Modelling Economic Capital under Basel II Pillar 2

Alexander J. McNeil
Heriot-Watt University
Edinburgh

9–10 June 2008, Concentric Finance Master Class, Milan, Italy

http://www.pupress.princeton/edu/titles/8056.html
http://www.ma.hw.ac.uk/~mcneil/

©2008 (McNeil)
Overview

I. Assembling the Components

II. Building Integrated Models for Economic Capital
Part I: Assembling the Components

A. Introduction to Risk and Economic Capital

B. Basic Elements of Risk Modelling

C. Market Risk Models

D. Credit Risk Models

E. Operational Risk Models
A. Introduction to Risk and Economic Capital

1. Risk and Why We Manage it

2. Basel II, Solvency II and Economic Capital

3. Important Themes
A1. Risk and Why We Manage it

What is Risk?

“hazard, a chance of bad consequences, loss or exposure to mischance” [OED]

“any event or action that may adversely affect an organization’s ability to achieve its objectives and execute its strategies”

“the quantifiable likelihood of loss or less-than-expected returns”
### Top 10 Trading Losses of Recent Times

<table>
<thead>
<tr>
<th>Year</th>
<th>Company</th>
<th>Country</th>
<th>Source</th>
<th>Loss</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Société Générale</td>
<td>France</td>
<td>index futures</td>
<td>€4.9</td>
<td>$7.10</td>
</tr>
<tr>
<td>2006</td>
<td>Amaranth Advisors</td>
<td>USA</td>
<td>gas futures</td>
<td>$6.5</td>
<td>$6.70</td>
</tr>
<tr>
<td>1998</td>
<td>LTCM</td>
<td>USA</td>
<td>IR derivatives</td>
<td>$4.6</td>
<td>$5.85</td>
</tr>
<tr>
<td>1996</td>
<td>Sumitomo Corp.</td>
<td>Japan</td>
<td>copper futures</td>
<td>¥285</td>
<td>$3.44</td>
</tr>
<tr>
<td>1994</td>
<td>Orange County</td>
<td>USA</td>
<td>IR derivatives</td>
<td>$1.7</td>
<td>$2.38</td>
</tr>
<tr>
<td>2000</td>
<td>BAWAG</td>
<td>Austria</td>
<td>FX trading</td>
<td>€1.4</td>
<td>$1.97</td>
</tr>
<tr>
<td>1993</td>
<td>Metallgesellschaft</td>
<td>Germany</td>
<td>oil futures</td>
<td>DM2.63</td>
<td>$1.96</td>
</tr>
<tr>
<td>1995</td>
<td>Barings Bank</td>
<td>UK</td>
<td>Nikkei futures</td>
<td>£0.827</td>
<td>$1.80</td>
</tr>
<tr>
<td>1995</td>
<td>Daiwa Bank</td>
<td>Japan</td>
<td>bonds</td>
<td>¥103</td>
<td>$1.50</td>
</tr>
<tr>
<td>2007</td>
<td>WestLB</td>
<td>Germany</td>
<td>share dealing</td>
<td>€0.60</td>
<td>$0.82</td>
</tr>
</tbody>
</table>


All figures are in billions. The **adjusted** loss is based on the USD inflation-adjusted amount.
Trading Losses: Market and Operational Risk

- **Société Générale.** A rogue trader, Jerome Kerviel, is alleged to have used index futures to make huge directional bets on European equities. He managed to evade risk management limits by his knowledge of back office systems. (echoes of Barings 1995)

- **Amaranth.** A star trader, Brian Hunter, placed huge, highly leveraged (8:1) spread trades in the natural gas futures market. (echoes of LTCM 1998)

Common to both cases: sophisticated firms; use of derivatives to take bets; market movement in opposite direction to that of bet; loss technically due to market risk; failure of risk management system a form of operational risk.
Risk Definitions

Market Risk. The risk of a change in the value of a financial position due to changes in the market value of the underlying components on which that position depends, such as stock and bond prices, exchange rates, commodity prices, etc.

Operational Risk. The risk of losses resulting from inadequate or failed internal processes, people and systems, or external events.

Credit Risk. The risk that the value of a portfolio changes due to unexpected changes in the credit quality of issuers or trading partners. This subsumes both losses due to defaults and losses due to downgradings of obligors in a rating system.
# Top 10 Asset Write-Downs & Credit Losses
(Since the Credit Crunch Began)

<table>
<thead>
<tr>
<th>Company</th>
<th>Country</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 UBS</td>
<td>Switzerland</td>
<td>$37.4bn</td>
</tr>
<tr>
<td>2 Merrill Lynch</td>
<td>USA</td>
<td>$22.0bn</td>
</tr>
<tr>
<td>3 Citibank</td>
<td>USA</td>
<td>$21.1bn</td>
</tr>
<tr>
<td>4 HSBC</td>
<td>UK</td>
<td>$17.2bn</td>
</tr>
<tr>
<td>5 Morgan Stanley</td>
<td>USA</td>
<td>$9.4bn</td>
</tr>
<tr>
<td>6 Deutsche Bank</td>
<td>Germany</td>
<td>$7.1bn</td>
</tr>
<tr>
<td>7 Bank of America</td>
<td>USA</td>
<td>$5.3bn</td>
</tr>
<tr>
<td>8 Bear Stearns</td>
<td>USA</td>
<td>$3.2bn</td>
</tr>
<tr>
<td>9 JPMorgan Chase</td>
<td>USA</td>
<td>$3.2bn</td>
</tr>
<tr>
<td>10 BayernLB</td>
<td>Germany</td>
<td>$3.2bn</td>
</tr>
</tbody>
</table>

Source: BBC website, 01/04/08

*write down* = a paper loss incurred by reducing the book value of an asset because it is overvalued compared to the market value
Think of a Number …

“It is a worry, though, that Merrill can justify a write-down of $4.5bn one week and $7.9bn just three weeks later. The sense that valuation is still a matter of **pick a number and divide by the chief trader’s golf handicap**, more than anything else, explains why the super-SIV proposed by Citigroup and others failed to reassure the market.”

Source: Financial Times
A Causes B Causes C

MBS = mortgage-backed security: an asset-backed security whose cash flows are backed by principal and payments of a set of mortgages

CDO = collateralized debt obligation: asset-backed security structure constructed from a portfolio of debt instruments which is divided into tranches of differing credit quality and sold to investors, usually investment banks and hedge funds

- Starting in late 2006 high interest rates and falling house prices in US ⇒ large scale default of sub-prime mortgages

- Sub-prime defaults ⇒ collapse of MBS market

- Collapse of MBS market ⇒ collapse in CDO market
Causes D Causes E

- Collapse in market for securitized debt $\Rightarrow$ write-offs and general nervousness in banks

- Nervousness about bad debt in banks $\Rightarrow$ drying up of interbank lending market and increase in cost of debt financing

- Increase in cost of debt financing $\Rightarrow$ liquidity problems at banks

- Liquidity problems $\Rightarrow$ bank runs, Northern Rock (09.07), Bear Stearns (03.08)

- Bank runs $\Rightarrow$ general financial malaise including falling stock markets led by bank shares

- General financial malaise $\Rightarrow$ ... recession etc.(?) ...
A2. Basel II, Solvency II and Economic Capital


1996 Amendment to Basel I. Internal VaR models for market risk in larger banks.

Basel II, 2001-08. Second Basel Accord, focussing on credit risk but also putting operational risk on agenda. Banks may opt for a more advanced, so-called internal-ratings-based approach to credit.
Three Pillars of Basel II

1 Pillar 1: minimal capital requirements (risk measurement)

2 Pillar 2: supervisory review of capital adequacy

3 Pillar 3: public disclosure
Solvency II

What is it? A fundamental review of the regulatory capital regime for insurance companies carried out by The European Commission and the Committee of European Insurance and Occupational Pensions Supervisors.

Aim: to establish an improved solvency system that protects interests of policyholders by reducing the likelihood of prudential failure. (Equitable Life!)

Structure: three-pillar system. Pillar 1 sets out the minimum capital requirements (MCR) for insurance, market, credit and operational risk; Pillar 2 defines the supervisory review process and Pillar 3 the disclosure and transparency requirements.
The UK Treasury on Solvency II

“There is a strong economic rationale for a reformed EU-wide solvency framework which is forward-looking in its assessment of risk and brings regulatory capital into line with economic capital. However, Solvency II cannot just be about capital requirements; no amount of capital can substitute for the capacity to understand, measure and manage risk and no formula or model can capture every aspect of the risks an insurer faces. The new framework should promote higher quality risk management, working with the grain of industry developments, and ensure that the assessment of regulatory capital is integrated with firms’ wider capital management processes.” [Treasury and FSA, 2006]
Economic Capital: What Is It?

- Economic capital is the capital required by a bank/insurer to limit the probability of insolvency to a given level over a given horizon.

- Whereas regulatory capital is based largely on external rules that are intended to ensure a level playing field, economic capital is an attempt to measure risk in terms of economic realities.

- Many companies see economic capital models as part of their response to Pillar II (supervisory review) of the regulatory regime.

- At its most general, economic capital should offer a firm-wide language for discussing and pricing risk and assessing the return on risk capital. A bank with a good economic capital model would hope to be able to use its capital more efficiently.
A 2007 study states:

“there is significantly increased experience in using Economic Capital across the whole financial services sector (e.g. for banking, frameworks have been in place an average of over 6 years and for insurance 4) and firms now feel broadly comfortable with the accuracy of outputs (75%+ for both insurance and banking). This in turn has meant that far more institutions feel sufficiently comfortable with their Economic Capital results to use them in discussions with external stakeholders, and there is increased use in business applications, albeit often as supplementary information rather than as a core driver.”

[IFRI Foundation and CRO Forum, 2007]
The 2006 IFRI/CRO Forum survey

17 banks taking part:

- **Europe:** ABN Amro, Allied Irish Banks, Barclays, Credit Suisse, Deutsche Bank, Dresdener Bank, EuroClear, Fortis. **America:** Citigroup, Goldman Sachs, Royal Bank of Canada, State Street. **Asia:** DBS Group, Macquarie Bank, National Australian Bank, Standard Chartered, Westpac.

17 insurance companies taking part:

- **Europe:** Aegon, Allianz, Assicurazioni Generali, Aviva, AXA, Fortis, ING, Munich Re, Prudential, Royal & Sun Alliance. **Switzerland:** Converium, Swiss Re, Winterthur, Zurich. **America:** American International Group, Royal Bank of Canada. **Australia:** Insurance Australia Group.
A3. Important Themes

Extremes Matter

“From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be balanced against the significant advantage of simplification. From the central bank’s corner, the consequences are even more serious because we often need to concentrate on the left tail of the distribution in formulating lender-of-last-resort policies. Improving the characterization of the distribution of extreme values is of paramount importance.”

[Alan Greenspan, Joint Central Bank Research Conference, 1995]
“With globalisation increasing, you’ll see more crises. Our whole focus is on the extremes now - what’s the worst that can happen to you in any situation - because we never want to go through that [LTCM] again. “


Much space is devoted in our book to models for financial risk factors that go beyond the normal (or Gaussian) model and attempt to capture the related phenomena of heavy tails, volatility and extreme values.
The Interdependence and Concentration of Risks

The multivariate nature of risk presents an important challenge. Whether we look at market risk or credit risk, or overall enterprise-wide risk, we are generally interested in some form of aggregate risk that depends on high-dimensional vectors of underlying risk factors such as individual asset values in market risk, or credit spreads and counterparty default indicators in credit risk.

