean of zero and the conditional variance series σ_i^2, that is, $r_i \sim N(0, \sigma_i^2)$ for $i = 1, \cdots, 500$. By using MLE, the optimal estimation of model parameters is such that the resulting σ_i^2 series maximizes the likelihood function $\mathcal{L}$, as shown in (28) (or $\ln \mathcal{L}$ in practice). The EWMA model in (27) is actually a special case of the GARCH(1,1) model in (30). GARCH(1,1) extends EWMA by adding constant term ω and relaxing the constraint that the coefficients $(\alpha + \beta)$ must sum to one. If the sum $(\alpha + \beta)$ is less than one (the more usual case), the volatility is mean-reverting, and the rate of mean reversion relates inversely to this sum. Unlike the EWMA model, the conditional variance in GARCH(1,1), in the absence of market shock, drifts toward its long-term variance, defined by:

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta} \tag{33}$$

In the spreadsheet attached [VaR_Volatility_Models.xls], we demonstrate calibrations of the EWMA and GARCH(1,1) model to the three years series (from 3 Jan 2011 to 31 Dec 2013) of daily log-returns of S&P500 equity index. The return series shown in Figure 1 reveals clustered volatilities across time. The models are calibrated to the return series by MLE, where Microsoft Excel Solver performs the optimization. The calibrated models are as follows (the numbers are the calibrated parameters):

$$\begin{aligned} &\text{EWMA model:} \\ &\qquad \sigma_i^2 = 0.9222 \times \sigma_{i-1}^2 + 0.0778 \times r_{i-1}^2 \\ &\text{GARCH}(1,1)\text{model}: \\ &\qquad \sigma_i^2 = 0.000004334 + 0.8197 \times \sigma_{i-1}^2 + 0.1357 \times r_{i-1}^2 \end{aligned} \tag{34}$$

The long-term (or steady-state) volatility in the GARCH(1,1) model is 0.986%, as per equation (33), which is close to the sample standard deviation (of returns) of 1.048%. Using the estimated model parameters, we can make one-step forward predictions of the volatility, also shown in Figure 1. Both models capture the volatility-clustering feature in the empirical data. The trends of the predicted volatilities are similar between models. Since the GARCH(1,1) model involves more free parameters, it shows more variations and responsiveness than the EWMA model.

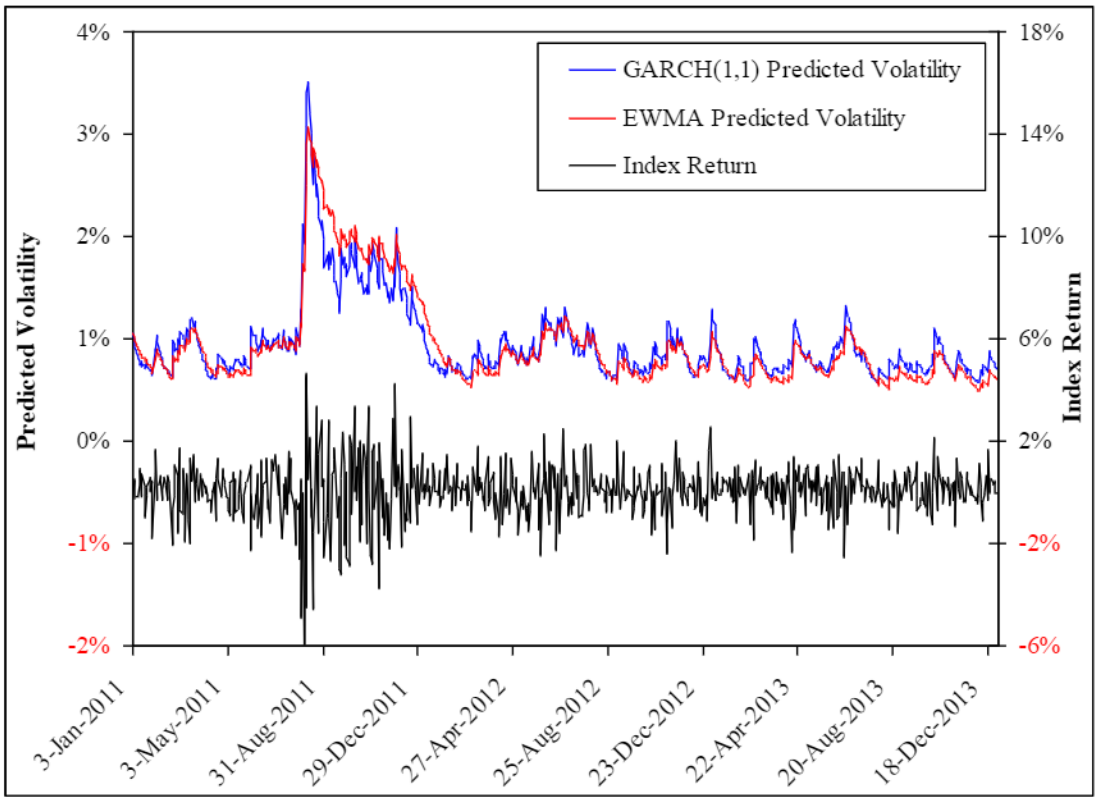

Figure 1. Volatility prediction: EWMA model versus GARCH(1,1) model (calibrated to S&P500 equity index daily returns from 2011 to 2013)

At this stage, there are no unique values for volatility, and hence also for VaR. They are statistical estimates, dependent on the choice of models. Instead of estimating this number from history, some analysts prefer to examine volatility implied from the options markets. This could be (arguably) forward-looking since it gives an instantaneous poll of a market's expectations through its price discovery.

Historical volatility and implied volatility

The models we discuss in previous sections estimate volatility from realized past returns of a risk factor. We call it historical volatility because it is a backward-looking measure, determined statistically from history. It differs conceptually from another volatility measure: the implied volatility. Implied volatility is used when calculating the price of an option. It is the volatility of the underlying asset that when input into the option pricing model (e.g., Black-Scholes model), returns a theoretical value equal to the current market price of the option. In the Black-Scholes model, the theoretical price of an option is a function of five parameters (for simplicity, dividend rate is ignored): spot price of underlying asset, strike price, expiry (i.e., time to maturity), volatility, and cost-of-carry (e.g., interest rates, storage cost, dividend rate, etc.). Given that the first four parameters are known for an option, the market price of the option gives a one-to-one translation to the volatility parameter. The value of the volatility backed out from traded price is the implied volatility. In fact, the market convention is to quote the implied volatility for trading. It is a forward-looking and subjective volatility measure that reflects the

market's expectation of the future dynamics of the underlying asset. Since options on an asset can be traded with different strike prices and expiries, implied volatilities derived from the market prices form a volatility surface defined by the grid of strike price versus expiry. For a given expiry, the implied volatilities at different strike prices usually show a 'smile'. This refers to a convex pattern of the implied volatility when plotted against the strike price. According to the Black-Scholes pricing model, the assumption of a normally distributed asset returns gives rise to a constant volatility across strikes. Hence, the phenomenon of volatility smile shows that the assumption is violated in practice. It can be shown mathematically that a distribution, which is fatter than normal (i.e., leptokurtic) and skewed, gives rise to a volatility smile that is slanted to one side. This occurs in nearly all options markets. The volatility smile also contains other information such as the liquidity premium across different strikes. The premium typically increases for strikes further away from the current spot price.

In risk management, we are interested in the fluctuation of risk factors or scenarios. This is captured objectively using historical volatility and other quantile measures such as VaR. In contrast, implied volatility is a risk factor by itself when option products are in the portfolio. Since most banks trade in option products, risk-factor mapping involves capturing the entire volatility surface. For example, if the FX options markets have 10 currencies with 5 strikes and 10 maturity expiries, the bank needs to map 500 risk factors for implied volatility. The VaR then involves measuring the fluctuation (or volatility) of these volatility risk factors to the extent that the bank is exposed to them.

VaR Specification

Most banks use their own in-house VaR systems to calculate VaR for risk management and reporting, and banks use different specifications for their VaR systems. They calculate a firm-wide VaR number, report it to the regulator, and periodically disclose it to investors. To make the comparison of VaR numbers meaningful and easy to understand, it is important to first specify the VaR system. A succinct way to do this is to use the format in Table 2.

Table 2. VaR system specification format

Item	*Possible Choices*
Product valuation	Full revaluation/delta approximation/delta-gamma approximation
VaR methodology	Parametric VaR/historical simulation/Monte Carlo simulation
Observation window	250 days/500 days/750 days/1000 days
Scenario weighting	Equally weighted/exponentially weighted
Confidence level	95%/97.5%/99%
Return calculation	Relative return/absolute return/log return
Return period	Daily/weekly/10-day
Mean adjustment	Yes/no
Scaling (if any)	Scaled to 10-day/scaled to 99% confidence level

A VaR system specification defines how the VaR number is calculated. Some of the items were mentioned. For example, we discussed the advantages and disadvantages of full revaluation and approximation methods for product valuation. Considering the size and complexity of their trading portfolios, banks choose either of the methods for their VaR systems as long as they are able to justify the appropriateness of its application. For example, if a portfolio consists primarily of linear products, the delta approximation is superior to the full revaluation in terms of computational speed, while its accuracy is still acceptable. However, if we have a portfolio of options or other nonlinear products, the delta-gamma approximation, or even the full revaluation method, must be considered (in practice, one might choose different revaluation methods for different products).

To choose a suitable length of observation window, we must find a balance between the sensitivity of VaR to recent structural changes in the market and the accuracy of VaR estimation. A short window allows better sensitivity (or reaction) to the recent market moves, but it might not contain sufficient data to produce a reliable VaR estimate. A long window such as 1000 days means the VaR would hardly move even as the market enters a stressful situation. This would mean that the risk measure is late in registering risks at the onset of major crises. This dilemma can be addressed, for example, by using weighting schemes such as exponentially decaying weighted scenarios in the VaR model. This scheme weighs the scenarios further in the past by increasingly smaller weights, so that the resulting volatility figure is influenced more by recent fluctuations in prices that fluctuations far in the past. Thus, when the market experiences a crisis, exhibiting large swings, the new information contributes to the volatility quicker (Wong (2013) [3]).

The calculation of returns can be different across risk factors. The choice of return calculation is determined by distributional type of the risk factor. Generally, if the size of the movements is independent of the value of the risk factor, using absolute returns is appropriate, and if the size of the moves scales with the risk-factor, using relative or log return is appropriate. Often, a large portfolio involves numerous risk factors, and hence a real-world VaR system might contain a mixture of return definitions.

Basel requires banks to estimate their VaRs for a time horizon of 10 days, and at a confidence level of 99%. If a 10-day (non-overlapping) return period is used, we need a very long observation window to provide enough historical data for VaR estimation. This not only reduces the predictability of VaR, it is also limited by data availability. Using overlapping, 10-day returns is not recommended since it introduces an autocorrelation bias in the VaR because each datum (i.e., daily return) is reused ten times [4]. Basel allows banks to calculate VaR using daily data and then scale the VaR to 10 days using the square-root-of-time rule. Strictly, this is valid only if the daily returns, $r_i, \cdots, r_n$, follow a Gaussian distribution and are independent from one another, where the n-day return has a volatility of $\sqrt{n}$ times that of the daily returns.

Since quantile is proportional to volatility in a Gaussian model, this is equivalent to scaling a daily VaR by $\sqrt{n}$ to arrive at a n-day VaR. One has to be cautious that when the market is in a crisis and the distribution becomes fatter than normal, scaling produces an understated VaR.

Advanced methods deal with scaling, for example, using power law scaling, or by assuming returns follow a Gaussian AR(n) autoregressive process. See Wong (2013) for more details [**Error! Bookmark not defined.**].

VAR Methods

As mentioned, VaR is no more than a loss quantile of the P&L distribution. There are many ways to generate the distribution and compute the quantile. Conventionally, the three most common methodologies used by banks are the parametric VaR, the Monte Carlo simulation VaR, and the historical simulation VaR. A recent survey showed that most banks (about 73%) that report their VaR for regulatory purposes use the historical simulation VaR [5].

Table 3. Hypothetical portfolio

Product Type	Equity	*Product Type*	Option	*Product Type*	Bond
Asset	SPX Index	*Asset*	SPX Index Call	*Asset*	5Y T-Bond
Risk Feature	Linear	*Risk Feature*	Non-linear	*Risk Feature*	Non-linear
Notional	$1,000,000	*Notional*	-$1,500,000	*Notional*	$500,000
Spot	1848.36	*Option Type*	Call	*Settlement Date*	12/31/2013
		Maturity (yrs)	1	*Maturity Date*	12/31/2018
		1Y Zero Rate	0.31%	*Coupon Rate*	2.00%
		Dividend Rate	0.00%	*Yield Rate*	1.74%
		Strike	1848.36	*Redemption Value*	100
		Volatility	15.23%	*Coupon Frequency*	2
		Spot	1848.36	*Day Count Basis*	Actual/360
Present Value	$1,000,000	*Present Value*	-$93,268	*Present Value*	$506,173
				Total Present Value of Portfolio	
					$1,412,905

To explain VaR estimation methods better, we constructed a hypothetical portfolio as of 31 Dec 2013, shown in Table 3. It consists of three component assets:

1. A long position on S&P500 equity index (denoted SPX Index) that is mapped to a single risk factor, SPX spot;
2. A short position on an equity index option (i.e., 1-year call options on SPX Index). Its value, given by the Black-Scholes option pricing formula, is mapped to three risk factors: the SPX spot, the 1-year at-the-money (ATM) volatility, and the 1-year zero rate (i.e., discount rate);
3. A long position on a U.S. Treasury bond (i.e., 5-year, fixed-coupon treasury bond) that is mapped to a single risk factor, 5-year yield rate.