A particular concern in our multivariate modelling is the phenomenon of dependence between extreme outcomes, when many risk factors move against us simultaneously.
Dependent Extreme Values: LTCM

“Extreme, synchronized rises and falls in financial markets occur infrequently but they do occur. The problem with the models is that they did not assign a high enough chance of occurrence to the scenario in which many things go wrong at the same time—the perfect storm scenario.”

[Business Week, September 1998]

In a perfect storm scenario the risk manager discovers that the diversification he thought he had is illusory; practitioners describe this also as a concentration of risk.
“Over the last number of years, regulators have encouraged financial entities to use portfolio theory to produce dynamic measures of risk. VaR, the product of portfolio theory, is used for short-run, day-to-day profit and loss exposures. Now is the time to encourage the BIS and other regulatory bodies to support studies on stress test and concentration methodologies. Planning for crises is more important than VaR analysis. And such new methodologies are the correct response to recent crises in the financial industry.”

[Scholes, 2000]
The Problem of Scale

A further challenge in QRM is the typical scale of our portfolios, which at their most general may represent the entire position in risky assets of a financial institution.

Calibration of detailed multivariate models for all risk factors is an almost impossible task and hence any sensible strategy involves dimension reduction, that is to say the identification of key risk drivers and a concentration on modelling the main features of the overall risk landscape with a fairly broad brush approach.

This applies both to market risk and credit risk models. In the latter, factor models for default dependence are at least as important as detailed models of individual default.
Interdisciplinarity

The quantitative risk manager of the future should have a combined skillset that includes concepts, techniques and tools from many fields:

- mathematical finance;
- statistics and financial econometrics;
- actuarial mathematics;
- non-quantitative skills, especially communication skills;
- humility — QRM is a small piece of a bigger picture!
B. Basic Elements of Risk Modelling

1. Simple Outline of the Problem

2. The Challenge of Valuation

3. Loss Distributions and Risk Measures
B1. Simple Outline of the Problem

We use the language of probability theory. Risks are represented by random variables mapping unforeseen future states of the world into values representing profits and losses. We consider a portfolio which might be:

- a collection of stocks and bonds;
- a book of derivatives;
- a collection of risky loans;
- a portfolio of life insurance policies;
- a financial institution’s overall position in risky assets.
Value

At time \( t \) the portfolio has a value given by \( V_t \).

If time \( t \) represents now and \( t + 1 \) represents the risk management horizon, the goal is to determine

\[
\Delta_{t+1} = V_{t+1} - V_t
\]

the change in value. The future being uncertain, we must estimate the distribution of \( \Delta_{t+1} \), the P&L distribution, or of \( L_{t+1} = -\Delta_{t+1} \), the loss distribution.

The capital a firm will hold is determined according to this distribution, usually as a quantile - this is the VaR idea. The capital serves as a cushion that allows a firm to meet its liabilities and remain solvent despite incurring losses.
Value-at-Risk

Profit & Loss Distribution (P&L)

95% VaR = 1.6

Mean profit = 2.4

5% probability

©2008 (McNeil)
Value-at-Risk II

Loss Distribution

Mean loss = -2.4

95% VaR = 1.6

95% ES = 3.3
B2. The Challenge of Valuation

How do I value the portfolio now? The ideal is market-consistent valuation. What would I get for the portfolio or its constituent parts if I attempted to dispose of it/them on the market now?

**mark-to-market** = the practice of assigning a value to a position based on current market prices for that kind of asset.

If there is no market for assets of a particular kind or if they are only thinly traded then a different method of valuation is required.

**mark-to-model** = the practice of assigning a value to a position based on a pricing model that makes reference to other underlying instruments whose values may be observed in the market.
Determining $V_t$

At time $t$ (now) ideally assets will be marked-to-market, although this really depends on their level of liquidity.

**easy:** stocks, bonds, foreign currency, exchange-traded derivatives

**more challenging:** over-the-counter derivatives

**difficult:** tranches of CDOs, loans, mortgages, life insurance products

For the difficult assets marking-to-model will typically be necessary.

But both methods of valuation may neglect liquidity risk.

**liquidity risk** = risk stemming from lack of marketability of an investment that cannot be sold quickly enough to prevent a loss

©2008 (McNeil)
Determining $V_{t+1}$

The market prices in the future are unknown and must be forecast. In practice we project forward a future market in the fundamental underlying risk factors: equities, interest rates, currencies, etc.

This is done on the basis of probability distributions, so that the future values of risk factors $Z_t = (Z_{t1}, \ldots, Z_{td})$ are treated as random variables arising from stochastic processes.

$$Z_t \Rightarrow \text{stochastic model for } (Z_t) \Rightarrow Z_{t+1}$$

Certain assets may be marked directly to this hypothetical market. Derivative positions will generally be marked to a model that takes the forecasts of these underlyings as inputs.
Think of a Number...

“It is a worry, though, that Merrill can justify a write-down of $4.5bn one week and $7.9bn just three weeks later. The sense that valuation is still a matter of pick a number and divide by the chief trader’s golf handicap, more than anything else, explains why the super-SIV proposed by Citigroup and others failed to reassure the market.”

It is no wonder that commentators have talked of valuation as mark-to-myth !!!
B3. Loss Distributions and Risk Measures

Risk measures attempt to quantify the riskiness of a portfolio and are used for a variety of purposes:

- **Determination of risk capital**: risk measure gives amount of capital needed as a buffer against (unexpected) future losses to satisfy a regulator.

- **Management tool**: risk measures are used in internal limit systems.

- **Insurance premia** can be viewed as measure of riskiness of insured claims.

Most commonly used risk measures (lik VaR) are based on the loss distribution of $L = L_{t+1}$. We denote the loss distribution function by $F_L$ so that $P(L \leq x) = F_L(x)$. 
Let $0 < \alpha < 1$. We use

- **Value at Risk** is defined as

\[
\text{VaR}_\alpha = q_\alpha(F_L) = F_L^{-1}(\alpha), \tag{1}
\]

where we use the notation $q_\alpha(F_L)$ or $q_\alpha(L)$ for a quantile of the distribution of $L$ and $F_L^{-1}$ for the (generalized) inverse of $F_L$.

- Provided $E(|L|) < \infty$ **expected shortfall** is defined as

\[
\text{ES}_\alpha = \frac{1}{1 - \alpha} \int_{\alpha}^{1} q_u(F_L) du. \tag{2}
\]
Losses and Profits

Loss Distribution

Mean loss = -2.4
95% VaR = 1.6
95% ES = 3.3

©2008 (McNeil)
Expected Shortfall

For continuous loss distributions there is a more intuitive way of understanding expected shortfall - it is the expected loss, given that the VaR is exceeded.

For any $\alpha \in (0, 1)$ we have

$$\text{ES}_{\alpha} = E(L \mid L \geq \text{VaR}_\alpha).$$

For a discontinuous loss distribution we have the more complicated expression

$$\text{ES}_{\alpha} = E(L \mid L \geq \text{VaR}_\alpha) + \text{VaR}_\alpha \frac{1 - \alpha - P(L \geq \text{VaR}_\alpha)}{1 - \alpha}.$$ 

[Acerbi and Tasche, 2002].
Coherent Measures of Risk

There are many possible measures of the risk of a portfolio, such as VaR, ES, or even variance or standard deviation. To decide which are reasonable risk measures a systematic approach is called for.

[Artzner et al., 1999] proposed a list of four axioms and defined a coherent risk measure to be a real-valued function $\varrho$ on a suitable space of random variables (representing losses) that satisfied the axioms.
The Axioms

1. **Monotonicity.** For two rvs with $L_1 \geq L_2$ we have $\varrho(L_1) \geq \varrho(L_2)$.

2. **Subadditivity.** For any $L_1, L_2$ we have $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$. This is the most debated property. Necessary for following reasons:
   - Reflects idea that risk can be reduced by *diversification* and that “a merger creates no extra risk”.
   - Makes *decentralized* risk management possible.
   - If a regulator uses a non-subadditive risk measure, a financial institution could reduce risk capital by splitting into subsidiaries.

3. **Positive homogeneity.** For $\lambda \geq 0$ we have that $\varrho(\lambda L) = \lambda \varrho(L)$. No diversification should imply equality in subadditivity axiom.

4. **Translation invariance.** For any $a \in \mathbb{R}$, $\varrho(L + a) = \varrho(L) + a$. 
Remarks

- VaR is in general not coherent. ES (as we have defined it) is coherent.

- Non-subadditivity of VaR is relevant in presence of skewed loss distributions (credit-risk management, derivative books), or if traders optimize against VaR.

- In the special idealized case that all losses may be represented as linear portfolios of a common set of underlying risk factors with a so-called elliptical distribution (for example a multivariate normal or Student $t$ distribution), then VaR may be shown to be sub-additive, and hence coherent.
Non-Coherence of VaR: an Example

Consider portfolio of 50 defaultable bonds with independent defaults. Default probability identical and equal to 2%. Current price of bonds equal to 95, face value equal to 100.

- Portfolio A: buy 100 units of bond 1; current value is \( V_0 = 9500 \).
- Portfolio B: buy 2 units of each bond; current value is \( V_0 = 9500 \).

Common sense. Portfolio B is less risky (better diversified) than Portfolio A. This is wrong if we measure risk with VaR!

Loss of each bond equals

\[
L_i := -(100(1 - Y_i) - 95) = 100Y_i - 5,
\]

where \( Y_i = 1 \) if default occurs, \( Y_i = 0 \) else. \( Y_i \) are iid Bernoulli(0.02).
Non-Coherence of VaR: an Example II

**Portfolio A:** \( L = 100L_1 \) and hence

\[
\text{VaR}_{0.95}(L) = 100 \text{VaR}_{0.95}(L_1) = -500,
\]

i.e. we may take 500 out of portfolio and still satisfy regulator.

**Portfolio B:** \( L = \sum_{i=1}^{50} 2L_i = 200 \sum_{i=1}^{50} Y_i - 500 \), and hence

\[
\text{VaR}_\alpha(L) = 200 q_\alpha \left( \sum_{i=1}^{50} Y_i \right) - 500.
\]

Inspection shows that \( q_{0.95}(\sum_{i=1}^{50} Y_i) = 3 \), so that \( \text{VaR}_{0.95}(L) = 100 \), i.e. extra capital is needed to hold the portfolio.

This is directly linked to non-coherence of VaR.
C. Market Risk

1. Mapping Portfolio Risks

2. Linearization of Portfolio Losses

3. Variance-Covariance Method

4. Historical Simulation Method

5. Monte Carlo Simulation Method

6. Comparison of Methods
C1. Mapping Portfolio Risks

Consider a portfolio and let $V_t$ denote its value at time $t$; we assume this random variable is observable at time $t$.

Suppose we look at risk from perspective of time $t$ and we consider the time period $[t, t+1]$. The value $V_{t+1}$ at the end of the time period is unknown to us.

The distribution of $(V_{t+1} - V_t)$ is known as the profit-and-loss or P&L distribution. We denote the loss by $L_{t+1} = -(V_{t+1} - V_t)$. By this convention, losses will be positive numbers and profits negative.

We refer to the distribution of $L_{t+1}$ as the loss distribution.
Introducing Risk Factors

The Value $V_t$ of the portfolio/position will be modelled as a function of time and a set of $d$ underlying risk factors. We write

$$V_t = f(\tau_t, Z_t) \quad (3)$$

where $Z_t = (Z_{t,1}, \ldots, Z_{t,d})'$ is the risk factor vector. This representation of portfolio value is known as a mapping. Examples of typical risk factors:

- (logarithmic) prices of financial assets
- yields
- (logarithmic) exchange rates
Risk Factor Changes

We define the time series of risk factor changes by

\[ X_t := Z_t - Z_{t-1}. \]

Typically, historical risk factor time series are available and it is of interest to relate the changes in these underlying risk factors to the changes in portfolio value.