In the remainder of this section, we use this hypothetical portfolio to illustrate the three VaR methodologies.

Discussed earlier, although an asset's P&L distribution can be derived directly from price history of the asset, we need to perform risk factor mapping to reduce the dimensionality and simplify the calculation. For example, in the case of a Treasury bond, the market convention is to quote

its price in the form of yield rate. Mapping the bond position to a risk factor, say a 5-year yield rate, is natural because bonds (of a specific issuer name and tenor) have a unique yield, whereas their prices depend on the coupons.

This is especially helpful if our portfolio involves a large number of bonds with different maturities. For example, a bond maturing in 4.75 years should be mapped to 4.75-year yield rate. However, the yield curve is made of discrete benchmark points. The 4.75-year yield rate is usually not included in the yield curve. Instead, the standard tenors at 4-year and 5-year points are the closest adjacent points to 4.75-year. One way of dealing with this issue is to interpolate the curve and retrieve the 4.75-year yield, and use it during calculations.

Another way, which is, in practice, faster and without much loss of accuracy, is to split the bond into standard maturities; instead of $100 of a 4.75-year bond, we assume $75 of a 5-year bond and $25 of a 4-year bond. The advantage of this method is that ultimately we have only one bond per standard tenor, which reduces the amount of calculation required greatly.

Parametric VaR

JP Morgan's Riskmetrics popularized parametric VaR (pVaR), or variance-covariance (VCV) VaR. The original methodology was published in 1994, and quickly took hold as the industry standard. Strictly, Riskmetrics proposed modeling of VaR using the normal distribution and the EWMA volatility measure [5]. There are two assumptions here. First, the valuation method of pVaR is sensitivity-based, and assumes a linear dependence on the risk factors. This is equivalent to a Taylor expansion, whereby all second and higher order terms are ignored. For nonlinear products (e.g., options), if the risk factors involve large moves, this assumption might introduce inaccuracies to pVaR. Second, the pVaR method assumes the returns of risk factors follow a multivariate Gaussian distribution. This assumption is often violated when markets are under stress; during a crisis, distribution of risk factors can be fat tailed and drastically skewed. Since VaR deals with tail losses of the P&L distribution, when these two assumptions are violated, it could introduce large errors in VaR estimation.

Table 4. pVaR specification

Item	*Choice*
Product valuation	delta approximation
VaR methodology	parametric VaR
Observation window	500 days
Scenario weighting	equally weighted
Confidence level	97.5%
Return calculation	log return
Return period	daily
Mean adjustment	no
Scaling (if any)	scaled to 10 days

Table 5. Level and volatility of risk factors

	SPX	*1Y Zero*	*5Y Yield*	*SPX Vol*
Current Level	1,848.4	31.4	174.1	15.2
Unit	point	basis point	basis point	% point
Volatility (%)	0.75%	2.26%	4.10%	2.00%

Table 6. Correlation matrix of risk factors

	SPX	*1Y Zero*	*5Y Yield*	*SPX Vol*
SPX	1.00	0.14	0.12	-0.80
1Y Zero	0.14	1.00	0.00	-0.13
5Y Yield	0.12	0.00	1.00	-0.12
SPX Vol	-0.80	-0.13	-0.12	1.00

Table 7. Sensitivities to risk factors

	SPX	*1Y Zero*	*5Y Yield*	*SPX Vol*
Equity	$541	-	-	-
Option	-$437	-$71	-	-$5,956
Bond	-	-	-$240	-
Portfolio	$104	-$71	-$240	-$5,956

At the portfolio level, there are four risk factors: SPX, SPX ATM volatility, 1-Year zero rate, and 5-Year yield rate. The specification for the pVaR calculation is given in Table 4. We want to calculate VaR at a confidence level of 97.5% and a 10-day horizon. A 500-day observation window is used, ending at 31 December 2013, the date on which the VaR is calculated. Referring to the spreadsheet [VaR_Methods.xls], the steps the pVaR calculation are:

1. For simplicity, we take daily log return $r_{i,t}$ for all four risk factors, where $i = 1, \cdots, 4$ indexes the risk factors, and $t = 1, \cdots, 500$ is the historical scenario number (i.e., time sequence) in the observation window;
2. We calculate risk factor volatilities (i.e., sample standard deviation) $\varsigma = (\varsigma_1, \cdots, \varsigma_4)$ and correlation matrix ρ from the return data. The results are shown in Table 5 and Table 6, respectively;
3. First-order sensitivities to the risk factors are computed using the central difference method for each asset in the portfolio, and summed across assets to the portfolio level. For example, if we denote $\delta = (\delta_1, \cdots, \delta_4)$ the portfolio level sensitivities, its first entry (i.e., sensitivity to SPX) is calculated as $\delta_1 = \$541 - \$437 + \$0 = \$104$. This is shown in Table 7;
4. The P&L volatility regarding the i-th risk factor is calculated as $\sigma_i = \delta_i \varsigma_i f_i$, where f_i is the current level of the i-th risk factor (e.g., the level on the date the VaR is estimated, 31 December 2013). For example, the P&L volatility with respect to SPX is $\sigma_1 = \$104 \times 0.75\% \times 1{,}848.4 = \$1{,}443$. We end up with a vector of P&L volatilities $\sigma = (\sigma_1, \cdots, \sigma_4)$ for the portfolio;

5. Due to the diversification effect among risk factors, P&L volatilities must be aggregated through correlation matrix ρ to yield a total P&L volatility of portfolio σ_p:

$$\sigma_p^2 = \sigma\rho\sigma^T = \sum_{i=1}^{4}\sum_{j=1}^{4}\sigma_i\sigma_j\rho_{ij} = \$3{,}328 \quad (35)$$

which quantifies the volatility of the joint P&L distribution over the next day;

6. Given the assumption of normality, the quantile (or VaR) relates to the volatility by a simple factor. For example, the 1-day VaR at confidence level $\alpha = 97.5\%$ can be estimated as:

$$\begin{aligned}\mathrm{VaR}_{1d} &= \Phi^{-1}(1-\alpha)\times\sigma_p \\ &= -1.96\times\$3{,}328 \\ &= -\$6{,}522\end{aligned} \quad (36)$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative density function of standard normal distribution. In Microsoft Excel, it is given by the function $\mathrm{NORMSINV}(\cdot)$;

7. The square-root-of-time-rule applies to derive the 10-day VaR:

$$\mathrm{VaR}_{10d} = \mathrm{VaR}_{1d}\times\sqrt{10} = -\$20{,}624 \quad (37)$$

In step (4), the value of risk factor level f_i is involved in the formula to calculate the P&L volatility because volatility ς_i in the example was estimated based on log returns of the risk factor. This yield based (in the sense of relative changes) volatility must be converted to norm based (in the sense of absolute changes) by multiplying ς_i with f_i to account for volatility of absolute changes in risk factor levels. Hence, this scaling by f_i is unnecessary if absolute returns have been used. In this case, the P&L volatility is simply $\sigma_i = \delta_i\varsigma_i$. We use only first-order (i.e., delta) approximations for P&L valuation. Accuracy improves if we include second-order corrections, and the approach is then called the delta-gamma approach.

Monte Carlo VaR

pVaR imposes a strict assumption on the distribution of the risk factor returns, and it cannot handle non-Gaussian distribution properly. It is limited to what kind of instruments it can treat; generally, every product with which the second-order approximation is insufficient is not a good fit (and even second order can be a stretch if a product exhibits unusual correlations among risk factors). An alternative is the Monte Carlo VaR (mcVaR), where in principle, any risk distribution can be simulated, and any valuation method can be applied. In the following, we use an example to illustrate the mcVaR method step-by-step.

Table 8. mcVaR specification

Item	*Choice*
Product valuation	Full revaluation
VaR methodology	Monte Carlo Simulation VaR
Observation window	500 days
Scenario weighting	Equally weighted
Confidence level	97.5%
Return calculation	Log return
Return period	Daily
Mean adjustment	No
Scaling (if any)	N/A

Table 8 shows the VaR specification for our example. The mcVaR calculation uses the same test portfolio as in Table 3, and the same observation window mentioned. For simplicity, we still assume a joint Gaussian distribution for the risk factors. However, during implementation, more realistic distributional assumptions can be applied, for example, we might sample from a Student *t* distribution to model fat tails. Referring to the spreadsheet [VaR_Methods.xls], we trace the following general steps during the mcVaR calculation:

1. We take daily log return $r_{i,t}$ for all four risk factors, where $i = 1, \cdots, 4$ indexes the risk factors and $t = 1, \cdots, 500$ denotes the numbering across historical scenarios in the observation window;
2. We calculate risk factor volatilities (i.e., sample standard deviation) $\varsigma = (\varsigma_1, \cdots, \varsigma_4)$ and correlation matrix ρ from the return data, using the 500 days return data, the same results as in Tables 5 and 6;
3. Multivariate normal random numbers $\epsilon_k = \left(\epsilon_{k,1}, \cdots, \epsilon_{k,4}\right)$ are generated using *Cholesky* decomposition of the correlation matrix ρ. Technically, *Cholesky* decomposition is an algorithm in linear algebra that decomposes a symmetrical, positive-definite matrix (e.g., a correlation matrix) into a product of a lower triangular matrix and its transpose. The spreadsheet shows how this is done. See also the Professional Risk Manager's Handbook, Volume II. In the example, we draw 2,000 simulated scenarios such that $k = 1, \cdots, 2000$. Obviously, the larger the number of simulations, the more precise the mcVaR becomes (not accounting for parameter estimation errors). Typically, banks simulate more than 10,000 scenarios;
4. Use ϵ_k to simulate the k-th scenario for the risk factors. We need to specify a model for this stochastic process. In the example, we assume the risk factors follow a geometric Brownian motion (GBM) process. In the case of GBM model (with no drift), the simulated shifted (or shocked) risk factor level is:

$$\hat{f}_{i,k} = f_i \exp\left(-\frac{1}{2}\varsigma_i^2 \Delta t + \varsigma_i \epsilon_{k,i} \sqrt{\Delta t}\right) \qquad (38)$$

   where Δt is the time increment, $i = 1, \cdots, 4$ is the index of risk factors, and f_i is the current level of the i-th risk factor (shortly, we discuss this formula in more detail).
5. The portfolio P&L of the k-th scenario is just the sum of the present values of the assets priced (using full revaluation or pricing function) at the *shifted* levels $\hat{f}_{i,k}$ of relevant risk

factors *minus* the sum of the present values of the assets priced at *current* levels f_i. Repeat this for all k to obtain a portfolio P&L vector with 2000 entries;

6. We take the 0.025-quantile of this P&L vector to give the 97.5% confidence level daily VaR.

In this illustration, we assume all risk factors follow a GBM stochastic process, which is a common simplification practiced in the industry:

$$df_t/f_t = \mu dt + \sigma dW \tag{39}$$

where μ is a deterministic drift (or long-term rate of return) and σ is the volatility. The dW is known as a Wiener process, and can be written $dW = Z\sqrt{dt}$, where Z is a random draw from a standard normal distribution. In other words, dW is normally distributed, with mean of zero and variance of dt. Substituting for dW in (39), we get:

$$df_t/f_t = \mu dt + \sigma Z\sqrt{dt} \tag{40}$$

The instantaneous rate of change in the risk factor, df_t/f_t, evolves according to its drift term μdt and the random term $\sigma Z\sqrt{dt}$. In practice, we would implement the model (or code it) in its discrete form. To reduce numerical instability, (40) is often discretized in logarithmic form (i.e., given Δt is a small time increment):

$$\Delta \ln f_t = \mu \Delta t - \frac{\sigma^2}{2}\Delta t + \sigma Z\sqrt{\Delta t} \tag{41}$$

where $\Delta \ln f_t = \ln f_{t+\Delta t} - \ln f_t$ is the change in the logarithm of the risk factor over the time interval Δt (the term $-\frac{1}{2}\sigma^2\Delta t$ comes from the Ito lemma to offset the bias introduced by the logarithm transformation). The shifted level of the risk factor then becomes:

$$f_{t+\Delta t} = f_t \exp\left(\mu \Delta t - \frac{1}{2}\sigma^2\Delta t + \sigma Z\sqrt{\Delta t}\right) \tag{42}$$

(41) assumes that the log return of the risk factor is normally distributed. Hence, our criticisms of parametric VaR with respect to the normality assumption also apply to the mcVaR methodology, unless we employ more realistic dynamics than the GBM model with constant volatility (e.g. stochastic volatility models).

Sometimes, we need to simulate risk factor f over time T up to the horizon of the VaR. We usually divide T into a number N of small time increments Δt (i.e., we set $\Delta t = T/N$). We take the current level of f as a starting value and draw a random sample to update f using (41). This gives the change in f over the first time increment, and we repeat the process to evolve the changes of f over all N increments up to risk horizon T. This is one simulated scenario. We then repeat the exercise many times to produce as many simulated scenarios as needed. Since we assume constant μ and σ in our GBM model, dividing T into N small time increments is unnecessary; the risk factor can be evolved in *one* step, with $\Delta t = T$. For a 10-day VaR, we let $\Delta t = 10$ in our Monte Carlo simulation given that our volatilities ς_i are estimated from daily returns (i.e., not annualized). Multistep simulation is needed in some cases in which the pricing of products depends on the nature of the paths taken by risk factors.

Monte Carlo simulation VaR offers many advantages over parametric VaR, including:

1. With the help of suitable distributional assumptions, it can capture a wider range of market behavior;
2. Since P&L is computed by full revaluation (and not sensitivity-based approximation), it can handle nonlinear products, including exotic options and structured financial instruments;
3. With a sufficient number of simulations, it can provide detailed insights into extreme losses that lie far out in the tails of the distributions, beyond the usual VaR cutoff.

However, there are also considerable drawbacks to mcVaR. Monte Carlo simulation is computer intensive not only because mcVaR uses full revaluation during P&L calculations, but also because of the large number of simulated scenarios needed to estimate mcVaR precisely. Generally, if we want to double the precision, we end up quadrupling the number of simulated scenarios (and this assumes Gaussian distributions; on more tailed distributions, the scaling behavior is even worse). This square root convergence is slow and greatly limits the power of Monte Carlo simulation.

mcVaR also depends highly on how the return distribution is modeled. Scenarios generated by Monte Carlo simulation must be consistent with the historical characteristics of the market. Although we know the Gaussian distribution is too idealistic to hold, the real joint distribution of a dynamic market remains unknown and is difficult to model. The stochastic process that drives the risk factor needs to be modeled. Although a variety of models have been developed in academia to describe the observed market dynamics of asset classes, none is perfect; each model has its advantages and disadvantages. With the exception of GBM, most models require calibration to determine their parameters. Unfortunately, such calibration is often limited by data availability, and the parameters themselves are seldom constant so recalibration is required periodically.

In practice, the banking industry often uses simple Gaussian simulations because it is too cumbersome to model each risk factor class separately. It is important that the student understands the limitation of simple models used in a realistic setting, and this often leads to understatement of risks.

Historical Simulation VaR

Historical simulation VaR (hsVaR) gained popularity in recent years. Most banks that disclose their VaR method report using historical simulation [5]. The hsVaR is different from previous two methods in that it does not assume a distribution on returns. It is a Monte Carlo simulation method that uses *historical* samples for its scenarios rather than samples drawn from *theoretical* distributions. It overcomes many drawbacks that plague pVaR and mcVaR. First, by sampling from empirical or historical data, it avoids the need to make any distribution assumption. This takes care of fat tails and skewness because we let the data decide the shape of the distribution. There is one distribution for each risk factor. Second, hsVaR usually computes the P&L of each product using full revaluation. Given the complexity of today's derivatives, most portfolios are non-linear. Full revaluation avoids errors arising from delta (or

delta-gamma) approximation. Third, risk aggregation is conducted by summation of P&L across products to form the portfolio P&L. The dependence structure among risk factors is accounted for in this way (i.e., you get the correlation for free), and there is no need to maintain a large correlation matrix as pVaR and mcVaR require. This simple cross summation is easy to understand intuitively and allows for easy risk decomposition during analysis.

Table 9. hsVaR specification

Item	*Choice*
Product valuation	full revaluation
VaR methodology	Historical Simulation VaR
Observation window	500 days
Scenario weighting	equally weighted
Confidence level	97.5%
Return calculation	log return
Return period	daily
Mean adjustment	no
Scaling (if any)	N.A.

Table 9 shows an hsVaR specification for our example. The hsVaR calculation uses the same test portfolio as in Table 3, and the same observation window as before. Referring to the spreadsheet [VaR_Methods.xls], we summarize the steps in computing hsVaR:

1. A scenario vector is formed from the observation window using the historical daily log returns $r_{i,t}$, where $i = 1, \cdots, 4$ indexes the risk factors and $t = 1, \cdots, 500$ numbers the historical scenarios along the observation period;
2. The shifted levels of the risk factors are calculated for the t-th scenario (e.g., we have $\hat{f}_{i,t} = f_i \exp(r_{i,t})$ for log returns or $\hat{f}_{i,t} = f_i + r_{i,t}$ for absolute returns, and $\hat{f}_{i,t} = f_i(1 + r_{i,t})$ for relative returns), where f_i is the current level of the i-th risk factor;
3. For an asset, P&L of the t-th scenario is just the present value priced (using full revaluation) at the *shifted* levels $\hat{f}_{i,t}$ of relevant risk factors *minus* the present value priced at the *current* levels f_i. Do this for all 500 scenarios to derive a P&L vector for that asset (i.e., a vector of 500 P&L's). Repeat this for all assets in the portfolio;
4. Sum the P&L vectors for all assets *by scenario* to derive the P&L vector for the entire portfolio, which is a distribution with 500 points. Take the 0.025-quantile of this P&L vector to produce the 1-day VaR at a confidence level of 97.5%.

In essence, we are trying to project the P&L one step (1 day) into the future from today, but using random draws (or scenarios) from the past sample period. Thus, we always apply the historical returns to current risk factor levels and current portfolio positions. In step (3), hsVaR might also use approximation methods (i.e., the sensitivity approach) to price assets for the purpose of reducing computational complexity. In step (4), the cross summation of asset P&L (by scenario) allows for a flexible risk decomposition and diagnosis. For example, if we are interested in knowing the risk due to equity spot alone, we can shift only the equity spot risk factor and repeat the steps above. If we are interested in a VaR breakdown by sub-portfolio, we

can do the P&L vector summation for each sub-portfolio separately. If we are interested in the impact of a single asset, we can exclude the P&L vector of that asset from the summation and look at the difference, known as the incremental VaR.

Although hsVaR has been used widely in industry, it remains subject to a few subtle weaknesses:

1. The returns are assumed independent for non-overlapping periods, and the historical scenarios are assumed to provide good guidance for tomorrow's risk. In reality, market risk parameters are not independent of time; they change dynamically, often suddenly, such as at the onset of a crisis. Hence, hsVaR lags major market shocks. This weakness (i.e., lateness) is generally true for any VaR method that relies on history;
2. The simulation uses a limited number of samples from historical scenarios; typically 1 to 2 years of data. Hence, there can be a large error in the estimated VaR. In fact, the higher the confidence level we set, the larger the error in the estimated VaR. As with mcVaR, precision can be improved by increasing the number of scenarios, but this is somewhat restricted for hsVaR because extending the window length reduces the sensitivity to detect regime changes in the market;
3. The aggregation of P&L vectors by scenario is affected by data quality. The price series must not be unusually quiet (or stale) and must not be too choppy (have spikes). Otherwise, the portfolio VaR becomes unstable because the tail of an empirical distribution is not as smooth as a theoretical one used with pVaR and mcVaR. This problem is pronounced for short observation periods, high confidence levels, or small portfolios. Thus, the sampling error (precision) of hsVaR is generally poorer;
4. Although the method appears objective, it is not; one still has to decide how a price change, say from $200 to $220, compares with another, say at $400 (e.g., to $440 or $420), and those choices have no trivial implications;
5. A major drawback is that new asset classes often get unrealistic risk figures. For example, in the run up to the 2008 crisis, ABS spreads had been constantly tightening, so much so that even at a very high quantile, the worst loss was a gain.

Simulation of Interest Rates

Interest rate markets are much more complex than markets for other asset classes such as equities, currencies, or (most) commodities. Even for a single currency, there are many interest rate markets of different credit worthiness, for example, the Treasury bond rates, inter-bank offered rates (LIBOR), corporate bond rates, mortgage rates, etc. Although these rates are affected by common macroeconomic factors, they do not move in sync due to structural and credit disparities. We cannot use a single rate to describe the overall interest rate level of the market; interest rate is generally a function of term (or tenor). This function is often called the interest rate term structure, or yield curve. The yield curve slopes positively during normal times, and its shape changes dynamically over time. The co-movements and dynamics of a yield curve complicate the risk measurement of a portfolio. Suppose a bank holds a bond portfolio that contains a collection of various bonds with different maturities. Each bond has a yield rate associated with its maturity. The portfolio has risk exposures along all the yield rates at various

maturities. When we perform risk analysis, it becomes important and natural to look at the shapes and movements of the whole yield curve.

Term Structure of Interest Rates

Figure 1 shows a typical yield curve of U.S. Treasury bonds observed on 31 December 2013. The yield curve is usually represented as a discrete set of benchmark rates of different standard terms (i.e., the points in the chart); these are often used for risk factor mapping. The shape of the yield curve indicates the current state of supply and demand and the cost of borrowing of that debt market. The longer dated the loan (or bond), the larger the credit and inflation risks. Lenders thus offer long-term loans at higher interest rates than they offer for short-term loans. This acts as a premium to compensate for the default and inflation risk the lender is exposed to. Hence, the yield curve typically slopes upward. Occasionally, a yield curve can become inverted as long-term yields fall below short-term yields, indicating the market anticipates lower interest rates in the future, and borrowers are seeking short-term loans more aggressively than long-term loans. This is often seen as a forward indication of economic downturn.

A borrower's creditworthiness (represented by a credit rating) is another driver of a yield curve. Governments often issue bonds in local currency, and a series of such bonds make up the government bond yield curve, or sovereign curve or govies. Governments are deemed to be risk-free and enjoy the highest credit rating in their own currencies since the central bank can always print local currencies to pay off maturing government debt. Thus, govies naturally have the lowest interest rates in the market. This working assumption breaks down in exceptional cases such as the Russian bond default in 1998 and the Eurozone crisis in 2009. Banks with a high credit rating (e.g., AA) borrow money from each other at LIBOR (or the interbank rate). LIBOR curves are typically a little higher than government curves to account for the lower creditworthiness of banks. Corporate curves are another category of yield curves, constructed from the yields of bonds issued by corporations. Since corporations typically have higher perceived credit risk than governments and banks, they have to offer higher yields on their bonds to attract investors.

Yield curves shift and evolve daily, reflecting market reactions to news and perceptions of supply and demand. The yield curve exhibits various degrees of freedom in its motion—parallel shifts, changes in slope (e.g., flattening or steepening), and changes in curvature (e.g., bowing). Short-term rates are more volatile than long-term rates because short rates are affected by central bank monetary policy actions/expectations, and central banks manage this. The long end is driven primarily by structural factors such as inflation expectations and supply/demand of bonds. The super-long end (more than 10 years) is also dominated by long-term (i.e., buy and hold) institutional investors such as pension funds and insurance companies.

When the shape of the yield curve changes, it affects the present value of the portfolio, and this gives rise to P&L for positions. A measure of how sensitive the position P&L is to unit change in yield curve movement is given by its PV01 (Table 1), but this considers only parallel movements in the yield curve. To study the full dynamics of the yield curve, a risk manager typically uses a common technique called Principal Component Analysis.

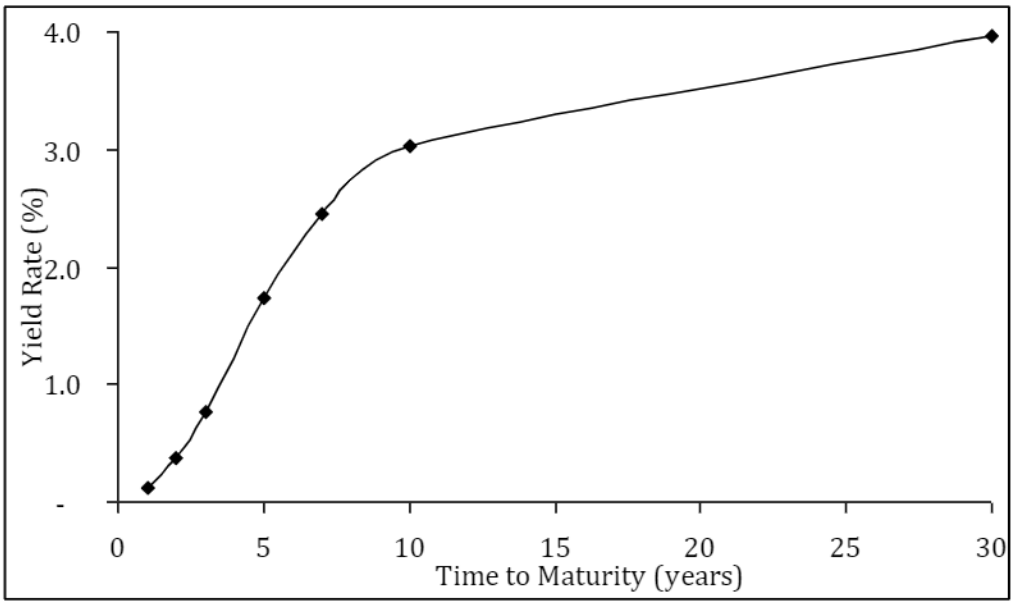

Figure 2. Yield curve of U.S. Treasury bonds as of 31 December 2013

Principal Component Analysis

Principal Component Analysis (PCA) is a mathematical procedure that performs an orthogonal linear transformation that changes a Gaussian vector with a non-trivial covariance matrix into a standard Gaussian vector in which variables are uncorrelated. The new variables are called Principal Components, and are sorted by descending variance. To reduce the dimensionality of a problem, people concentrate on, say, the first three risk factors that capture, say, 80% of the overall variance, reducing a high dimensional problem to a three dimensional one. For example, when modeling a yield curve, the first factor tends to be a parallel shift, the second factor a change in slope, and the third a change in the convexity of the curve. This describes the majority of moves seen in practice, and so the 20-tenor/20-dimensional problem reduces to a 3-dimensional problem, where all 20 tenors are reconstructed from those 3 factors.