We have

\[ L_{t+1} = -(V_{t+1} - V_t) \]
\[ = -(f(\tau_{t+1}, Z_{t+1}) - f(\tau_t, Z_t)) \]
\[ = -(f(\tau_{t+1}, Z_t + X_{t+1}) - f(\tau_t, Z_t)) \]

(4)
The Loss Operator

Since the risk factor values $Z_t$ are known at time $t$ the loss $L_{t+1}$ is determined by the risk factor changes $X_{t+1}$.

Given realisation $z_t$ of $Z_t$, the loss operator at time $t$ is defined as

$$l_{[t]}(x) := -(f(\tau_{t+1}, z_t + x) - f(\tau_t, z_t)),$$  \hspace{1cm} (5)

so that

$$L_{t+1} = l_{[t]}(X_{t+1}).$$

From the perspective of time $t$ the loss distribution of $L_{t+1}$ is determined by the multivariate distribution of $X_{t+1}$.

But which distribution exactly? **Conditional** distribution of $L_{t+1}$ given history up to and including time $t$ or **unconditional** distribution under assumption that $(X_t)$ form stationary time series?
Conditional or Unconditional Models?

This issue is related to the time series properties of \((X_t)_{t \in \mathbb{N}}\), the series of risk factor changes. If we assume that \(X_t, X_{t-1}, \ldots\) are iid random vectors, the issue does not arise. But, if we assume that they form a strictly stationary multivariate time series then we must differentiate between conditional and unconditional.

Many standard accounts of risk management fail to make the distinction between the two.

If we cannot assume that risk factor changes form a stationary time series for at least some window of time extending from the present back into intermediate past, then any statistical analysis of loss distribution is difficult.
The Conditional Problem

Let $\mathcal{F}_t$ represent the history of the risk factors up to the present.

More formally $\mathcal{F}_t$ is sigma algebra generated by past and present risk factor changes $(X_s)_{s \leq t}$.

In the conditional problem we are interested in the distribution of $L_{t+1} = l_{[t]}(X_{t+1})$ given $\mathcal{F}_t$, i.e. the conditional (or predictive) loss distribution for the next time interval given the history of risk factor developments up to present.

This problem forces us to model the dynamics of the risk factor time series and to be concerned in particular with predicting volatility. This seems the most suitable approach to market risk.
The Unconditional Problem

In the unconditional problem we are interested in the distribution of \( L_{t+1} = l_{[t]}(X) \) when \( X \) is a generic vector of risk factor changes with the same distribution \( F_X \) as \( X_t, X_{t-1}, \ldots \).

When we neglect the modelling of dynamics we inevitably take this view. Particularly when the time interval is large, it may make sense to do this. Unconditional approach also typical in credit risk.

More Formally

Conditional loss distribution: distribution of \( l_{[t]}(\cdot) \) under \( F_{X_{t+1}|\mathcal{F}_t} \).

Unconditional loss distribution: distribution of \( l_{[t]}(\cdot) \) under \( F_X \).
Simple Equity Portfolio

Consider $d$ stocks; let $\alpha_i$ denote number of shares in stock $i$ at time $t$ and let $S_{t,i}$ denote price.

The risk factors: following standard convention we take logarithmic prices as risk factors $Z_{t,i} = \log S_{t,i}, 1 \leq i \leq d$.

The risk factor changes: in this case these are $X_{t+1,i} = \log S_{t+1,i} - \log S_{t,i}$, which correspond to the so-called log-returns of the stock.

The Mapping (3)

$$V_t = \sum_{i=1}^{d} \alpha_i S_{t,i} = \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}}. \quad (6)$$
BMW and Siemens Data: 1972 days to 23.07.96.
Respective prices on evening 23.07.96: 844.00 and 76.9. Consider portfolio in ratio 1:10 on that evening.
Risk Factor Returns

BMW

Siemens

BMW and Siemens Log Return Data: 1972 days to 23.07.96.
Example Continued

The Loss (4)

\[ L_{t+1} = - \left( \sum_{i=1}^{d} \alpha_i e^{Z_{t+1,i}} - \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}} \right) \]

\[ = - \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}} (e^{X_{t+1,i}} - 1) \] (7)

The loss operator (5)

\[ l_{[t]}(x) = - \sum_{i=1}^{d} \alpha_i e^{z_{t,i}} (e^{x_i} - 1), \]

Numeric Example: \[ l_{[t]}(x) = - (844(e^{x_1} - 1) + 769(e^{x_2} - 1)) \]
C2. Linearization of Loss

Recall the general formula (4) for the loss $L_{t+1}$ in time period $[t, t + 1]$. If the mapping $f$ is differentiable we may use the following first order approximation for the loss

$$L_{t+1}^\Delta = - \left( f_\tau(\tau_t, Z_t) \Delta t + \sum_{i=1}^{d} f_{z_i}(\tau_t, Z_t) X_{t+1,i} \right), \quad (8)$$

- $f_{z_i}$ is partial derivative of mapping with respect to risk factor $i$
- $f_\tau$ is partial derivative of mapping with respect to time
- The term $f_\tau(\tau_t, Z_t) \Delta t$ only appears when mapping explicitly features time, and is sometimes neglected. $\Delta t$ is the length of time horizon expressed in the time unit of the mapping.
Recall the loss operator (5) which applies at time $t$. We can obviously also define a linearized loss operator

$$l_{[t]}^\Delta(x) = - \left( f_\tau(\tau_t, z_t) \Delta t + \sum_{i=1}^d f_{z_i}(\tau_t, z_t)x_i \right) ,$$

(9)

where notation is as in previous slide and $z_t$ is realisation of $Z_t$.

Linearization is convenient because linear functions of the risk factor changes may be easier to handle analytically. It is crucial to the variance-covariance method. The quality of approximation is best if we are measuring risk over a short time horizon and if portfolio value is almost linear in risk factor changes.
Here there is no explicit time dependence in the mapping (6). The partial derivatives with respect to risk factors are

\[ f_{zi}(\tau_t, z_t) = \alpha_i e^{z_{t,i}}, \quad 1 \leq i \leq d, \]

and hence the linearized loss (8) is

\[ L^\Delta_{t+1} = -\sum_{i=1}^{d} \alpha_i e^{Z_{t,i}} X_{t+1,i} = -V_t \sum_{i=1}^{d} \omega_{t,i} X_{t+1,i}, \]

where \( \omega_{t,i} = \alpha_i S_{t,i}/V_t \) is relative weight of stock \( i \) at time \( t \). This formula may be compared with (7).

**Numeric Example:** \( l^\Delta_{[t]}(x) = -(844x_1 + 769x_2) \)
European Call Option

Consider portfolio consisting of one standard European call on a non-dividend paying stock $S$ with maturity $T$ and exercise price $K$.

The Black-Scholes value of this asset at time $t$ is $C^{BS}(\tau_t, S_t, r, \sigma)$ where

$$C^{BS}(\tau, S; r, \sigma) = S\Phi(d_1) - Ke^{-r(T-\tau)}\Phi(d_2),$$

$\Phi$ is standard normal df, $r$ represents risk-free interest rate, $\sigma$ the volatility of underlying stock, and where

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T - \tau)}{\sigma\sqrt{T - \tau}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T - \tau}.$$ 

While in BS model, it is assumed that interest rates and volatilities are constant, in reality they tend to fluctuate over time; they should be added to our set of risk factors.
Example Continued

The risk factors: \( Z_t = (\log S_t, r_t, \sigma_t)' \).

The risk factor changes: \( X_t = (\log(S_t/S_{t-1}), r_t - r_{t-1}, \sigma_t - \sigma_{t-1})' \).

The mapping (3)
\[
V_t = C^{BS}(\tau_t, S_t; r_t, \sigma_t),
\]

The loss/loss operator could be calculated from (4). For derivative positions it is quite common to calculate linearized loss.

The linearized loss (8)
\[
L_{t+1}^{\Delta} = - \left( f_{\tau}(\tau_t, Z_t) \Delta t + \sum_{i=1}^{3} f_{z_i}(\tau_t, Z_t) X_{t+1,i} \right).
\]
The Greeks

It is more common to write the linearized loss as

\[ L_{t+1}^\Delta = - \left( C_{\tau}^{BS} \Delta t + C_{S}^{BS} S_t X_{t+1,1} + C_{r}^{BS} X_{t+1,2} + C_{\sigma}^{BS} X_{t+1,3} \right) , \]

in terms of the derivatives of the BS formula.

- \( C_{S}^{BS} \) is known as the delta of the option.
- \( C_{\sigma}^{BS} \) is the vega.
- \( C_{r}^{BS} \) is the rho.
- \( C_{\tau}^{BS} \) is the theta.
C3. Variance-Covariance Method

Further Assumptions

- We assume $X_{t+1}$ has a multivariate normal distribution (either unconditionally or conditionally).

- We assume that the linearized loss in terms of risk factors is a sufficiently accurate approximation of the loss. We consider the problem of estimating the distribution of

$$L^\Delta = l_{[t]}^\Delta(X_{t+1}),$$
Theory Behind Method

Assume $X_{t+1} \sim N_d(\mu, \Sigma)$.

Assume the linearized loss operator (9) has been determined and write this for convenience as

$$l^\Delta_{[t]}(x) = -\left(c + \sum_{i=1}^{d} w_i x_i\right) = -(c + w'x).$$

The loss distribution is approximated by the distribution of

$L^\Delta = l^\Delta_{[t]}(X_{t+1}).$

Now since $X_{t+1} \sim N_d(\mu, \Sigma) \Rightarrow w'X_{t+1} \sim N(w'\mu, w'\Sigma w)$, we have

$L^\Delta \sim N(-c - w'\mu, w'\Sigma w).$
Implementing the Method

1. The constant terms in $c$ and $w$ are calculated.

2. The mean vector $\mu$ and covariance matrix $\Sigma$ are estimated from data $X_{t-n+1}, \ldots, X_t$ to give estimates $\hat{\mu}$ and $\hat{\Sigma}$.

3. Inference about the loss distribution is made using distribution $N(-c - w'\hat{\mu}, w'\hat{\Sigma}w)$.

4. Estimates of the risk measures $\text{VaR}_\alpha$ and $\text{ES}_\alpha$ are calculated from the estimated distribution of $L^\Delta$. 

©2008 (McNeil)
Estimating Risk Measures

- Value-at-Risk. $\text{VaR}_\alpha$ is estimated by

$$\hat{\text{VaR}}_\alpha = -c - w'\hat{\mu} + \sqrt{w'\hat{\Sigma}w} \cdot \Phi^{-1}(\alpha).$$

- Expected Shortfall. $\text{ES}_\alpha$ is estimated by

$$\hat{\text{ES}}_\alpha = -c - w'\hat{\mu} + \sqrt{w'\hat{\Sigma}w} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}.$$

**Remark.** For a rv $Y \sim N(0, 1)$ it can be shown that $E(Y \mid Y > \Phi^{-1}(\alpha)) = \phi(\Phi^{-1}(\alpha))/(1 - \alpha)$ where $\phi$ is standard normal density and $\Phi$ the df.
Pros and Cons, Extensions

• Pros. In contrast to the methods that follow, variance-covariance offers analytical solution with no simulation.