There are two issues with this approach however. First, 80% of the variance captured, a typical value, is not as impressive as it sounds because the standard deviations, which are much more relevant figures in this context, are the square roots of the variance. So, in this example, if we explain 80% of the variance (and hence do not explain 20% of it), this means, for example, that we have two random variables, one with volatility 100 and the other (independent) one with volatility 50, and we ignore the second. The ignored risk is bigger than it sounds when we say, "We explain 80% of the risk."

The second issue is that ultimately, it does not matter what percentage of the variance of our risk factors we explain, but we are interested in the portfolio variance. Let us stay in the interest rate space and assume we are using a two-factor model, meaning we explain only level and slope. Any convexity product (e.g., long 1-year and 5-year, double short 3-year) shows up as zero risk in this model, which is evidently incorrect.

PCA is best suited for highly correlated systems and systems that have common economic drivers. Typical examples include the stock market, interest rate yield curves, futures prices across tenors, and option volatility surfaces. The way PCA works in practice is we start from a covariance matrix Σ. Since Σ is symmetrical and positive (semi-) definite, eigen-decomposition can be used to decompose the Σ into:

$$\Sigma = E\Lambda E^T \tag{43}$$

where Λ is a diagonal matrix with non-negative entries being the eigenvalues in descending order and E is an orthogonal matrix with columns being the eigenvectors of a unit length. A non-zero (column) vector e is an eigenvector of a square matrix Σ if and only if it satisfies linear equation:

$$\Sigma e = \lambda e \tag{44}$$

where λ is a non-zero scalar called the eigenvalue corresponding to the eigenvector e. In other words, the linear transformation by matrix Σ merely elongates or shrinks the eigenvectors, and the amount they elongate or shrink by is the eigenvalue. The orthonormal matrix E resulting from the eigen-decomposition can be used to transform the n risk factors into n latent variables, called the principal components, such that:

$$p = rE \tag{45}$$

where $p = (p_1, \cdots, p_n)$ is a n-dimensional row vector. Since E is orthonormal, thinking geometrically, the transformation is just a simple rotation of the coordinate system. The principal components, $p_1, \cdots, p_n$, are uncorrelated and decreasingly responsible for the overall variation in the risk factors. For example, $p_1 = \sum_{i=1}^{n} r_i E_{i,1}$ is uncorrelated with $p_2 = \sum_{i=1}^{n} r_i E_{i,2}$ (i.e., the covariance between p_1 and p_2 is zero). The variance of p_1 is given by the first and largest diagonal entry of matrix Λ, that is, $\sigma_{p_1}^2 = \Lambda_{1,1}$, while the variance of p_2 is given by the second diagonal entry $\Lambda_{2,2}$, etc.. Figure 3 shows the two principal components derived from a 2-D joint normal distribution. Seen geometrically, the PCA transformation is a rotation of the coordinate system from the chart's axes to an orthogonal system represented by the slanted L. Referring to the spreadsheet [VaR_PCA.xls], the joint normal distribution is formed by two marginal distributions with volatility of 2.0 and 1.0 respectively, correlated by a coefficient of 0.8. Eigen-decomposition shows that the first principal has a variance of 4.69, and the second of only 0.31. In contrast, at correlation equal to zero, the two margins are already orthogonal; the principal components are the margins themselves (Figure 4).

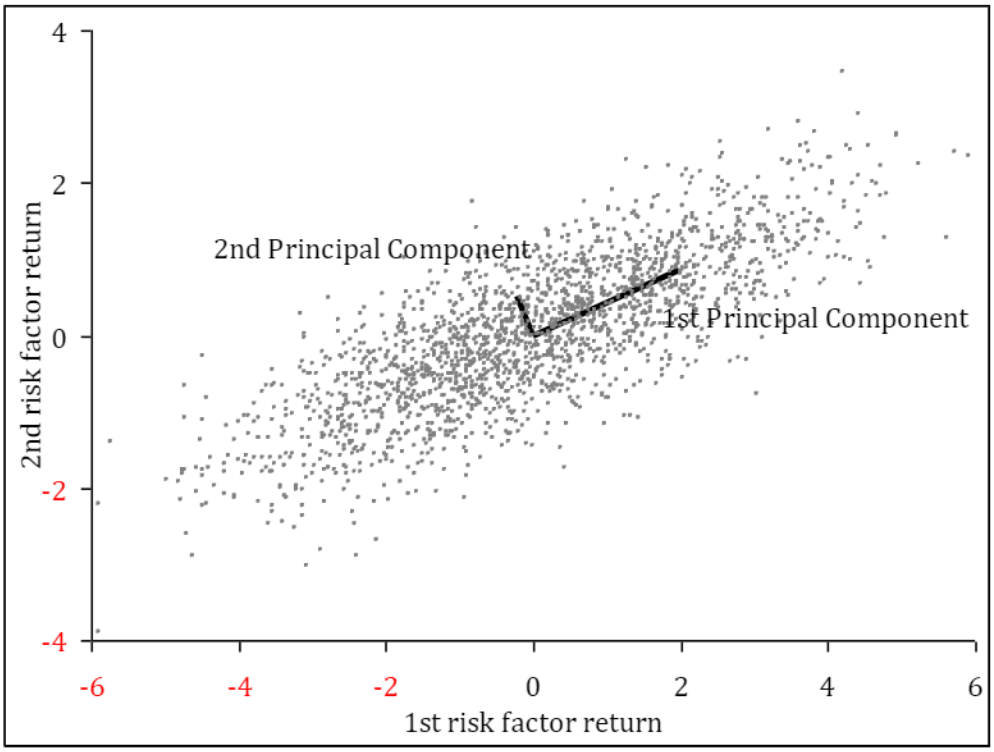

Figure 3. Principal components from a correlated 2-D joint normal distribution. The standard deviations of the cloud along the x-axis and y-axis give the volatilities of the two marginal distributions (2.0 and 1.0 respectively).

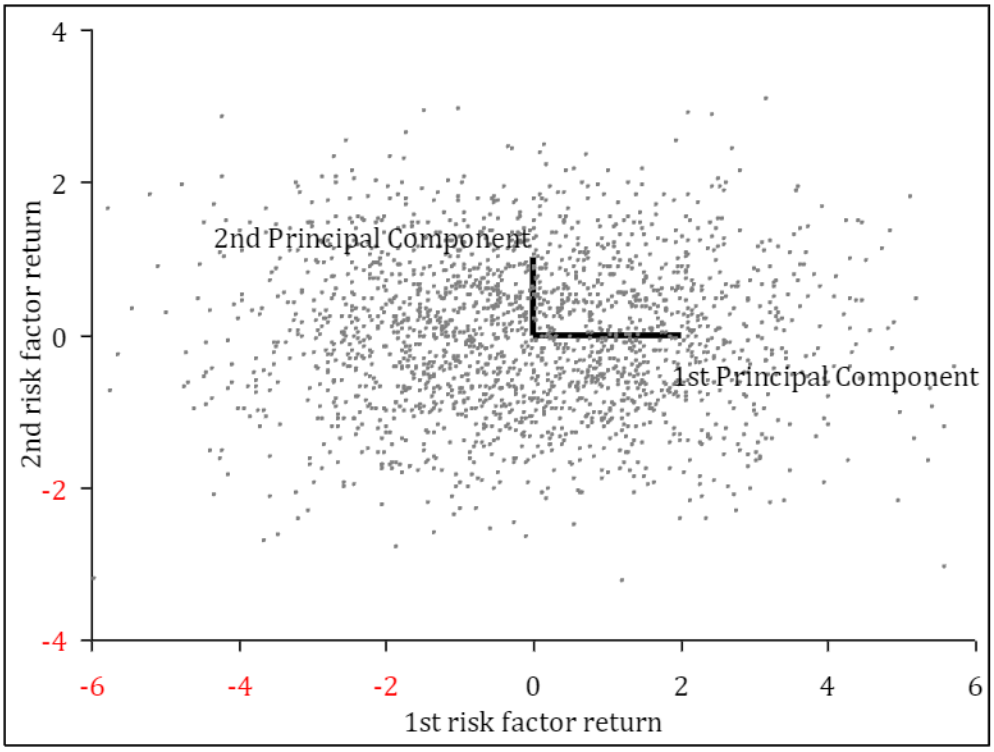

Figure 4. Principal components from an uncorrelated 2-D joint normal distribution

Reconstruction of the risk factors, r, from principal components, p, is straightforward. Since E is orthonormal, the inverse of E is its transpose. The principal components can then be transformed back to original risk factors by:

$$r = pE^T \tag{46}$$

To reduce the dimensionality, we might form a vector, $\hat{p}$, that includes only a few, but not all, of the most significant principal components. For example, $\hat{p} = (p_1, \cdots, p_k, 0, \cdots, 0)$ includes only the first k principal components. The $\hat{p}$ is then transformed by matrix E^T to give an approximation, $\hat{r} = \hat{p}E^T$, of the original risk factors, r. Since the ignored principal components, $p_{k+1}, \cdots, p_n$, are insignificant, $\hat{r}$ still captures most of the variation in $\hat{p}$. In other words, we reduce the dimension of the problem from n to k without sacrificing much accuracy. This is why the entire yield curve can be represented by only three components. PCA is also useful for scenario analysis and stress testing. Through reconstruction, we can generate scenarios of r consistent with its historical observations by introducing shocks to the main principal components. For example, we can separately quantify how large a parallel shift, steepening, and bowing can be in history, and then shock the three principal components by these magnitudes.

PCA performs effectively when risk factors correlate strongly. If the relationship is weak among risk factors, PCA does not work well to reduce dimensionality. Generally, if most of the correlation coefficients are smaller than 0.3, PCA will not help. In the remainder of this section, we illustrate the performance of PCA under various conditions using an example. Referring to the spreadsheet [VaR_PCA.xls], three risk factors form a row vector $r = (r_1, r_2, r_3)$, each following a univariate normal distribution with mean of zero and volatility of 4.0, 3.0 and 2.0, respectively. To illustrate the performance of PCA for a strongly correlated system, we simulate a joint distribution for the risk factors using the correlation matrix shown in Table 10. The correlation matrix shows that r_1 and r_3 correlate strongly with r_2, whereas the correlation between them is weak. The correlation matrix, along with the risk factor volatilities, defines a covariance matrix (Table 10), which PCA takes as input to perform eigen-decomposition. The resulting eigenvectors and eigenvalues of the covariance matrix are shown in Table 11. The first principal component given by the first eigenvector $(0.78,\ 0.58,\ 0.23)^T$ has a variance of 23.86, which is greater than the second principal component, and overwhelmingly larger than the third, whose variance is only 0.07. Since the first two principal components capture most of the variation in the correlated system, we can rely only on them to reconstruct an approximation of the system. The omission of the third principal component is trivial since the covariance and correlation matrix of the reconstructed risk factors, shown in Table 12, differ only slightly from the original ones (Table 10).

Table 10. Correlation and covariance matrix of risk factors (strongly correlated system)

Correlation Matrix			*Covariance Matrix*		
1.00	0.80	0.30	16.00	9.60	2.40
0.80	1.00	0.80	9.60	9.00	4.80
0.30	0.80	1.00	2.40	4.80	4.00

Table 11. Eigenvalues and eigenvectors of covariance matrix (strongly correlated system)

Eigenvalues, Λ			*Eigenvectors,* E		
23.86	-	-	0.78	-0.54	0.32
-	5.08	-	0.58	0.43	-0.69
-	-	0.07	0.23	0.72	0.65

Table 12. Covariance and correlation matrix of reconstructed risk factors using the first two principal components (strongly correlated system)

Correlation Matrix			*Covariance Matrix*		
1.00	0.80	0.30	15.99	9.61	2.39
0.80	1.00	0.81	9.61	8.97	4.83
0.30	0.81	1.00	2.39	4.83	3.97

In a weakly correlated system, PCA is generally less effective. We consider a correlation matrix, shown in Table 13, in which risk factors are weakly correlated. The eigen-decompostion of the covariance matrix shows all principal components (Table 14) are significant, which means omission of even the least significant principal component in reconstruction might result in large errors. This can be seen in Table 15, in which both the covariance and correlation matrix of the reconstructed risk factors differ from the original ones (Table 13). The impact is especially large for the third risk factor, whose variance reduced from 4.00 to 0.45.

Table 13. Correlation and covariance matrix of risk factors (weakly correlated system)

Correlation Matrix			*Covariance Matrix*		
1.00	0.20	0.10	16.00	2.40	0.80
0.20	1.00	0.20	2.40	9.00	1.20
0.10	0.20	1.00	0.80	1.20	4.00

Table 14. Eigenvalues and eigenvectors of covariance matrix (weakly correlated system)

Eigenvalues, Λ			*Eigenvectors,* E		
16.84	-	-	0.95	-0.32	-0.02
-	8.44	-	0.30	0.93	-0.21
-	-	3.72	0.09	0.19	0.98

Table 15. Covariance and correlation matrix of reconstructed risk factors using the first two principal components (weakly correlated system)

Correlation Matrix			*Covariance Matrix*		
1.00	0.20	0.33	16.00	2.38	0.88
0.20	1.00	0.99	2.38	8.83	1.97
0.33	0.99	1.00	0.88	1.97	0.45

PCA in Interest Rate Simulations

To demonstrate application of PCA to interest rate simulations, we examine the U.S. Treasury bond curve with tenors 1, 2, 3, 5, 7, 10, and 30 years. Daily returns for the rates are calculated using four years of historical data, from 2010 to 2013. The covariance matrix derived from the return data is subjected to PCA analysis. Results are shown in Table 16. There are 7 principal components derived from the covariance matrix, ranked from left to right in terms of the amount of variance they explain. The first principal component is denoted PC1, the second PC2, etc. Denoting the yield returns as $r_{1Y}, r_{2Y}, \cdots, r_{10Y}$, the first principal component, for example, is formed by a linear combination of the returns, with factor loadings shown in the column associated with PC1:

$$PC_1 = 0.37 \times r_{1Y} + 0.53 \times r_{2Y} + 0.53 \times r_{3Y} + \cdots + 0.13 \times r_{10Y} \qquad (47)$$

The remainder of the principal components can be constructed similarly. The formula above suggests that if we introduce a one-basis-point change in PC1, it corresponds to a 0.37 basis-points change in r_{1Y}, a 0.53 basis-points change in r_{2Y}, etc.

Table 16. Summary of PCA on U.S. Treasury bond yield curve

Summary of Principal Components

		PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*
Variances, λ (bps)		92.37	21.97	6.95	1.83	0.64	0.23	0.07
Factor Loadings	*1Y*	0.37	-0.90	0.21	-0.02	-0.01	0.00	0.00
	2Y	0.53	0.05	-0.66	0.52	-0.08	-0.00	0.00
	3Y	0.53	0.19	-0.17	-0.72	0.37	-0.07	-0.01
	5Y	0.39	0.25	0.33	-0.10	-0.63	0.52	0.03
	7Y	0.29	0.21	0.41	0.18	-0.14	-0.73	0.33
	10Y	0.21	0.16	0.36	0.28	0.29	-0.02	-0.80
	30Y	0.13	0.11	0.29	0.31	0.60	0.43	0.50
					Total Variance (in basis points):			124.06

Percentage of Variance Explained

PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*
74.45%	17.71%	5.60%	1.48%	0.51%	0.19%	0.06%

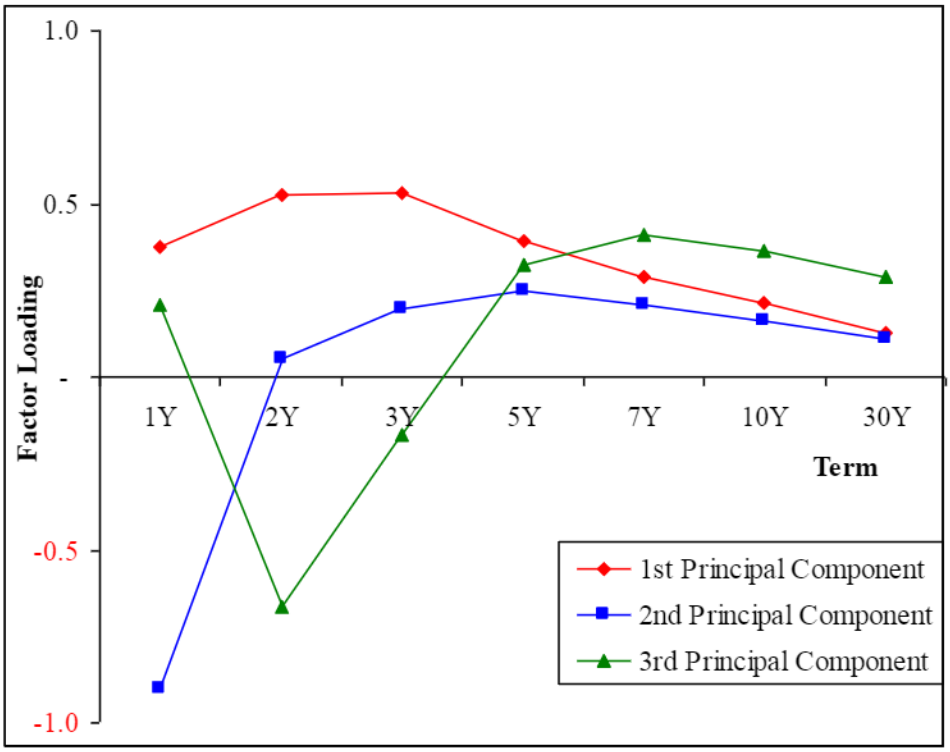

Figure 5. Factor loadings of the top 3 principal components

Figure 5 shows the factor loadings of the top 3 principal components. The PC1 has a positive and nearly flat shape of factor loadings for all terms. This can be interpreted as a parallel shift in the yield curve (all coefficients of the first eigenvector always have the same sign). Table 16 shows that PC1 alone explains 74.45% of the total variance observed in the yield curve (i.e., a variance of 92.37 for PC1 out of a total variance of 124.06 for the entire yield curve), implying the yield curve tends to move in parallel (as defined by the eigenvector, with moves in the short term roughly twice the size of moves in the long term) most of the time. The factor loadings for PC2 form an upward-sloping shape, negative for short terms and positive for longer terms. In other words, PC2 correlates negatively with short-term rates and positively with long-term rates, which corresponds to flattening or steepening movements of the yield curve. The factor loadings for PC3 are positive at both short and long terms, but negative at mid-terms, corresponding to a change of curvature in the yield curve.

The top three principal components combined explain 97.76% (=74.45% + 17.71% + 5.60%) of total variance. Hence, the dynamics of the yield curve can be approximated well using only three principal components. That is, for VaR estimation, we map the portfolio positions to the three principal components as risk factors rather than the seven yield rates, which reduces the dimensionality of the risk factors, without distorting the distribution and its risk features.

With pVaR, the remapping of risk factors is achieved by transforming the risk sensitivities concerning the seven yield rates to the three principal components. This can be done using a similar formula shown in (47). For example, if we have the first-order sensitivities (i.e., PV01) of

a portfolio, $\delta_{1Y}, \delta_{2Y}, \cdots, \delta_{10Y}$, corresponding to the yield rates, we can transform them into the sensitivities with respect to the principal components using the factor loadings. In the case of PC1, the transformation reads:

$$\delta_{PC1} = 0.37 \times \delta_{1Y} \times y_{1Y} + 0.53 \times \delta_{2Y} \times y_{2Y} + \cdots + 0.13 \times \delta_{10Y} \times y_{10Y} \qquad (48)$$

where δ_{PC1} is the PV01 with respect to PC1 and $y_{1Y}, y_{2Y}, \cdots, y_{10Y}$ are the current levels of the rates (e.g., at the time the VaR is estimated). We again scale the PV01 by the corresponding rate level because the principal components are derived using log returns; its sensitivity must be scaled to a yield based value. If absolute returns were used, the scaling could be omitted. Since the principal components are uncorrelated, risk aggregation is a simple summation. The portfolio P&L volatility is:

$$\sigma_p^2 = \delta_{PC1}^2 \times \lambda_{PC1} + \delta_{PC2}^2 \times \lambda_{PC2} + \cdots + \delta_{PC3}^2 \times \lambda_{PC3} \qquad (49)$$

where λ_{PC1} is the variance of PC1, λ_{PC2} the variance of PC2, etc. We know the top three principal components dominate the variance, so we can calculate $\delta_{PC1}, \delta_{PC2}, \text{and } \delta_{PC3}$, and ignore higher components terms in (49) to estimate an approximate portfolio P&L volatility $\hat{\sigma}_p$. The $\hat{\sigma}_p$ can then be used to estimate the pVaR using formula(36).

With hsVaR, we want to use historical scenarios of the principal components to simulate the yield curve evolution. This involves three steps: a) construct return series for principal components, b) use the return series of principal components to reconstruct return series for yield rates, and c) use the reconstructed yield rate returns to evolve the yield curve. The first step can be done using formula (47). For every historical scenario, we use the factor loadings to transform returns of yield rates into returns of principal components. We retain only the top three principal components and exclude the rest. We reconstruct the (approximated) rate returns from them using formula (46) for each scenario. The yield curve is then evolved using the reconstructed rate returns. For a single scenario date, the evolved yield curve can then be used to reprice the portfolio to give the full revaluation P&L. By repeating repricing for all historical scenarios, we obtain a P&L distribution. hsVaR is just the quantile of this distribution. The same strategy applies to mcVaR. In this case, normal distributions are assumed for the principal components. Hence, we sample the principal components from independent normals with respective variances $\{\lambda_{PC1}, \lambda_{PC2}, \lambda_{PC3}\}$ rather than sample from the historical scenarios.

Weaknesses and Limitations of the Value-at-Risk Model

The 2008 global financial crisis was an expensive lesson for the banking industry, revealing that the risk models many banks used, particularly VaR, were inadequate at capturing risks. The VaR model was criticized as being 'too little, too late' in that it is often underestimated and late at forecasting. It was also found that the model breaks down during a crisis, and is useful only as a peacetime tool. Numerous model weaknesses were revealed during that stressful period,

leading to development of more sophisticated Basel III risks models, and exciting risks research in academia. This section gives a high-level summary of weaknesses known regarding the VaR model.

Not All Risks are Modelable

After the crisis, banks were criticized for over-relying on models for risk management. Clearly, not all risks can be modeled. A good working assumption is that any risk that is not of an actuarial nature (i.e., not statistically observable) cannot be modeled with VaR. Some of these risks include the risk of currency controls, global event risks (such as abandonment of the dollar as a reserve currency), reputation risks, war and its impact on cross-border payment system, banking scandals, etc. For such risks, stress testing is a possible solution (see Chapter 6 written by David Rowe in this Handbook).

Nevertheless, due to regulatory requirements, the banking industry (or to be exact, some qualified banks) still model VaR for operational risks, where the data set that supports its calculation is often scarce. This is not an ideal situation since it means that the measurement error for such OpVaR is huge (possibly in the same order of magnitude as OpVaR itself).

Liquidity Effects

Liquidity risks are absent from VaR input. These include the effects of bid-offer spreads dynamics and price impact. The latter refers to impact of the volume of trades on price stability. For example, the sale of a $1 billion trade at once pressures a price to shift downward more strongly in comparison to the sales of a $100 million trade ten times over the course of a day. Before the crisis, there were research papers from academia that tried to incorporate liquidity risk (i.e., bid-offers) into VaR, the so-called liquidity-VaR or LVAR. After the crisis, the Basel Committee somewhat influenced the course of model development by requiring that liquidity risk be categorized by product and modeled in terms of risk horizons. Such Basel III models are still being developed (or fine-tuned) at major global banks, and the industry has yet to agree on a standard best practice.

Losses Beyond VaR

VaR is the quantile/cutoff point at the left tail of the loss distribution. Thus, VaR is oblivious to the loss points larger than VaR. Two distributions with different tail shapes can have the same VaR value but have very different extreme risk profiles. Another measure that has long been used in this context is expected shortfall (i.e., expected loss), conditional within a quantile in the tail.

Mapping Issues and Historical Data Limitations

One issue when running market risk VaR models concerns historical data. In a typical bank, risk factor mapping contains an order of magnitude of 100,000 risk factors, each a time series. This is a tremendous amount of data to clean, process, and maintain in the bank's system. Among data issues, the most insidious is the problem of data asynchrony. That means risk factors' time series are discordant, and hence the portfolio correlation structure is broken (and VaR blatantly incorrect). This can happen for a number of reasons:

1. Two data series are snapped at different times. For example, a USD 10-year swap snapped at London close versus NY close. Using data from January 2011 to December 2012, the effect on 99% VaR of a typical portfolio can be 13% due to this effect alone;
2. Two data series are snapped simultaneously but their markets close at different time zones. For example, a USD-denominated Asian bond closes at Tokyo close, but its hedge, a USD swap, closes at NY time;
3. Two data series from markets from the same time zone are snapped at the same closing time, but one is less liquid and hence its quotes are updated less frequently. For example, a more liquid CDS hedging a less liquid cash bond of the same obligor name.

The problem of data integrity often affects hedged positions, and that causes misrepresentation of basis risks. The LTCM debacle in 1998 that involved a consortium bailout organized by the FED is a classic case illustrating the difficulty in modeling and managing basis risks. In that incident, the basis spread between long off-the-run bonds positions versus short, more liquid benchmark bonds blew apart. It is often difficult to model such basis risk because of the absence of reliable data.