• Cons. Linearization may be crude approximation. Assumption of normality may seriously underestimate tail of loss distribution.

• Extensions. Instead of assuming normal risk factors, the method could be easily adapted to use multivariate Student t risk factors or multivariate hyperbolic risk factors, without sacrificing tractibility. (Method works for all elliptical distributions.)
C4. Historical Simulation Method

The Idea

Instead of estimating the distribution of $L = l_{[t]}(X_{t+1})$ under some explicit parametric model for $X_{t+1}$, estimate distribution of the loss operator under empirical distribution of data $X_{t-n+1}, \ldots, X_t$.

The Method

1. Construct the historical simulation data

$$\{\tilde{L}_s = l_{[t]}(X_s) : s = t - n + 1, \ldots, t\}$$ (10)

2. Make inference about loss distribution and risk measures using these historically simulated data: $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$. 
Inference about loss distribution

There are various possibilities in a simulation approach:

- Use **empirical quantile estimation** to estimate the VaR directly from the simulated data. But what about precision?

- Fit a parametric univariate distribution to \( \tilde{L}_{t-n+1}, \ldots, \tilde{L}_t \) and calculate risk measures from this distribution. But which distribution, and will it model the tail?

- Use the techniques of **extreme value theory** to estimate the tail of the loss distribution and related risk measures.
Theoretical Justification

If $X_{t-n+1}, \ldots, X_t$ are iid or more generally stationary, convergence of empirical distribution to true distribution is ensured by suitable version of law of large numbers.

Pros and Cons

- **Pros.** Easy to implement. No statistical estimation of the distribution of $X$ necessary.

- **Cons.** It may be difficult to collect sufficient quantities of relevant, synchronized data for all risk factors. Historical data may not contain examples of extreme scenarios.
C5. The Monte Carlo Method

Idea

We estimate the distribution of $L = l_{[t]}(X_{t+1})$ under some explicit parametric model for $X_{t+1}$.

In contrast to the variance-covariance approach we do not necessarily make the problem analytically tractible by linearizing the loss and making an assumption of normality for the risk factors.

Instead we make inference about $L$ using Monte Carlo methods, which involves simulation of new risk factor data.
The Method

1. With the help of the historical risk factor data $X_{t-n+1}, \ldots, X_t$ calibrate a suitable statistical model for risk factor changes and simulate $m$ new data $\tilde{X}_{t+1}^{(1)}, \ldots, \tilde{X}_{t+1}^{(m)}$ from this model.

2. Construct the Monte Carlo data
   \[ \{\tilde{L}_i = l_t(\tilde{X}_t^{(i)}), \ i = 1, \ldots, m\}. \]

3. Make inference about loss distribution and risk measures using the simulated data $\tilde{L}_1, \ldots, \tilde{L}_m$. We have similar possibilities as for historical simulation.
Pros and Cons

• Pros. Very general. No restriction in our choice of distribution for $X_{t+1}$.

• Cons. Can be very time consuming if loss operator is difficult to evaluate, which depends on size and complexity of portfolio.

Note that MC approach does not address the problem of determining the distribution of $X_{t+1}$. 
C6. Comparison of Methods

> Xdata <- DAX[(5147:6146),c("BMW","SIEMENS")]
> X <- seriesData(Xdata)

# Set stock prices and number of units
> alpha <- cbind(1,10)
> Sprice <- cbind(844,76.9)

#1. Implement variance-covariance analysis

> weights <- alpha*Sprice
> muhat <- apply(X,2,mean)
> Sigmahat <- var(X)
> meanloss <- -sum(weights*muhat)
> varloss <- weights %*% Sigmahat %*% t(weights)
> VaR99 <- meanloss + sqrt(varloss)*qnorm(0.99)
> ES99 <- meanloss +sqrt(varloss)*dnorm(qnorm(0.99))/0.01

#2. Implement a historical simulation analysis

> loss.operator <- function(x,weights){
   -apply((exp(x)-1)*matrix(weights,nrow=dim(x)[1],ncol=length(weights),byrow=T),1,sum)}
> hsdata <- loss.operator(X,weights)
> VaR99.hs <- quantile(hsdata,0.99)
> ES99.hs <- mean(hsdata[hsdata > VaR99.hs])

©2008 (McNeil)
#3a. Implement a Monte Carlo simulation analysis with Gaussian risk factors

> X.new <- rmnorm(10000,Sigma=Sigmahat,mu=muhat)
> mcdata <- loss.operator(X.new,weights)
> VaR99.mc <- quantile(mcdata,0.99)
> ES99.mc <- mean(mcdata[mcdata > VaR99.mc])

#3b. Implement alternative Monte Carlo simulation analysis with t risk factors

> model <- fit.t(X, nu=NA)
> X.new <- rmt(10000, df=model$nu, Sigma=model$Sigma, mu=model$mu)
> mcdatat <- loss.operator(X.new,weights)
> VaR99.mct <- quantile(mcdatat,0.99)
> ES99.mct <- mean(mcdatat[mcdatat > VaR99.mct])

#Draw pictures

> hist(hsdata,nclass=20,prob=T)
> abline(v=c(VaR99,ES99))
> abline(v=c(VaR99.hs,ES99.hs),col=2)
> abline(v=c(VaR99.mc,ES99.mc),col=3)
> abline(v=c(VaR99.mct,ES99.mct),col=4)
D. Credit Risk

1. Introduction to Credit Risk
2. Basics of Credit Risk Modelling
3. Default Dependence
4. Merton’s Model and the KMV Model
5. Models Based on Credit Ratings
6. Bernoulli Mixture Models
7. KMV/Creditmetrics as a Mixture Model
8. Large Portfolios and Basel II
D1. Introduction to Credit Risk

“Credit risk is the risk that the value of a portfolio changes due to unexpected changes in the credit quality of issuers or trading partners. This subsumes both losses due to defaults and losses due to downgradings of obligors in a rating system.” (paraphrased from MFE)

obligor = a counterparty who has a financial obligation to us.

default = failure to fulfil that obligation, for example, failure to pay interest on a loan or the coupon on a bond; generally due to insolvency; may entail bankruptcy.
Credit-Risky Instruments

• Loans. Secured loans like mortgages or unsecured loans like credit-card debt. Residential lending; consumer lending; commercial lending. Terminology: principal and interest.

• Bonds. A debt security issued by a company or country for a fixed period entitling investor to repayment of principal plus interest (generally in installments known as coupons).

• OTC Derivatives. OTC (over-the-counter as opposed to exchange-traded) derivatives, such as swaps, carry the risk that the counterparty may default affecting the pay-off of the transaction.

• Credit Derivatives. OTC derivatives specifically designed to transfer credit risk from one party to another; they take a myriad of forms.
Specific Issues in Credit Risk Management

In credit risk management there are a number of specific challenges:

• Data. There is a lack of public information on credit quality of obligors. Rating agencies play a role but not all risks rated.

• Longer time horizon - typically at least one year.

• Loss distributions are typically strongly skewed with long upper tail, leading to frequent small gains and occasional large losses.

• Modelling of dependence across a portfolio is even more important than in market risk management, as tail of loss distribution is influenced strongly by specification of dependence between defaults.
Basel II and Credit Risk

**Rationale** for the New Accord: more flexibility and risk sensitivity. Banks may apply the standardized approach to credit risk introduced in Basel I but sophisticated banks now have the option of IRB approach.

**Structure** of the New Accord: Three-pillar framework:

1. Pillar 1: minimal capital requirements (risk measurement)
2. Pillar 2: supervisory review of capital adequacy
3. Pillar 3: public disclosure
The Standardized Approach

- Compute **risk-weighted assets** by multiplying individual credit exposures by appropriate risk weights.

- Risk weights determined by external ratings of obligors, where available. For example AAA to AA- corporate bonds have weight 20% and BBB+ to BB- bonds have weight 100%.

- Retail products (overdrafts and credit cards) have weight 75% and claims secured by residential property have weight 35%.

- Sum risk-weighted assets across the credit portfolio.

- Determine regulatory capital by multiplying total risk-weighted assets by fraction known as **Cooke ratio** or **McDonough ratio**, which is currently 0.08.
The IRB Approach

This comes in various flavours (foundation IRB and advanced IRB) and has been implemented in different ways in different countries.

The basic idea is that banks make their own internal assessments of the default probabilities (PDs) of obligors. In some cases they may also make their own assessments of losses-given-defaults (LGDs). These assessments should be based on quantitative models that are satisfactory to the regulator.

The capital to be held is determined by the **Basel II formula** which takes as basic inputs the estimated PDs, LGDs and also exposure information (EAD - exposure at default). We will study and understand the formula in this course.
There are various classifications of credit risk models. For example it is common to talk of structural models and reduced-form models.

The former, also known as firm-value models, attempt to explain the mechanism by which default takes place. The progenitor of most of these models is Merton’s model, which postulates a mechanism for default in terms of the relationship between a firm’s assets and its liabilities.

The latter leave the mechanism unspecified and concentrate on the modelling of times of default using stochastic models for durations or survival times.
Static and Dynamic Models

For calculation of regulatory or economic capital a fixed time period of (typically) one year is considered. We are concerned only with the numbers of defaults taking place over the time period and the magnitudes of the losses. Static (distributional) models suffice.

**Latent or critical variable models** are simplified static versions of the structural models: default (or downgrading) occurs if a critical variable (generally interpreted as asset value) falls below a critical level (liabilities) by the end of the time period.

**Mixture models** are simple static versions of reduced form models which assume conditional independence of defaults given factors.

Fully dynamic models are required in the pricing of credit derivatives.
PDs, EADs and LGDs

Consider a fixed time period \([0, T]\) and let \(\tau_i\) be the (random) time-to-default for obligor \(i\). The default indicator \(Y_i\) is a Bernoulli random variable defined by \(Y_i = I_{\{\tau_i \leq T\}}\) which obviously satisfies

\[
P(Y_i = 1) = 1 - P(Y_i = 0) = P(\tau_i \leq T) = F_{\tau_i}(T) =: PD_i.
\]

In a simple model that considers only losses caused by default we have

\[
\text{Portfolio Loss} = \sum_{i=1}^{m} \text{EAD}_i \times Y_i \times \text{LGD}_i.
\]

Assuming deterministic exposures and losses given default we get

\[
E(\text{Portfolio Loss}) = \sum_{i=1}^{m} \text{EAD}_i \times PD_i \times \text{LGD}_i.
\]
Expected and Unexpected Losses

A bank effectively covers the expected losses by charging them to its obligors in the form of a risk premium. This happens explicitly in loan pricing where customers pay premiums that are determined by their creditworthiness. With bonds the coupon payments contain the implicit risk premiums.

Risk managers are more concerned with losses that exceed EL. One measure of this risk is the so-called unexpected loss - the standard deviation of the portfolio loss.

However to measure the full potential for losses we need to consider the whole loss distribution and apply a variety of risk measures to this.

The most popular risk measures like Value-at-Risk (VaR) and expected shortfall (ES) describe the right tail of the portfolio loss distribution.
The Loss Distribution

Using more mathematical notation, we write the portfolio loss as

\[ L = \sum_{i=1}^{m} e_i \delta_i Y_i. \]

where \( e_i \) denotes the exposure, \( \delta_i \) denotes the LGD and \( Y_i \) is the default indicator as before. Probabilities of default will be denoted \( p_i := P(Y_i = 1) \).
D3. Default Dependence

Dependence between defaults is a key issue in credit risk management.