The more risk factors a bank includes into a VaR model, the more imprecise the model becomes. This is due to the curse of dimensionality. A typical VaR system in a bank uses 500 points (or 2 years' worth of data) but runs more than 50,000 positions (i.e., dimensions). It is mathematically impossible a simulate this series without reducing the dimensions; a covariance matrix in 50,000 dimensions has about 1.25bn coefficients, but we have only 25m data points to calculate them, so the system is greatly underdetermined and the resulting covariance matrix has much degenerated dimensions with zero variance.

Pro-cyclical Risks

VaR, or volatility in general, is low during a market rally and abruptly high during a market crash. This phenomenon is called the leverage effect, and can easily be seen by plotting the S&P

500 index against the VIX index (which is the volatility of the U.S. equity market implied by the option market; Figure 6).

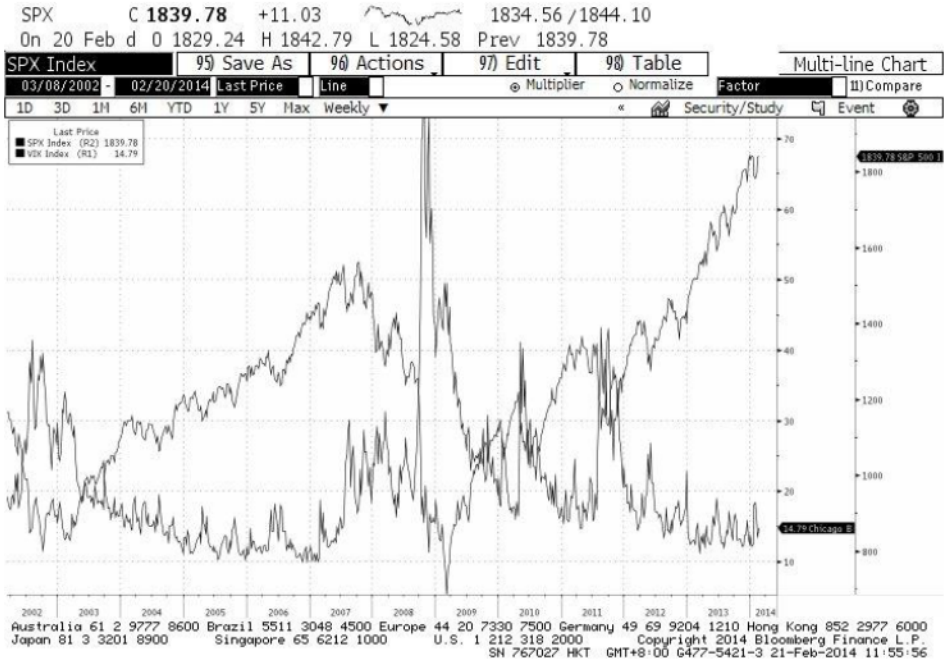

Figure 6. Bloomberg screenshot of S&P 500 index versus VIX index

The problem is that Basel regulation requires that a bank's capital be based on VaR. Hence, when the market is in a boom phase, VaR and capital requirements become benign (i.e., low), encouraging balance sheet expansion of banks and the purchase of riskier assets. There is a degree of herding at work because most banks want to maximize their use of capital to compete with other banks. Conversely, during the bust phase, the market falls; VaR and capital requirements increase abruptly, causing banks to liquidate risky positions to manage required, regulatory capital. In effect, the regulatory rules inadvertently cause banks to buy rallies and sell breaks. This herding behavior pressures the markets and amplifies the boom-bust cycle. This undesirable effect is called procyclical risk and has been getting much attention since the crisis. Basel III responded somewhat to those concerns, for example, by introducing countercyclical capital buffers and a stressed VaR figure whose risk parameters are frozen at stress-scenario values.

Extremistan and Black Swans

Nassim Taleb, author of the New York Times best seller The Black Swan (2007) and a strong critic of VaR models, made Extremistan popular. Extremistan refers to a class of probability structures that are immeasurable at the extreme tails of the distribution. Such distributions exhibit characteristic fat tails and rare events that make up the tails are atypical, meaning past occurrences offer no guidance on the magnitude of future occurrences and are hence not amenable to statistics. Examples of extremistan phenomena include destruction from flu pandemics, world wars, Ponzi schemes, the wealth creation of the superrich, technological

breakthroughs, etc. Taleb termed such unpredictable events Black Swans. In the social sphere, Black Swans are often caused by scalability and positive feedback of thinking participants. In contemporary electronic trading, banking interconnectedness, opaque derivatives markets, and fiat money, the unconstrained use of leverage can create the dire possibility of massive financial losses and crisis contagion, as in 2008.

If one believes that such financial crises are extremistan events, then use of VaR models to predict low probability events is questionable. Due to finite sampling, the risk modeler can always estimate VaR to the stated degree of confidence level, but in the presence of extremistan, such a number is misleading and dangerous because it gives the user a false sense of security; the true risk could be much more severe. Taleb's idea of extremistan provides a wise warning of the misuse of models and statistics.

Dangerous Nonlinearity

A common mistake in risk management is to assume that extreme losses that threaten a bank's survival occur only when the market is under stress. However, a bank is hurt by the payoff $g(x)$ of its positions instead of market movement x. Since the payoff $g(\cdot)$ can be nonlinear or convex, the losses can be large even though movement in the underlying market (x) is regular.

Payoff of bonds, for example, is convex, but this convexity is mild except for extremely long-dated and high-coupon bonds. More convex are the payoffs of option products, especially exotic options with discontinuous payoffs. An option causes the P&L distribution to be skewed to one side and fatter than normal (leptokurtic). Unfortunately, skewness and kurtosis are not measured well statistically, even for models such as historical simulation VaR, because of the scarcity of sampled data and because these moments are prone to data outliers. Hence, even if the market is in a regular (non-crisis) state, nonlinear risk is something a risk manager should monitor closely using a variety of other tools such as stress scenario matrices and gamma (or convexity) limits.

More dangerous is nonlinearity that arises because of market impact. When a market is under stress or is illiquid, selling 100 blocks of securities at once creates a much larger price impact on the market (and P&L swing) than selling 10 blocks of securities ten times over a period. Such dangerous nonlinearity occurs during crises (even for linear products), and is missed by VaR models (see the Liquidity Effects section above).

Inconsistency in Estimation across Banks

The inability to fit the distribution accurately at the extreme loss tail of the sample, and the data challenges discussed above, mean banks often obtain different VaR results for identical portfolios. In practice, banks build their own internal models (i.e., VaR) with different methodologies, parameters, and calibrations such as observation window length, weighting schemes, risk factor mappings, pricing models, parameter calibrations, system implementations, data sources, etc.

A study by Basel in December 2013 suggests banks compute varying VaR results for the same test portfolios. The analysis covered 17 banks in 6 jurisdictions. Each bank was given 35 test portfolios covering a range of asset classes. The VaR from each bank was compiled, normalized to 100%, and plotted as a point in Figure 7.

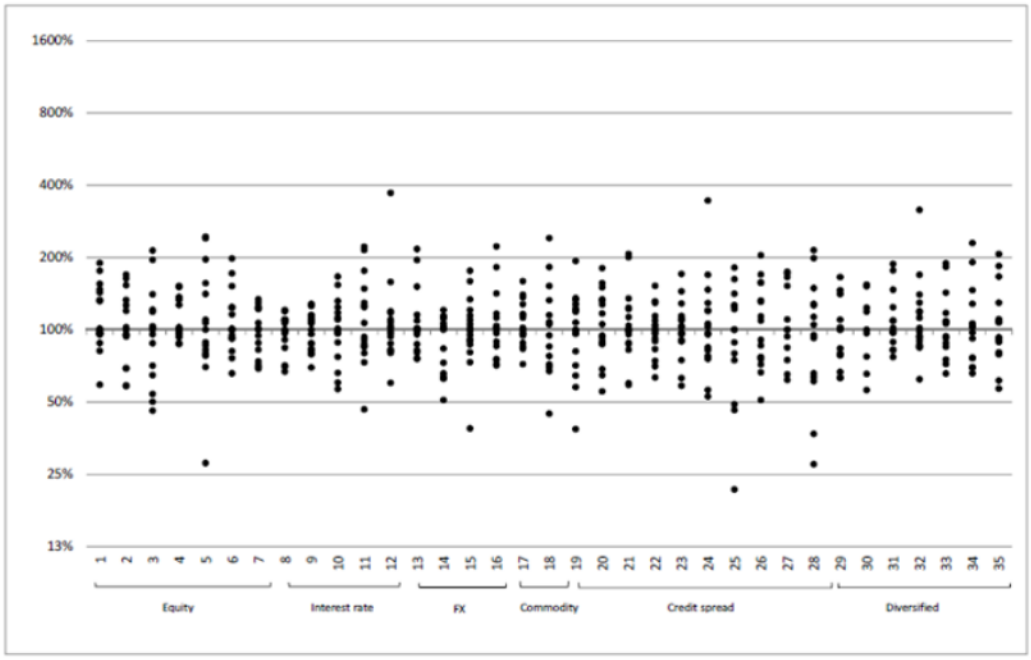

Figure 7. Dispersion of normalized VaR results for all portfolios
Source: BCBS Dec 2013, "Regulatory consistency assessment program: Second report for RWA for market risk"

The figure revealed that the VaR across banks ranged from half to double the median value. Such inconsistency in reported VaR is worrying because it questions the efficacy of the metric, and regulators cannot apply appropriate and consistent capital charges to safeguard banks.

Correlation Framework is Fragile

Linear (or Pearson's) correlation is a fundamental concept in Markowitz portfolio theory, and is used commonly in risk management to sum risks, but correlation is a minefield for the unaware. The key is to remember that estimation of correlation involves drawing the best straight line across the bivariate scatter plot (i.e., correlation measures the clustering around a straight line). The drawing is good only if the data cloud is elliptical. Otherwise, the estimation is biased, and correlation sometimes behaves unintuitively. To describe the relationship between variables fully, one would normally need to use copulas (see the section on Copula), but due to the paucity of data, it is impossible to determine which copula is correct. Even where copulas are used, the industry typically uses the simplest copula (the Gaussian copula) in selected applications such as to model default risks.

Systemic Risks and Breakdown of Model during Crisis

One drawback of the VaR model is that by design, it does not capture the dynamics of feedback loops in prices. Thus, during a crisis, VaR models fail. The topic of systemic risk gained much

attention after the 2008 crisis. It helps to understand the mechanism of a market crisis. The financial markets are seldom in a state of equilibrium, as suggested by classical economic theory. As early as 1987, George Soros introduced the idea of reflexivity in his book The Alchemy of Finance, which suggests market prices (because of feedback loops) are always in an unstable state of overshooting, and large and sudden corrective moves constitute a crisis. Due to structural features of the market, participants often build up crowded trades and exhibit herding mentality. Examples include the systemic involvement of global banks in credit derivatives in the years leading up to the credit crisis in 2008, and the collective speculation in Yen-carry trades by investment banks/hedge funds in 1998. In the latter case, financial institutions purchased high-yielding currencies (such as AUD dollar) funded by near-zero Yen interest rates. Both ended badly in massive destruction of wealth and fallouts as participants rushed to exit trades when the market reversed.

Such panic selling gives rise to two effects: a) contagion, the spillover effect among large players or institutions, and b) positive feedback loops within the market, panic selling that depresses prices, which then leads to more panic selling. Contagion is worsened by the fact that financial institutions are often counterparties to each other, and hence their balance sheets interconnect. A contagion causes institutions to tighten credit lines and margin requirements to other institutions that are exposed heavily to a falling market. An institution might even sell into the same market to hedge itself, or cut losses to meet margin calls. These actions give rise to feedback loops, which break the stationary assumption of VaR models and create serial correlations in data series. This string of losses, or drawdowns, hurts the banks most, and is not contained well by VaR capital. Such a regime shift typically occurs quickly, which makes it difficult for VaR models to capture it. As an extreme example, the flash crash on Thursday May 6, 2010, in which the Dow Jones index plunged about 1000 points (roughly 9%) only to recover within minutes, was an incident caused by high-frequency trading.

This sudden appearance of serial correlation gives rise to stochastic (or changing) volatility; it is mathematically shown that stochastic volatility causes fat-tail distributions and the phenomenon of volatility smiles reflected in option markets. Unfortunately, standard VaR models use a slow observation window; the VaR metric is late in picking up the new regime. At the portfolio level, a crisis regime shift is typically accompanied by a breakdown of correlation (i.e., relationships in the past no longer apply). Markets tend to fall together and become highly correlated during a contagion.

Over the years, academia introduced conditional models (such as the GARCH model, 1986), which are more adept and responsive to changing volatilities. However, such models are difficult to implement for large portfolios typical at banks, and are too complex to maintain and communicate.

Regulatory Impetus on Model Development

The BCBS is instrumental in guiding development of a framework for banking regulation and models for capital calculations. A core element of supervision is the idea of capital adequacy—a

bank must hold sufficient capital for the business in which it is engaged to buffer itself against unforeseen, extreme losses.

The first VaR model was introduced into regulatory capital requirements in the 1996 Market Risk Amendment. The Basel II model extended the VaR concept to credit risk and operational risk, with a different time horizon (1 year) and confidence level (99.9%). The trading book approach was not changed substantially from the 1996 Amendment.

The idea of Basel II was that sophisticated banks could use their own internal models (obeying constraints) for risk calculations on the credit side, as was already the case on the market risk side. Basel II thus was meant to provide economic incentives for banks to research and develop better models to measure risks, and obtain more efficient ways to utilize capital.