Reasons for dependence between defaults:

- Dependence caused by common factors (for example, interest rates and changes in economic growth) affecting all obligors

- Default of company A may have direct impact on default probability of company B and vice versa because of direct business relations, a phenomenon known as contagion
Empirical Evidence for Default Dependence

Standard and Poor’s default data from 1980 to 2000 show clear evidence of cycles; we expect within-time-period and between-time-period dependence.
Comparison of the loss distribution of a homogeneous portfolio of 1000 loans with a default probability of 1% assuming (i) independent defaults and (ii) a default correlation of 0.5%.
**Default Correlation**

**Definition.** Default or event-correlation of firms $i$ and $j$ is given by $\rho(Y_i, Y_j)$.

Denote by $p_i = P(Y_i = 1) = E(Y_i)$ unconditional default probabilities. We have $\text{cov}(Y_i, Y_j) = E(Y_i Y_j) - p_i p_j$ and $\text{var}(Y_i) = p_i (1 - p_i)$. Hence

$$\rho(Y_i, Y_j) = \frac{E(Y_i Y_j) - p_i p_j}{\sqrt{p_i (1 - p_i)} \sqrt{p_j (1 - p_j)}}$$

so that default correlation can be computed from joint default probability $E(Y_i Y_j) = P(Y_i = 1, Y_j = 1)$. 

©2008 (McNeil)
D4. Merton’s Model and the KMV Model

Merton’s model [Merton, 1974] is the prototype of all firm value models; it is an influential benchmark even today.

Modelling of Default: Consider firm with stochastic asset-value \((V_t)\), financing itself by equity (i.e. by issuing shares) and debt. Denote by \(S_t\) and \(F_t\) value at time \(t \leq T\) of equity and debt so that \(V_t = S_t + F_t, 0 \leq t \leq T\). Assume that

- Debt consists of single zero coupon bond with face value \(F\) and maturity \(T\).
- Default occurs if the firm misses a payment to its debt holders and hence only in \(T\).
Modelling of Default continued

In $T$ we have two possible cases

i) $V_T \geq F$. In that case the debtholders receive $F$; shareholders receive residual value $S_T = V_T - F$, and there is no default.

ii) $V_T < F$. In that case the firm cannot meet its financial obligations, and shareholders hand over control to the bondholders, who liquidate the firm; hence we have $F_T = V_T, S_T = 0$.

In summary we obtain

$$S_T = [V_T - F]^+ \text{ and } F_T = \min(V_T, F) = F - [F - V_T]^+. \quad (11)$$
Dynamics of \((V_t)\) and Default Probability

It is assumed that asset value \((V_t)\) follows a diffusion of the form
\[
dV_t = \mu_V V_t dt + \sigma_V V_t dW_t
\]
for constants \(\mu_V \in \mathbb{R}, \sigma_V > 0\), and a Brownian motion \((W_t)_{t \geq 0}\), so that
\[
V_T = V_0 \exp\left((\mu_V - \frac{1}{2}\sigma_V^2)T + \sigma_V W_T\right);
\]
in particular \(\ln V_T \sim N\left(\ln V_0 + (\mu_V - \frac{1}{2}\sigma_V^2)T, \sigma_V^2 T\right)\).

Under this assumption the default probability of our firm is readily computed. We have
\[
P(V_T < F) = P(\ln V_T < \ln F) = \Phi\left(\frac{\ln \frac{F}{V_0} - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}}\right). \quad (12)
\]

In line with economic intuition this is increasing in \(F\) and \(\sigma_V\) and decreasing in \(V_0\) and \(\mu_V\).
The KMV Model

The KMV model, developed in 1990s, is a popular extension of Merton’s model, which is now maintained by Moody’s KMV.

**Contributions of KMV:** extends Merton’s model to give better empirical performance; implementation using proprietary database.

**Key concept: EDF (expected default frequency).** This is simply the one-year default probability of a given firm. In Merton’s model we get from (12)

\[
EDF_{Merton} = 1 - \Phi \left( \frac{\ln V_0 - \ln F + (\mu_V - \frac{1}{2} \sigma_V^2)}{\sigma_V} \right) .
\]  

In KMV model EDF has similar structure, but \(1 - \Phi\) is replaced by estimated function, and \(V_0\) and \(\sigma_V\) are determined from equity data.
The KMV Model continued

**Determination of** \( V_0 \) **and** \( \sigma_V \).

Market value \( V_0 \) of a firm’s assets is not fully observable.

- Market value and book value can differ widely.
- Only equity and at most a part of liabilities is traded.

Therefore KMV backs out \( V_0 \) from observable value of a firm’s equity using the Merton model by inverting pricing formula. In an iterative procedure asset volatility \( \sigma_V \) (which differs in general from equity volatility) is determined from equity data.

**Remark.** KMV uses more sophisticated pricing model for equity; often \( F \) is taken as sum of short-term and half of long-term debt.
Calculation of EDFs

**Distance to Default.** Relation (13) for $EDF_{Merton}$ might be too simplistic. Therefore KMV defines in an intermediary step the distance to default or DD by

$$DD := \frac{(V_0 - \tilde{F})}{\sigma_V V_0},$$

(14)

where $\tilde{F}$ represents the default threshold (typically liabilities payable within one year). (14) is an approximation of the argument of (13), since $\mu_V$ and $\sigma^2_V$ are small and $\ln V_0 - \ln \tilde{F} \approx (V_0 - \tilde{F}' )/V_0$.

**Calculation of EDFs.** KMV uses DD as state variable and assumes that firms with equal DD have equal EDFs. Functional relation between EDF and DD is not postulated, but estimated empirically.
Applying the KMV Approach: an Example

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market value of equity</td>
<td>$110,688</td>
<td>Share Price × # Outstanding.</td>
</tr>
<tr>
<td>Book Liabilities</td>
<td>$64,062</td>
<td>From balance sheet.</td>
</tr>
<tr>
<td>Market value of assets</td>
<td>$170,558</td>
<td>From option-pricing model.</td>
</tr>
<tr>
<td>Asset volatility</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Default point</td>
<td>$47,499</td>
<td>Liabilities payable within 1 year.</td>
</tr>
<tr>
<td>Distance-to-default</td>
<td>3.5</td>
<td>Given by ratio $\frac{170 - 47}{0.21 \times 170}$.</td>
</tr>
<tr>
<td>EDF (one year)</td>
<td>0.25 %</td>
<td>Determined using empirical mapping between DD and EDF.</td>
</tr>
</tbody>
</table>

The example is concerned with situation of Phillip Morris Inc. at the end of April 2001. Financial quantities in million USD. Example provided by KMV.
D5. Models Based on Credit Migration

Overview. In credit migration approach each firm is given a particular credit rating measuring its credit quality; moreover, probability of moving from one rating to another (including default) is specified.

Credit rating serves as state variable, i.e. firms with equal rating have equal transition probabilities.

Credit ratings for major companies or sovereigns and rating transition matrices are provided by rating agencies such as Moody’s or Standard & Poor’s (S&P).

Standard industry model in this class is CreditMetrics developed by J.P. Morgan and the RiskMetrics Group.
### Transition Probabilities: an Example from S&P

<table>
<thead>
<tr>
<th>Initial rating</th>
<th>Rating at year-end(%)</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>90.81</td>
<td>8.33</td>
<td>0.68</td>
<td>0.06</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>0.70</td>
<td>90.65</td>
<td>7.79</td>
<td>0.64</td>
<td>0.06</td>
<td>0.14</td>
<td>0.02</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.09</td>
<td>2.27</td>
<td>91.05</td>
<td>5.52</td>
<td>0.74</td>
<td>0.26</td>
<td>0.01</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>BBB</td>
<td>0.02</td>
<td>0.33</td>
<td>5.95</td>
<td>86.93</td>
<td>5.30</td>
<td>1.17</td>
<td>1.12</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>0.03</td>
<td>0.14</td>
<td>0.67</td>
<td>7.73</td>
<td>80.53</td>
<td>8.84</td>
<td>1.00</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.11</td>
<td>0.24</td>
<td>0.43</td>
<td>6.48</td>
<td>83.46</td>
<td>4.07</td>
<td>5.20</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>0.22</td>
<td>0.00</td>
<td>0.22</td>
<td>1.30</td>
<td>2.38</td>
<td>11.24</td>
<td>64.86</td>
<td>19.79</td>
<td></td>
</tr>
</tbody>
</table>

Empirical probabilities of migrating from one rating to another within 1 year. Source: Standard & Poor’s CreditWeek. For example, the 1 year default probability of a B-rated firm is 5.2%.
## Cumulative Default Probabilities according to S&P

<table>
<thead>
<tr>
<th>Term</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5...</th>
<th>7...</th>
<th>10...</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.15</td>
<td>0.24...</td>
<td>0.66...</td>
<td>1.40...</td>
</tr>
<tr>
<td>AA</td>
<td>0.00</td>
<td>0.02</td>
<td>0.12</td>
<td>0.25</td>
<td>0.43...</td>
<td>0.89...</td>
<td>1.29...</td>
</tr>
<tr>
<td>A</td>
<td>0.06</td>
<td>0.16</td>
<td>0.27</td>
<td>0.44</td>
<td>0.67...</td>
<td>1.12...</td>
<td>2.17...</td>
</tr>
<tr>
<td>BBB</td>
<td>0.18</td>
<td>0.44</td>
<td>0.72</td>
<td>1.27</td>
<td>1.78...</td>
<td>2.99...</td>
<td>4.34...</td>
</tr>
<tr>
<td>BB</td>
<td>1.06</td>
<td>3.48</td>
<td>6.12</td>
<td>8.68</td>
<td>10.97...</td>
<td>14.46...</td>
<td>17.73...</td>
</tr>
<tr>
<td>B</td>
<td>5.20</td>
<td>11.00</td>
<td>15.95</td>
<td>19.40</td>
<td>21.88...</td>
<td>25.14...</td>
<td>29.02...</td>
</tr>
<tr>
<td>CCC</td>
<td>19.79</td>
<td>26.92</td>
<td>31.63</td>
<td>35.97</td>
<td>40.15...</td>
<td>42.64...</td>
<td>45.10...</td>
</tr>
</tbody>
</table>

Average cumulative default-rates (%). Source: Standard & Poor’s CreditWeek. To be consistent these rates should be approximately equal to default rates obtained from $n$-fold product of one-year transition probabilities.
Credit Migrations and KMV Approach Compared

Advantages of KMV Approach.

• The KMV EDF reacts quickly to changes in economic prospects of a firm, whereas agencies are often slow to adjust ratings.

• EDFs tend to reflect the current macroeconomic environment and tend to be better predictors of defaults over short time horizons.

Advantages of Credit migration approach.

• KMV approach sensitive to over- and underreactions in equity markets. If KMV model is widely followed this might have destabilizing effects on real economy.

• As rating agencies focus on average credit quality “through the business cycle”, risk capital requirements based on rating transitions fluctuate less, helping to provide liquidity in credit markets.
Credit Migration Models as Firm-Value Models

Given $n + 1$ rating classes $j \in \{0, 1, \ldots, n\}$ of increasing quality (0 is default). Suppose that for a given firm the probability that the firm is in class $j$ at year-end is given by probabilities $q_j$, $0 \leq j \leq n$. Suppose that asset value $V_T$ is lognormal. Choose thresholds $-\infty = \tilde{d}_0 < \tilde{d}_1 < \cdots < \tilde{d}_n < \tilde{d}_{n+1} = \infty$ such that

$$P \left( \tilde{d}_j < V_T \leq \tilde{d}_{j+1} \right) = q_j \text{ for all } j.$$

**Definition.** Firm belongs to rating class $j \in \{0, \ldots, n\}$ at $T$ if and only if $\tilde{d}_j < V_T \leq \tilde{d}_{j+1}$.