Under Basel II, market risk (MR) is divided into general market risk (GMR) and specific risk (SR), based on a 99% confidence level, 10-day risk horizon. For example, a bond issued by IBM would contain the interest rate risk (GMR) and risk coming from movement of the credit spreads of IBM, the issuer (SR).

Basel 2.5 was the first BCBS policy response to the global financial crisis. Released in July 2009, it was the precursor to Basel III, and the Basel 2.5 changes were eventually subsumed into Basel III. The new rules pertaining to market risk capital calculations and models are:

1. Incremental Risk Charge (IRC): applicable to any bank that has internal models approval for specific risks. Basel provided high-level guidance on how this should be modeled:
    a. Based on 99.9% confidence level, 1-year risk horizon, must have basic framework similar to the IRB model. Hence, the Vasicek model is implied.
    b. Introduced multiple liquidity horizons for products of different liquidities. The most liquid products have a liquidity horizon of at least 3 months.
    c. Assumes a constant level of risks (i.e., positions are assumed to be rebalanced at each liquidity horizon to ensure constant VaR), and for many, steps up the risk horizon. Many banks chose to implement a multistep, one-factor Gaussian copula model.
    d. Rating migration, default risk, optionality and their cross correlations must be modeled.
    e. Concentration risks must be reflected by having a granular classification/differentiation of positions.
2. Credit securitization products (e.g., tranches, CDOs, and credit correlation instruments) are excluded from the IRC and modeled instead using a Comprehensive Risk Model (CRM). The CRM, in theory, captures the myriad of risk factors pertinent to these toxic products. Since 2008, most banks chose to sell such legacy businesses instead of facing the arduous task of designing/maintaining the capital-punitive CRM.
3. Stressed VaR (SVaR): an additional capital requirement to cover the weakness that the standard VaR underestimates risks during a crisis, and as a first buffer against procyclical risk. SVaR is just the 99%/10-day VaR calculated using a 1-year (fixed) observation period of high stress.

4. Basel III was released in December 2010. It contained a number of important changes, for example, with respect to capital buffers and counterparty requirements, but it did not introduce changes to the market risk framework other than those that had been introduced by Basel 2.5.

An ongoing process (as of 2014) with the BCBS is a fundamental review of the trading book. Key proposals include:

1. Replacement of VaR with 97.5% expected shortfall, which empirically is close to the 99% VaR on real-world distributions (ES). The latter captures the shape of the tail of the loss distribution, and is a coherent risk measure.
2. Use of stressed calibration for risk models for the purpose of capital. This recognizes the BCBS objective of reduction of cyclicality of risk measures.
3. Internal models are approved at the more granular (desk) level rather than at the bankwide level, and are conditional on good backtesting and P&L attribution.
4. Additional charge to cover non-modelable risks such as those arising from data issues.
5. Impose constraints on diversification benefits in internal models. This recognizes the breakdown of correlations during crises.
6. Comprehensive incorporation of liquidity risk into ES. This uses the liquidity horizon construct like IRC, but horizons are prescribed by BCBS based on asset classes.

Advanced VAR Models - Univariate

Backtesting

From the previous section, it is obvious that banks can choose from many VaR methodologies and parameters settings such as observation periods, weighting schemes, etc. Accuracy of the VaR model depends critically on the quality of the data collected by the bank and the IT system implementation. In practice, many factors lead to inaccurate, imprecise, or inconsistent VaR numbers. Since VaR is used for computation of regulatory capital, there is an important implication on banks' safety buffers and costs of capital. Thus, it is essential to perform internal testing, validation, and review of VaR models

To ensure VaR models are fit for purpose, regulators require that banks backtest their VaR models regularly. In the case of market risk, two variations of backtesting serve several purposes. These involve comparing *ex ante* VaR estimates with *ex post* values of a) real P&L in the applicable periods, and b) hypothetical P&L, assuming constant positions for the applicable periods. For example, the VaR computed for positions at time T is compared with P&L of the same position at T+1. This is done for all T's in the sample period. In the case of a), the position changes across the sample, and in the case of b), the latest position is used and assumed to hold every day in the past sample. The first approach is required of banks under Basel rules.

The correctness of recorded P&L over various cycles is what risk systems are ultimately designed to achieve. Hence, comparing VaR estimates to real P&L (based on the bank's

changing positions) is a useful part of any backtesting. A problem with such a test is that there can be many reasons real P&Ls exceed risk estimates more frequently than expected theoretically. This might occur not only because of weaknesses in the VaR model, but due to contamination of the past P&L by factors such as intraday P&L and reserve adjustments that were not removed properly from the P&L for the purpose of backtesting, an old IT bug, etc.

When backtesting reveals weaknesses in the VaR system, it is important to diagnose the source of the problem as per the system and positions today. This is where the second form of backtesting is a useful complement. VaR assumes that the portfolio position remains static over the risk horizon, an idea consistent with the use of hypothetical P&L during backtesting. Put another way, we should be more interested in the correctness of the VaR model for today's positions than the correctness on the VaR model in the distance past.

Exception Measurement and Basel Rules

The primary consideration during backtesting is whether the P&L series exceeds the corresponding *ex ante* VaR estimate within the predicted frequency. This can and should be repeated for VaR with various confidence levels. In the simplest form, backtesting counts the number of times the real portfolio P&L breached the VaR estimate, and compares that number to the confidence level. For example, if a confidence level were 99% for a daily VaR, we would expect, on average, a 1% chance of a daily P&L breach of the VaR. The breach is also called exception or exceedance. We observe the number of exceptions that occurred in the past. If the number of exceptions is approximately 1% of the total number of days in the observation period, we conclude that the VaR model is adequate. Otherwise, the VaR model is aggressive if the percentage of exception is evidently greater than 1% or conservative if it is evidently smaller than 1%.

Current regulation requires banks to use the traffic light approach (TLA) as a standard method to backtest their VaR systems. This approach counts N, the number of 1% VaR violations (i.e., exceptions) in the previous 250 trading days. Regulators rely on this approach to justify the soundness of a bank's internal VaR model, and to adjust the multiplier for capital charges accordingly. If a bank qualifies for Basel's internal models, its market risk capital requirement is:

$$MRC_t = \max\left(\mathrm{VaR}_t^{99\%},\ K_t \frac{1}{60} \sum_{i=0}^{59} \mathrm{VaR}_{t-i}^{99\%} \right) + SRC_t \qquad (50)$$

which involves taking the larger of the most recent VaR and the 60-day average VaR. The multiplier, K_t, is determined by classifying the number N into three categories:

$$K_t = \begin{cases} 3.0 & \text{if } N \le 4 \quad \text{Green} \\ 3.0 + 0.2 \times (N - 4.0) & \text{if } 5 \le N \le 9 \quad \text{Yellow} \\ 4.0 & \text{if } N \ge 10 \quad \text{Red} \end{cases} \qquad (51)$$

If the VaR is an unbiased estimate of the quantile, we expect 2.5 violations in 250 days, statistically by definition. The multiplier remains at its lowest level of 3 if the VaR exceptions are

fewer than 4. If the VaR is violated more frequently, a bank is penalized more by higher capital charges through a bigger multiplier. In the red zone, the VaR model is deemed inaccurate (i.e., systematically understates risk), and immediate corrections are required to improve the VaR system. In contrast, too few violations (<2) implies that the bank's model is overly conservative; it overstates risk systemically.

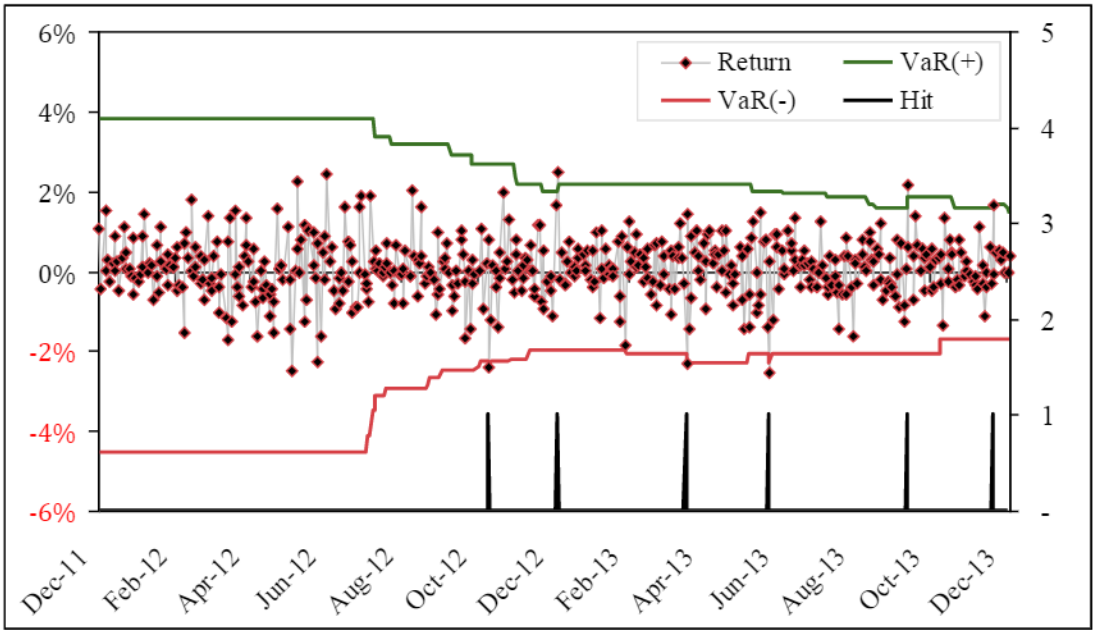

Figure 8. Backtesting using TLA for a simple equity portfolio

As an illustration, the spreadsheet [VaR_Backtesting.xls] shows TLA backtesting using a portfolio comprised of an equity index asset (SPX index). The VaR calculation and backtesting windows are both 250 days. With 3-year P&L (expressed in percent return) data from 2011 to 2013, we have approximately 500 VaR estimates. Results appear in Figure 8. There are in total six VaR violations (including positive and negative violations) plotted as hits. Overall, TLA shows that the VaR model falls into the green zone most of the time, but turns yellow briefly in late 2013.

Frequency Based Backtests

To determine whether a VaR model is specified well, it suffices to test that it satisfies the hypothesis of unconditional coverage and independence. The hypothesis of unconditional coverage states that the expected frequency of observed exceptions does not differ from the assumed probability p (e.g., $p = 1 - \alpha = 0.01$ for a 99% VaR). The hypothesis of independence means the exceptions should occur independently across time; the exception sequence should be spread evenly, not clustered. Otherwise, serial dependence might arise in the P&L (i.e., in the market) that is not capture by the VaR model, and hence the VaR model is misspecified.

The TLA is simple but is merely an empirical formula provided by the regulator, and applies only to 99% VaR. To improve this counting test, a few approaches have been developed based on statistical hypothesis testing. One is called the Kupiec test, proposed by Kupiec (1995) [6], and is a cousin of TLA, which examines the null hypothesis of unconditional coverage rate p^* (e.g., $\mathcal{H}_0: p^* = 0.01$). The Kupiec test is analogous to TLA since both focus on testing the proportion of exceptions (i.e., coverage rate). The difference is that TLA is a one-sided test, with the alternative hypothesis $\mathcal{H}_1: p^* > 0.01$, and the Kupiec test is a two-sided test with alternative hypothesis $\mathcal{H}_1: p^* \neq 0.01$. This means the Kupiec test fails if the VaR model is either too aggressive or too conservative, whereas the TLA test fails only if the VaR model is too aggressive.

The idea of the Kupiec test is to test whether the observed violation frequency is consistent with the frequency predicted by the model. For example, consider a daily series of *ex post* portfolio return r_t and a corresponding series of *ex ante* VaR forecast VaR_t^{1-p}, with coverage rate p (e.g., $\mathrm{VaR}_t^{99\%}$, with $p = 0.01$). This yields a probability $\mathbb{P}\left[r_t < \mathrm{VaR}_t^{1-p}\right] = p^*$ at a one-sided $1 - p$ confidence level. Under the null hypothesis, $\mathcal{H}_0: p^* = p$, the number of violations, x, follows a binomial distribution. Given a total of n P&L observations, the probability of x violations is calculated as:

$$\mathbb{P}[x|n,p] = \binom{n}{x} p^x (1-p)^{n-x} \tag{52}$$

Rather than calculate probabilities from the discrete binomial distribution directly, Kupiec proposes a proportion-of-failures (POF) coverage test, a variant of likelihood ratio test, relying on the statistic $-2\ln\Lambda$, where Λ is the likelihood ratio constructed as:

$$\Lambda = \frac{\text{likelihood for null model}}{\text{likelihood for alternative model}} = \frac{p^x(1-p)^{n-x}}{\left(\frac{x}{n}\right)^x \left(1-\frac{x}{n}\right)^{n-x}} \tag{53}$$

By assuming $\mathcal{H}_0$, the statistic $-2\ln\Lambda$ asymptotically follows a centered chi-squared distribution $\chi^2(1,0)$, with 1 degree of freedom as n increases. For a given significance level, α (e.g. 95%), we can construct a non-rejection interval $[l,\ u]$ such that $\mathbb{P}[x < l] \leq \frac{\alpha}{2}$ and $\mathbb{P}[x > u] \leq \frac{\alpha}{2}$, where the interval bounds l and u are the two solutions of x, making the $-2\ln\Lambda$ equal to the α quantile of the $\chi^2(1,0)$.