**Remark.** Transition probabilities invariant under simultaneous transformation of $V_T$ and thresholds, $\Rightarrow$ work with normally distributed $X = \ln V_T$; Portfolio versions of KMV/CreditMetrics are similar.
D6. Bernoulli Mixture Models

Many static credit risk models can be considered more abstractly as Bernoulli mixture models. There are a number of advantages to this presentation: a common framework; easy to simulate; possible to study large portfolio behaviour; possible to estimate through GLMM methods.

**Definition. (Bernoulli mixture model).** Given some $p < m$ and a $p$-dimensional random vector $\Psi = (\Psi_1, \ldots, \Psi_p)'$, the default indicator vector $Y$ follows a Bernoulli mixture model with factor vector $\Psi$ if there are functions $p_i : \mathbb{R}^p \rightarrow (0, 1)$, such that conditional on $\Psi$ the components of $Y$ are independent Bernoulli rvs with $P(Y_i = 1 \mid \Psi = \psi) = p_i(\psi)$. 
Distribution of Defaults

For \( y = (y_1, \ldots, y_m)' \) in \( \{0, 1\}^m \) we get

\[
P(Y = y \mid \Psi = \psi) = \prod_{i=1}^{m} p_i(\psi)^{y_i}(1 - p_i(\psi))^{1-y_i},
\]

and the unconditional distribution is given by

\[
f(y) = P(Y = y) = \int_{\mathbb{R}^p} \prod_{i=1}^{m} p_i(\psi)^{y_i}(1 - p_i(\psi))^{1-y_i} g(\psi) d\psi,
\]

where \( g(\psi) \) is the probability density of the factors.

By adding exposures and assumptions about losses given default we complete the specification of a one-period model.
CreditRisk$^+$

CreditRisk$^+$ may be represented as a Bernoulli mixture model. Distribution of the default indicators is given by

$$p_i(\psi) = 1 - \exp(-w_i'\psi).$$

Here $\Psi = (\Psi_1, \ldots, \Psi_p)'$ is a vector of independent gamma distributed macroeconomic factors and $w_i = (w_{i,1}, \ldots, w_{i,p})'$ is a vector of positive factor weights.

**Remark:** CreditRisk$^+$ is usually presented as a Poisson mixture where, conditional on $\Psi$, the default of counterparty $i$ occurs independently of other counterparties with a Poisson intensity $\lambda_i(\Psi) = w_i'\Psi$. This leads to the above default probabilities. The model assumptions mean that the distribution of the number of defaults is a sum of independent negative binomials.
One-Factor Mixture Models

Often it is useful to work in a one factor model ($p = 1$):

- Fitting to default data relatively easy
- Behaviour of large portfolios easy to understand

In the exchangeable special case the conditional default probabilities $p_i(\Psi)$ are identical, $\forall i$, making $\mathbf{Y}$ exchangeable. Define the rv $Q := p_1(\Psi)$ with df $G(q)$:

$$
\pi := P(Y_i = 1) = E(Y_i) = E(E(Y_i | Q)) = E(Q)
$$

$$
\pi_k := P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) = E(Q^k) = \int_0^1 q^k dG(q), 
$$

Unconditional default probabilities and joint default probabilities are moments of the mixing distribution.
Exchangeable Bernoulli Mixtures

Default correlations: It follows that, for $i \neq j$,
\[
\text{cov} (Y_i, Y_j) = \pi_2 - \pi^2 = \text{var} \ Q \geq 0.
\]
Hence default correlation is given by
\[
\rho_Y := \text{corr} (Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2}.
\]

Examples of Mixing Distributions

- **Beta** $Q \sim \text{Beta}(a, b), g(q) = \beta(a, b)^{-1} q^{a-1} (1 - q)^{b-1}, a, b > 0$. Corresponds to one-factor CreditRisk$^+$ [Frey and McNeil, 2002].

- **Probit-Normal** $Q = \Phi(\mu + \sigma \Psi), \Psi \sim N(0, 1)$ (CreditMetrics/KMV)

- **Logit-Normal** $Q = (1 + \exp(-\mu - \sigma \Psi))^{-1}, \Psi \sim N(0, 1)$ (CreditPortfolioView)
Parameterizing Mixing Distributions

In these 2-parameter examples, if we fix default probability $\pi$ and default correlation $\rho_Y$ (or $\pi_2$) we fully calibrate the model. For instance, in the exchangeable Beta–Bernoulli mixture model, we have $\pi = a/(a + b)$ and $\pi_2 = \pi(a + 1)/(a + b + 1)$.

Beta Density $g(q)$ of mixing variable $Q$ in exchangeable Bernoulli mixture model with $\pi = 0.005$ and $\rho_Y = 0.0018$. 
Comparison of Exchangeable Models ($\pi$ and $\pi_2$ fixed)

The tail of distribution of $Q$ determines tail of loss distribution of portfolio.

Horizontal line at 0.01 shows that models only diverge at 99th percentile of $Q$; for given $\pi$ and $\pi_2$ there is little model risk.
One-Factor Model with k Homogeneous Groups

Exchangeability too restrictive (e.g. obligors belonging to different rating classes). Hence more general one-factor models are needed.

**Example:** (homogeneous group-model) We have $k$ groups and $r(i) \in \{1, \ldots, k\}$ gives the group membership of firm $i$. We model default probabilities by

$$p_i(\Psi) = h(\mu_{r(i)} + \sigma \Psi), \text{ for } \sigma > 0 \text{ and a real-valued rv } \Psi. \quad (15)$$

If $h$ is increasing, the conditional default probabilities are comonotonic. Possible choices: $h = \Phi$, $\Psi \sim \mathcal{N}(0, 1)$.

This model is akin to GLMM (generalised linear mixed model) and can be fitted by ML.
D7. KMV/CreditMetrics as a Mixture Model

These industry models belong to the class of structural or firm-value models descending from Merton’s influential credit risk model.

We assume that default occurs for counterparty \( i \) if a critical variable \( X_i \) (often interpreted as asset value) lies below a critical threshold \( d_i \) (often interpreted as liabilities) at the end of the time period of interest. We assume \( X = (X_1, \ldots, X_m)' \) satisfies:

- \( X \) has a multivariate normal distribution.

- \( X_i \sim N(0, 1) \) for all \( i \) (since we can standardize \( X_i \) and \( d_i \) without altering the default probability.)

- \( X_i \) follows a standard linear factor model.
Factor Model for Critical Variables

\[ X_i = z_i' \Psi + \varepsilon_i, \quad \text{where} \]

- \( \Psi \sim N_p(0, \Omega) \) is a random vector of normally distributed common economic factors,
- \( z_i \) is a vector of loadings for the \( i \)th counterparty,
- and \( \varepsilon_i \) is a normally distributed error, which is independent of \( \Psi \) and of \( \varepsilon_j \) for \( j \neq i \).

The term \( z_i' \Psi \) is the **systematic risk component** for counterparty \( i \) and has variance \( \beta_i := z_i' \Omega z_i \), whereas \( \varepsilon_i \) is the **idiosyncratic risk component** with variance \( 1 - \beta_i \).
KMV/CreditMetrics as a Mixture Model

KMV/Creditmetrics is a Bermoulli mixture model with factor vector $\Psi$. The conditional independence of defaults given $\Psi$ follows from the independence of the idiosyncratic terms $\varepsilon_1, \ldots, \varepsilon_m$. The conditional default probabilities are

$$p_i(\psi) = P(Y_i = 1 \mid \Psi = \psi) = P(X_i < d_i \mid \Psi = \psi)$$

$$= P(\varepsilon_i < d_i - z_i'\psi) = \Phi \left( \frac{d_i - z_i'\psi}{\sqrt{1 - \beta_i}} \right).$$

We can reparametrize in terms of the default probability $p_i$ of counterparty $i$:

$$p_i(\psi) = \Phi \left( \frac{\Phi^{-1}(p_i) - z_i'\psi}{\sqrt{1 - \beta_i}} \right).$$
Special Cases

One-Factor Model

\[ p_i(\psi) = \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\beta_i} \psi}{\sqrt{1 - \beta_i}} \right), \]

where \( \Psi \) is a standard normally distributed factor.

Homogeneous Correlation Model

\[ p_i(\psi) = \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho} \psi}{\sqrt{1 - \rho}} \right), \quad (16) \]

where \( \rho \) turns out to be the asset correlation between any two critical variables \( X_i \neq X_j \).
Motivation. Consider exchangeable Bernoulli default model of KMV/CreditMetrics type, (defaults are conditionally iid with conditional default probability $Q := p_1(\Psi)$ where $p_1$ has the form (16).) Suppose $e_i = \delta_i = 1$ for all companies. Let $L(m) = \sum_{i=1}^{m} Y_i$ and consider increasing the portfolio size.

Conditional on some realization $\psi$ of the common factor the SLLN says that, almost surely,

$$\lim_{m \to \infty} \frac{L(m)}{m} = q = p_1(\psi).$$

In other words the distribution of $Q = p_1(\Psi)$ can be thought of as the distribution of the portfolio loss (expressed as a fraction) in an infinitely large exchangeable portfolio.
Remarks

• In this example the asymptotic portfolio loss distribution, the distribution of

\[ Q = \Phi \left( \frac{\Phi^{-1}(p) - \sqrt{\rho} \Psi}{\sqrt{1 - \rho}} \right), \]

has been called the Vasicek loss distribution. [Vasicek, 1997]. (It is a probit-normal distribution.)

• The idea of looking at infinitely fine-grained portfolios where the individual risks become negligible and the systematic factor(s) dominate(s) has been taken up by other researchers in more complicated models and has influenced regulation. [Gordy, 2003]
Of course the practical relevance of the large portfolio results is for making computations when \( m \) is large. Consider again the simple example.

- For tail probabilities we have

\[
P \left( L^{(m)} > l \right) \approx P(Q > l/m) = \Phi \left( \frac{\Phi^{-1}(p) - \sqrt{1 - \rho} \Phi^{-1}(l/m)}{\sqrt{\rho}} \right).
\]

- For Value-at-Risk we have

\[
\text{VaR}_\alpha(L^{(m)}) \approx mq_\alpha(Q) = m\Phi \left( \frac{\Phi^{-1}(p) + \sqrt{\rho} \Phi^{-1}(\alpha)}{\sqrt{1 - \rho}} \right).
\]
Implications for Basel II

The formulas coming from a large-portfolio analysis of the KMV/CreditMetrics model have been influential in deriving capital rules for Basel II.

Consider the model in (16). For $m$ large we have

$$q_\alpha \left( L^{(m)} \right) \approx \sum_{i=1}^{m} \delta_i e_i p_i (\Phi^{-1}(\alpha)).$$

In the internal-ratings-based (IRB) approach the capital required for risk $i$ is proportional to

$$\delta_i e_i p_i (\Phi^{-1}(0.999)) = \delta_i e_i \Phi \left( \frac{\Phi^{-1}(p_i) + \sqrt{\rho} \Phi^{-1}(0.999)}{\sqrt{1 - \rho}} \right).$$
Implications for Basel II ctd.

Hence the IRB capital charge can be considered as 99.9% quantile of the portfolio loss in a large homogeneous portfolio following a one-factor KMV-type model.

Parameters.

- Bank specifies $p_i$.
- Correlation parameter $\rho$ and one-factor assumption is imposed by the regulator independently of portfolio considered.

In particular IRB-approach is not a fully internal model.
E. Modelling Operational Risk

**Definition.** The risk of losses resulting from inadequate or failed internal processes, people and systems, or external events.

**Remark:** This definition includes legal risk, but excludes strategic and reputational risk.

**Some examples.**

- 1995: Nick Leeson/Barings Bank, £1.3b
- 2001: September 11
- 2001: Enron (largest US bankruptcy so far)
- 2008: Soc Gen
Risk Measurement Methods for Operational Risk

Pillar 1 regulatory minimal capital requirements for operational risk:

Three distinct approaches:

1. Basic Indicator Approach
2. Standardised Approach
3. Advanced Measurement Approach (AMA)
Basic Indicator Approach

• Capital charge:
  \[ C_{\text{OP}}^{\text{BIA}} = \alpha \times GI \]

• \( C_{\text{OP}}^{\text{BIA}} \): capital charge under the Basic Indicator Approach

• \( GI \): average annual gross income over the previous three years

• \( \alpha = 15\% \) (set by the Committee based on CISs)
Standardised Approach

• Similar to the BIA, but on the level of each business line:

\[ C_{OP}^{SA} = \sum_{i=1}^{8} \beta_i \times GI_i \]

\( \beta_i \in [12\%, 18\%], \ i = 1, 2, \ldots, 8 \) and 3-year averaging

• 8 business lines:

  Corporate finance (18%)  Payment & Settlement (18%)
  Trading & sales (18%)    Agency Services (15%)
  Retail banking (12%)    Asset management (12%)
  Commercial banking (15%) Retail brokerage (12%)
Advanced Measurement Approach (AMA)

- Allows banks to use their internally generated risk estimates
- Preconditions: Bank must meet qualitative and quantitative standards before being allowed to use the AMA
- Incorporation of risk diversification benefits allowed
- “Given the continuing evolution of analytical approaches for operational risk, the Committee is not specifying the approach or distributional assumptions used to generate the operational risk measures for regulatory capital purposes.”
- Example: Loss distribution approach
Loss Distribution Approach (LDA)

- OP-losses are divided into cells corresponding to 8 business lines and 7 loss types

- For each business line/loss type cell \((i, k)\) one models

\[
L_{i,k}^{T+1}: \quad \text{OP risk loss for business line } i, \text{ loss type } k \text{ over the future (one year, say) period } [T, T + 1]
\]

\[
L_{i,k}^{T+1} = \sum_{\ell=1}^{N_{i,k}^{T+1}} X_{i,k}^\ell \quad \text{(next period’s loss for cell } (i, k)\text{)}
\]

This is a compound sum, a typical object from non-life-insurance mathematics.
LDA: continued

<table>
<thead>
<tr>
<th></th>
<th>$RT_1$</th>
<th>$\ldots$</th>
<th>$RT_k$</th>
<th>$\ldots$</th>
<th>$RT_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BL_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BL_i$</td>
<td>$L_{i,k}^{T+1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BL_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$L^{T+1}$
Part II: Integrated Models for Economic Capital

A. Modelling Dependent Risks

B. The Modular Approach

C. The Fully Integrated Approach

D. Capital Allocation
A. Modelling Dependent Risks

1. Introduction to Dependence
2. Introduction to Copulas
3. Sampling Copulas
4. Correlation
5. Tail Dependence
6. Rank Correlation
A1. Introduction to Dependence

Three Extreme Days

History

New York, 19th October 1987

Berlin Wall 16th October 1989

The Kremlin, 19th August 1991
Some Notation

A $d$-dimensional random vector of risk-factor changes $\mathbf{X} = (X_1, \ldots, X_d)'$ has joint df

$$F(x) = F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

The marginal dfs $F_i$ of the individual risks are given by

$$F_i(x_i) = P(X_i \leq x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty).$$

In some cases we work instead with joint survival functions

$$\bar{F}(x) = \bar{F}(x_1, \ldots, x_d) = P(X_1 > x_1, \ldots, X_d > x_d),$$

and marginal survival functions

$$\bar{F}_i(x_i) = P(X_i > x_i) = \bar{F}(-\infty, \ldots, -\infty, x_i, -\infty, \ldots, -\infty).$$
Some Notation II

**Densities.** Joint densities $f(x) = f(x_1, \ldots, x_d)$, when they exist, are related to joint dfs by

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, u_d) du_1 \cdots du_d.$$

**Independence.** The components of $X$ are said to be mutually independent if and only if

$$F(x) = \prod_{i=1}^{d} F_i(x_i), \quad \forall x \in \mathbb{R}^d,$$

or, when $X$ possesses a joint density, if and only if

$$f(x) = \prod_{i=1}^{d} f_i(x_i), \quad \forall x \in \mathbb{R}^d.$$
Multivariate Normal (Gaussian) Distribution

This distribution has joint density

\[ f(x) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{(x - \mu)'\Sigma^{-1}(x - \mu)}{2} \right\}, \]

where \( \mu \in \mathbb{R}^d \) and \( \Sigma \in \mathbb{R}^{d\times d} \) is a positive definite matrix.

- If \( X \) has density \( f \) then \( E(X) = \mu \) and \( \text{cov}(X) = \Sigma \), so that \( \mu \) and \( \Sigma \) are the mean vector and covariance matrix respectively. A standard notation is \( X \sim N_d(\mu, \Sigma) \).

- Clearly, the components of \( X \) are mutually independent if and only if \( \Sigma \) is diagonal. For example, \( X \sim N_d(0, I) \) if and only if \( X_1, \ldots, X_d \) are iid \( N(0, 1) \).
Bivariate Standard Normals

In left plots $\rho = 0.9$; in right plots $\rho = -0.7$. 
Properties of Multivariate Normal Distribution

- The marginal distributions are univariate normal.

- Linear combinations $a'X = a_1X_1 + \cdots + a_dX_d$ are univariate normal with distribution $a'X \sim N(a'\mu, a'\Sigma a)$.

- Conditional distributions are multivariate normal.

- The sum of squares $(X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi^2_d$ (chi-squared).

Simulation.
1. Perform a Cholesky decomposition $\Sigma = AA'$
2. Simulate iid standard normal variates $Z = (Z_1, \ldots, Z_d)'$
3. Set $X = \mu + AZ$. 
Deficiencies of Multivariate Normal for Risk Factors

- Tails of univariate margins are very thin and generate too few extreme values.

- Simultaneous large values in several margins relatively infrequent. Model cannot capture phenomenon of joint extreme moves in several risk factors.

- Very strong symmetry (known as elliptical symmetry). Reality suggests more skewness may often be present.
The Multivariate t Distribution

This has density

\[ f(x) = k_{\Sigma, \nu, d} \left( 1 + \frac{(x - \mu)'\Sigma^{-1}(x - \mu)}{\nu} \right)^{-\frac{(\nu+d)}{2}} \]

where \( \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d} \) is a positive definite matrix, \( \nu \) is the degrees of freedom and \( k_{\Sigma, \nu, d} \) is a normalizing constant.

- If \( X \) has density \( f \) then \( E(X) = \mu \) and \( \text{cov}(X) = \frac{\nu}{\nu-2} \Sigma \), so that \( \mu \) and \( \Sigma \) are the mean vector and dispersion matrix respectively. For finite variances/correlations \( \nu > 2 \). Notation: \( X \sim t_d(\nu, \mu, \Sigma) \).

- If \( \Sigma \) is diagonal the components of \( X \) are uncorrelated. They are not independent.

- The multivariate \( t \) distribution has heavy tails.
Bivariate Normal and $t$ 

$\rho = -0.7$, $\nu = 3$, variances all equal 1.
Fitted Normal and $t_3$ Distributions

Simulated data (2000) from models fitted by maximum likelihood to BMW-Siemens data.

©2008 (McNeil)
A2. Introduction to Copulas

- Copulas help in the understanding of dependence at a deeper level;
- They show us potential pitfalls of approaches to dependence that focus only on correlation;
- They allow us to define useful alternative dependence measures;
- They express dependence on a quantile scale, which is natural in QRM;
- They facilitate a bottom-up approach to multivariate model building;
- They are easily simulated and thus lend themselves to Monte Carlo risk studies.
What is a Copula?

A copula is a multivariate distribution function with standard uniform margins.

Equivalently, a copula if any function $C : [0, 1]^d \rightarrow [0, 1]$ satisfying the following properties:

1. $C(u_1, \ldots, u_d)$ is increasing in each component $u_i$.

2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0, 1]$.

3. For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0, 1]^d$ with $a_i \leq b_i$ we have:

\[
\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,
\]

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.
Probability and Quantile Transforms

**Lemma 1: probability transform**

Let $X$ be a random variable with **continuous** distribution function $F$. Then $F(X) \sim U(0, 1)$ (standard uniform).

$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \quad \forall u \in (0, 1).$$

**Lemma 2: quantile transform**

Let $U$ be uniform and $F$ the distribution function of any rv $X$. Then $F^{-1}(U) \overset{d}{=} X$ so that $P(F^{-1}(U) \leq x) = F(x)$.

These facts are the key to all statistical simulation and essential in dealing with copulas.
Sklar’s Theorem

Let $F$ be a joint distribution function with margins $F_1, \ldots, F_d$. There exists a copula $C$ such that for all $x_1, \ldots, x_d$ in $[-\infty, \infty]$:

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

If the margins are continuous then $C$ is unique; otherwise $C$ is uniquely determined on $\text{Ran} F_1 \times \text{Ran} F_2 \ldots \times \text{Ran} F_d$.

And conversely, if $C$ is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then $F$ defined above is a multivariate df with margins $F_1, \ldots, F_d$. 
Sklar’s Theorem: Proof in Continuous Case

Henceforth, unless explicitly stated, vectors $\mathbf{X}$ will be assumed to have continuous marginal distributions. In this case:

$$F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$$
$$= P(F_1(X_1) \leq F_1(x_1), \ldots, F_d(X_d) \leq F_d(x_d))$$
$$= C(F_1(x_1), \ldots, F_d(x_d)).$$

The unique copula $C$ can be calculated from $F, F_1, \ldots, F_d$ using

$$C(u_1, \ldots, u_d) = F\left(F_1^\leftarrow(u_1), \ldots, F_d^\leftarrow(u_d)\right).$$
Copulas and Dependence Structures

In the case of continuous margins, Sklar’s theorem shows how a unique copula $C$ describes in a sense the dependence structure of the multivariate distribution function $F$ of a random vector $X$. This unique copula is referred to as the copula of $F$ (or $X$).

**Invariance**

$C$ is invariant under strictly increasing transformations of the marginal distributions. If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ has the same copula as $(X_1, \ldots, X_d)$. 
The Fréchet Bounds

For every copula $C(u_1, \ldots, u_d)$ we have the important bounds

$$\max \left\{ \sum_{i=1}^{d} u_i + 1 - d, 0 \right\} \leq C(u) \leq \min \{u_1, \ldots, u_d\}.$$ \hspace{1cm} (17)

The upper bound represents perfect positive dependence or comonotonicity and is the copula of any random vector $(v_1(Z), \ldots, v_d(Z))$ where $Z$ is a random variable and $v_1, \ldots, v_d$ are increasing functions.

The lower bound represents perfect negative dependence or countermonotonicity and is the copula of $(v_1(Z), v_2(Z))$ where $v_1$ is increasing and $v_2$ decreasing, or vice versa.