Both the Kupiec and TLA methods focus only on testing frequency, ignoring potential serial dependence in VaR exceptions. The latter weakness can be detected by a duration-based approach developed by Christoffersen and Pelletier (2004) [7]. The idea is that if a 1-day VaR model is specified well for a frequency of p, each day, the unconditional expected duration should always be $1/p$ days. This defines the null hypothesis that the duration has no memory on the length it has already undergone, and it always has an expected mean of $1/p$ days. Since exponential distribution is the only memory-free continuous random distribution, under the null hypothesis, the duration must follow an exponential distribution. To establish a statistical

test for duration independence, an alternative hypothesis must be specified that allows for duration dependence. Hence, the test must be conducted on a distribution that nests the exponential distribution as a special case (e.g., a Weibull distribution). A likelihood ratio test is then conducted to assess whether the special case holds. If it holds, the null hypothesis is passed and the VaR model has the duration independence.

Many other statistical tests based on backtesting techniques have been developed in recent years. A good reference for these advanced tests is Wehn (2008) [8].

Distributional Equality Based Backtests

All tests mentioned above check for frequency of VaR exceptions, which are counts of rare events in general, and are certainly less informative than the entire P&L distribution. To use the entire P&L data fully, new approaches have been developed to backtest the whole distribution. Such tests first transform realized P&L in terms of forecast probability CDF values and then use the transformed data to test the equality of the probability distributions. Since tests are performed on the entire distributions, they have more diagnostic power than frequency-based approaches. Full distribution backtesting exploits the *Rosenblatt transformation*, expressed as:

$$u_t = F_t(x_t) \tag{54}$$

where x_t is the realized P&L at time t and $F_t(\cdot)$ the CDF of the forecasting P&L distribution (i.e., the P&L distribution used to derive the VaR for day t; with hsVaR, it is the 250-day P&L vector up to day $t-1$). Since true density function $F_t(\cdot)$ is unknown and estimated by a VaR model *ex ante* (i.e., using data from one day earlier), we use the estimated $\hat{F}_t(\cdot)$ to perform the transformation, which gives a series of $\hat{u}_t$. For parametric VaR models, the functional form of $\hat{F}_t(\cdot)$ is known and can be used analytically. For nonparametric VaR models (e.g., hsVaR), the $\hat{F}_t(\cdot)$ is estimated as an empirical CDF, which is simply:

$$\hat{F}_t(x) = \frac{\text{number of elements in the P\&}L\ vector \leq x}{\text{total number of elements in the P\&}L\ vector + 1} \tag{55}$$

We transform the series of x_t to $\hat{u}_t$ for the past n days (say, $n = 500$). For the VaR model to be well behaved, the null hypothesis in the series of $\hat{u}_t$ must be distributed uniformly between zero and 1. The uniformity in the transformed $\hat{u}_t$ series can be tested using Kolmogorov-Smirnov statistics, defined as:

$$\begin{aligned} D_n &= \sqrt{n} \max_{t=1,\cdots,n} |\hat{u}_t - u_t| \\ &= \sqrt{n} \max_{t=1,\cdots,n} \left|\hat{F}_t(x_t) - F_t(x_t)\right| \end{aligned} \tag{56}$$

The steps to construct the statistics are:

1. Define a backtesting window of n-days such that $t = 1, \cdots, n$. For example, let $n = 500$.
2. For each day t, we transform the realized P&L x_t to $\hat{u}_t$ using the estimated density function $\hat{F}_t(\cdot)$ of the P&L distribution (the empirical CDF in hsVaR or parametric CDF in pVaR using *ex ante* [one day behind] dataset).
3. Sort the $\hat{u}_t$ in ascending order to obtain the empirical uniform CDF. This gives a vector $\hat{u}_i$ for $i = 1, \cdots, n.$, with $\hat{u}_1$being the smallest number closest to 0.0. For comparison, we define another vector u_i for $i = 1, \cdots, n$, the true uniform CDF such that:

$$u_i = \frac{i-1}{n} \tag{57}$$

   The Kolmogorov-Smirnov test is performed to examine the uniformity of the $\hat{u}_i$ series by benchmarking against the u_i series.
4. The Kolmogorov-Smirnov statistic is defined as the largest absolute difference between the vector $\hat{u}_i$ and u_i multiplied by the square root of n. If the statistic is smaller than the critical value at a given confidence level, the null hypothesis is passed (or not rejected), and the VaR model is considered to be well behaved.

Other statistics test this such as the Anderson-Darling test, but this is outside the scope of this book.

The spreadsheet [VaR_Backtesting.xls], found on the handbook resources webpage, illustrates this advanced backtest using a simple example. Data used in the example are 3-year log-returns of SPX index from 2011 to 2013. For simplicity, we assume a rolling observation period of 250-day returns (ending at date *T*) offers a good estimate of the P&L distribution (at date *T*), and use it to perform the out-of-sample Rosenblatt transformation of the P&L at date $(T + 1)$. After deriving the $\hat{u}_t$ series (roughly 500 points), we plot a histogram (shown in Figure 9) to visualize its uniformity. It clearly shows a slightly lower than uniform frequency at both loss and gain tails. This indicates that the P&L distribution generally forecasts a conservative (or overstated) VaR. If the tails of the frequency distribution are higher than the body, any VaR forecast (that comes from that distribution) is likely underestimated. An unbiased VaR comes from a P&L distribution in which its Rosenblatt-transformed distribution is reasonably uniform. Such underestimation or overestimation of VaR is likely caused by regime shifts in the dataset and stochastic volatility of the markets, in general. We also draw the empirical uniform CDF versus the theoretical uniform CDF as a QQ plot (Figure 10). The discrepancy in the two tails suggests the empirical tails are fatter than in theory, and hence, computed risk measures such as VaR are overestimated.

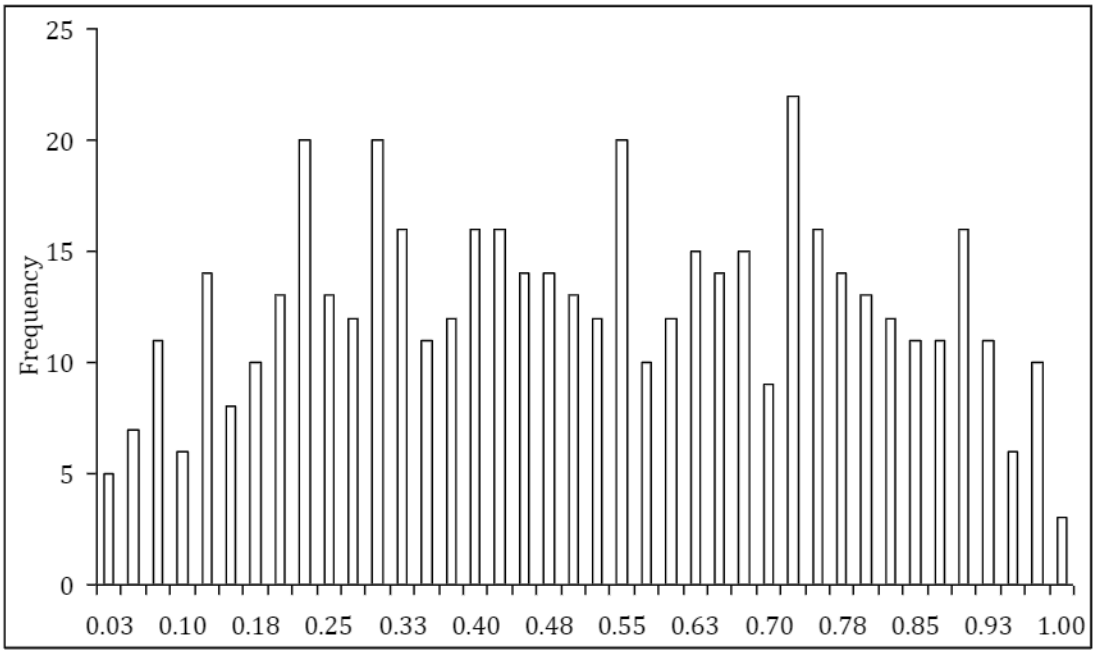

Figure 9. Out-of-sample Rosenblatt transformation

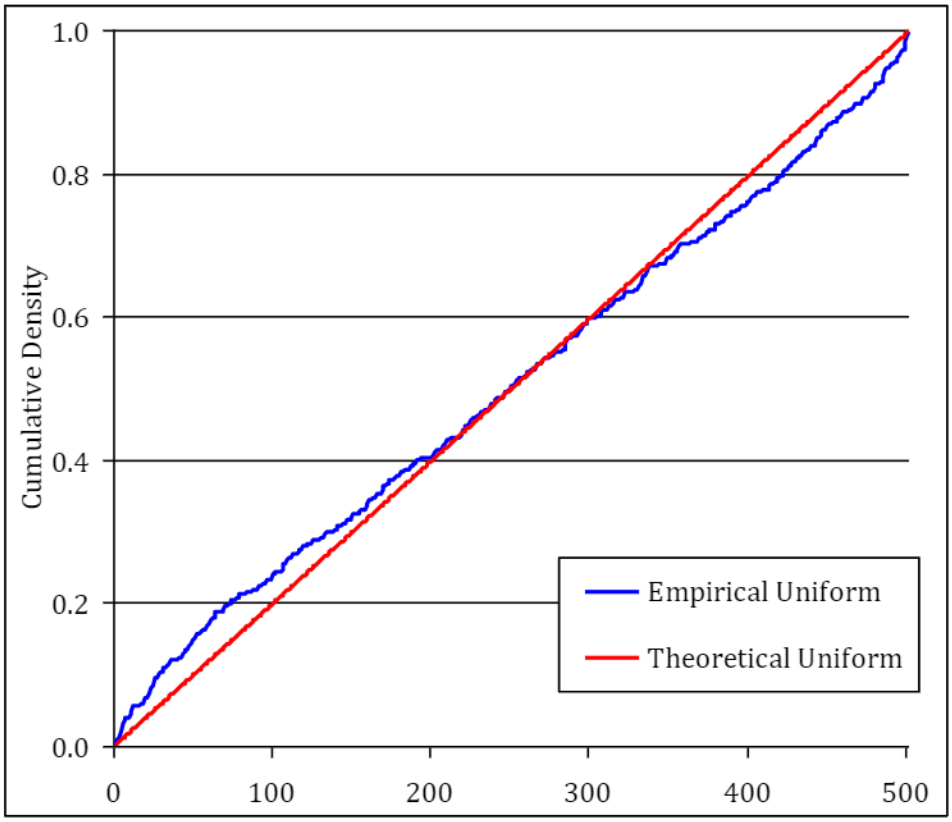

Figure 10. Empirical uniform CDF (Rosenblatt transformation) versus theoretical uniform CDF

Extreme Value Theory

Extreme Value Theory (EVT) is a field of applied statistics traditionally used in the insurance industry and now increasingly applied to operational risk. It provides estimation techniques to forecast extreme events with low probability of occurrence. This section provides a brief introduction of what EVT can do in the area of VaR. We end with a simple Microsoft Excel example of EVT VaR. EVT does not attempt to model the tail process; hence, one can think of it as a tool to fit the tail distribution in a statistically correct manner. The reasoning is that if the rare event lies outside the range of available observation, it seems essential to rely on good fundamental methodology. The reason EVT gained acceptance in risk management is because return distributions in financial markets are severely fat-tailed during times of crises.

Classical EVT

The fundamental model of EVT describes the behavior of the *maxima* of a distribution (the theory discussed here also applies to the minima because properties of the minima can be obtained from those of the maxima by a simple sign change). Consider a collection of n observed daily returns $\{r_i\}$ for $i = 1, \cdots, n$, where we ignore the sign of returns and express losses as positive numbers. With VaR, we are interested in modeling extreme losses, l_n. So, consider the worst-case loss such that $l_n = \max\{r_i\}$. We denote the cumulative distribution function (CDF) of a random variable x as $F(x)$. Assuming the returns are i.i.d., the CDF of l_n, (i.e., $F_n(x)$) can be derived easily:

$$\begin{aligned} F_n(x) &= \mathbb{P}[l_n < x] \\ &= \mathbb{P}[r_1 \leq x, \cdots, r_n \leq x] \\ &= \prod_{i=1}^{n} \mathbb{P}[r_i \leq x] \\ &= \prod_{i=1}^{n} F(x) \\ &= F(x)^n \end{aligned} \tag{58}$$

However, this CDF becomes degenerated as n increases to infinity; it becomes a Heaviside step function translated to a value u, that is:

$$F_n(x) \xrightarrow[n\to\infty]{} \begin{cases} 0 & \text{if } x < u \\ 1 & \text{if } x \geq u \end{cases} \tag{59}$$

Since the degenerated CDF is useless in practice, EVT identifies an asymptotic distribution of the normalized maximum:

$$l_n^* = \frac{l_n - \mu_n}{\sigma_n} \tag{60}$$