The copula representing independence is $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$. 
Parametric families

- Gauss copula - implicit in multivariate Gaussian model using Sklar’s Theorem:

\[ C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \]

- t copula (implicit in multivariate Student model)

- generalized hyperbolic copulas (implicit in similarly named distributions)

- Archimedean copulas (implicit in simplex distributions)
Gauss and $t$ Copulas

Gaussian Copula

$$C_P^{Ga}(u) = \Phi_P \left( \Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d) \right),$$

where $\Phi$ denotes the standard univariate normal df, $\Phi_P$ denotes the joint df of $X \sim N_d(0, P)$ and $P$ is a correlation matrix. Write $C_{\rho}^{Ga}$ when $d = 2$.

$P = I_d$ gives independence and $P = J_d$ gives comonotonicity.

$t$ Copula

$$C_{\nu, P}^{t}(u) = t_{\nu, P} \left( t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_d) \right),$$

where $t_{\nu}$ is the df of a standard univariate $t$ distribution, $t_{\nu, P}$ is the joint df of the vector $X \sim t_d(\nu, 0, P)$ and $P$ is a correlation matrix. Write $C_{\nu, \rho}^{t}$ when $d = 2$.

$P = J_d$ gives comonotonicity, but $P = I_d$ does not give independence.
Archimedean Copulas

**Gumbel Copula**

\[ C_{\theta}^{\text{Gu}}(u_1, \ldots, u_d) = \exp \left( - \left( ( - \log u_1)^{\theta} + \cdots + ( - \log u_d)^{\theta} \right)^{1/\theta} \right). \]

\( \theta \geq 1: \theta = 1 \) gives independence; \( \theta \to \infty \) gives comonotonicity.

**Clayton Copula**

\[ C_{\theta}^{\text{Cl}}(u_1, \ldots, u_d) = \left( u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1 \right)^{-1/\theta}. \]

\( \theta > 0: \theta \to 0 \) gives independence; \( \theta \to \infty \) gives comonotonicity.
Meta-Distributions

By the converse of Sklar’s Theorem we know that if $C$ is a copula and $F_1, \ldots, F_d$ are univariate dfs, then $F(x) = C(F_1(x_1), \ldots, F_d(x_d))$ is a multivariate df with margins $F_1, \ldots, F_d$.

We refer to $F$ as a meta-distribution with the dependence structure represented by $C$. For example, if $C$ is a Gaussian copula we get a meta-Gaussian distribution and if $C$ is a t copula we get a meta-t distribution.
A3. Sampling Copulas

Simulating Gaussian copula $C_{P}^{Ga}$

- Simulate $\mathbf{X} \sim N_{d}(\mathbf{0}, \mathbf{P})$

- Set $\mathbf{U} = (\Phi(X_{1}), \ldots, \Phi(X_{d}))'$ (probability transformation)

Simulating $t$ copula $C_{\nu, P}^{t}$

- Simulate $\mathbf{X} \sim t_{d}(\nu, \mathbf{0}, \mathbf{P})$

- Set $\mathbf{U} = (t_{\nu}(X_{1}), \ldots, t_{\nu}(X_{d}))'$ (probability transformation)
  $t_{\nu}$ is df of univariate t distribution.

Simulation of Archimedean copulas is less obvious, but also turns out to be fairly simple in the majority of cases.
Simulating Copulas II

Gaussian

Gumbel

Clayton

t4

Gauss: $\rho = 0.7$, Gumbel: $\theta = 2$, Clayton: $\theta = 2.2$, t: $\rho = 0.71, \nu = 4$
By the converse of Sklar’s Theorem we know that if $C$ is a copula and $F_1, \ldots, F_d$ are univariate dfs, then $F(x) = C(F_1(x_1), \ldots, F_d(x_d))$ is a multivariate df with margins $F_1, \ldots, F_d$.

We refer to $F$ as a meta-distribution with the dependence structure represented by $C$. For example, if $C$ is a Gaussian copula we get a meta-Gaussian distribution and if $C$ is a t copula we get a meta-t distribution.

If we can sample from the copula $C$, then it is easy to sample from $F$: we generate a vector $(U_1, \ldots, U_d)$ with df $C$ and then return

$$(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)).$$
Simulating Meta Distributions

Linear correlation \( \rho(X_1, X_2) \approx 0.7 \) in all cases.
A4. Correlation

Denote the linear correlation of two random variables $X_1$ and $X_2$ by $\rho(X_1, X_2)$. We should be aware of the following.

- Linear correlation only gives a scalar summary of (linear) dependence and $\text{var}(X_1), \text{var}(X_2)$ must exist.

- $X_1, X_2$ independent $\Rightarrow \rho(X, Y) = 0$.
  But $\rho(X_1, X_2) = 0 \not\Rightarrow X_1, X_2$ independent.
  Example: spherical bivariate t-distribution with $\nu$ d.f.

- Linear correlation is not invariant with respect to strictly increasing transformations $T$ of $X_1, X_2$, i.e. generally

  $$\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2).$$
Correlation Confusion

“Among nine big economies, stock market correlations have averaged around 0.5 since the 1960s. In other words, for every 1 per cent rise (or fall) in, say, American share prices, share prices in the other markets will typically rise (fall) by 0.5 per cent.”

The Economist, 8 November 1997

“A correlation of 0.5 does not indicate that a return from stockmarket A will be 50% of stockmarket B’s return, or vice-versa...A correlation of 0.5 shows that 50% of the time the return of stockmarket A will be positively correlated with the return of stockmarket B, and 50% of the time it will not.”

The Economist (letter), 22 November 1997
Fallacies in the Use of Correlation

We consider random vectors \((X_1, X_2)\).

**Fallacy 1**

“Marginal distributions and correlation determine the joint distribution”.

- True for the class *bivariate* normal distributions (or more generally for elliptical distributions).
- Wrong in general, as we have seen.
Fallacy 1 continued

Sometimes Fallacy 1 is hidden in statements like: “If two random variables $X_1$ and $X_2$ are uncorrelated, they may be considered as approximately independent”.

Consider two portfolios of risks. Set

$$X_1 = Z \quad \text{(Profit&Loss Country A)},$$
$$X_2 = V \cdot Z \quad \text{(Profit&Loss Country B)},$$

$V, Z$ independent, $Z \sim N(0, 1)$,

$$P(V = +1) = P(V = -1) = 1/2.$$

$V$ switches between perfect positive and negative dependence.

$$X_2 \sim N(0, 1) \quad \text{and} \quad \rho(X_1, X_2) = 0.$$

But $(X_1, X_2)'$ is not bivariate normal.
VaR (Quantile) for two different dependence models

\[ \text{VaR}_\alpha (X_1 + X_2) \] for \( X_1, X_2 \) independent and \( X_1, X_2 \) dependent.
Fallacy 2

“Given marginal distributions $F_1$ and $F_2$ for $X_1$ and $X_2$, all linear correlations between -1 and +1 can be attained through specification of the joint distribution”.

- This is again true for elliptical distributions but not true in general. If $F_1$ and $F_2$ are not of the same type, then $\rho(X_1, X_2) < 1$.

- **Theorem** (Höffding 1940)

  1. The set of possible correlations is a closed interval $[\rho_{\text{min}}, \rho_{\text{max}}]$.
  2. $\rho_{\text{max}}$ is attained iff $X_1, X_2$ comonotonic. $\rho_{\text{min}}$ is attained iff $X_1, X_2$ countermonotonic.
Example of Attainable Correlations

Take $X_1 \sim \text{Lognormal}(0, 1)$, and $X_2 \sim \text{Lognormal}(0, \sigma^2)$. Observe how interval of attainable correlations varies with $\sigma$. Upper boundary represents comonotonicity. See [McNeil et al., 2005] for details.
Additivity of VaR for Comonotonic Risks

Let $0 < \alpha < 1$ and $L_1, \ldots, L_d$ be comonotonic random variables (so that pairwise correlations are maximal). Then

$$\text{VaR}_\alpha(L_1 + \cdots + L_d) = \text{VaR}_\alpha(L_1) + \cdots + \text{VaR}_\alpha(L_d)$$

It is clear from its definition (2) that expected shortfall shares this property of comonotonic additivity.
Fallacy 3

“VaR for the sum of two risks is at its worst when these two risks have maximal correlation, i.e. are comonotonic”.

- A correlation fallacy or a Value-at-Risk fallacy?

- True when the risks have a bivariate normal, or elliptical distribution.

- Any superadditive VaR example shows that it is wrong in general, since in that case we have

  \[ \text{VaR}_\alpha(L_1 + L_2) > \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2) \]

  for some risks \( L_1 \) and \( L_2 \) and some \( \alpha \). But the right hand side is the VaR of \( (L_1 + L_2) \) when the risks are comonotonic and corresponds to the maximal correlation case.
Distributions with the same correlation may show quite different dependence in the tails and this is a matter of concern for quantitative risk managers.

Some distributions show strong tail dependence - the tendency for joint extreme values to fall together. Use of a model with strong tail dependence is one way of capturing the phenomenon that correlations tend to go to one in times of stress.

Generally a model with strong tail dependence can be understood in terms of extreme values in a common risk driver.
Tail Dependence Corefficients

When limit exists, coefficient of upper tail dependence is

$$\lambda_u(X_1, X_2) = \lim_{q \to 1} P(X_2 > F_2^{-}(q) \mid X_1 > F_1^{-}(q)),$$

Analogously the coefficient of lower tail dependence is

$$\lambda_l(X_1, X_2) = \lim_{q \to 0} P(X_2 \leq F_2^{-}(q) \mid X_1 \leq F_1^{-}(q)).$$

These are functions of the copula given by

$$\lambda_u = \lim_{q \to 1} \frac{\overline{C}(q, q)}{1 - q} = \lim_{q \to 1} \frac{1 - 2q + C(q, q)}{1 - q},$$

$$\lambda_l = \lim_{q \to 0} \frac{C(q, q)}{q}.$$
Tail Dependence

Clearly $\lambda_u \in [0, 1]$ and $\lambda_l \in [0, 1]$.

For copulas of elliptically symmetric distributions $\lambda_u = \lambda_l =: \lambda$. This is true, more generally, for all copulas with radial symmetry.

Terminology:

$\lambda_u \in (0, 1]$: upper tail dependence,

$\lambda_u = 0$: asymptotic independence in upper tail,

$\lambda_l \in (0, 1]$: lower tail dependence,

$\lambda_l = 0$: asymptotic independence in lower tail.
Examples of Tail Dependence

The Gaussian copula is asymptotically independent for \(|\rho| < 1\).

The t copula is tail dependent when \(\rho > -1\).

\[
\lambda = 2 \bar{t}_{\nu+1} \left( \sqrt{\nu + 1} \sqrt{1 - \rho / \sqrt{1 + \rho}} \right).
\]

The Gumbel copula is upper tail dependent for \(\theta > 1\).

\[
\lambda_u = 2 - 2^{1/\theta}.
\]

The Clayton copula is lower tail dependent for \(\theta > 0\).

\[
\lambda_l = 2^{-1/\theta}.
\]

All formulas are derived in the book.
Gaussian and t3 Copulas Compared

Normal Dependence

$t$ Dependence

Copula parameter $\rho = 0.7$; quantiles lines 0.5% and 99.5%.
## Joint Tail Probabilities at Finite Levels

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>C</th>
<th>95%</th>
<th>99%</th>
<th>99.5%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>N</td>
<td>$1.21 \times 10^{-2}$</td>
<td>$1.29 \times 10^{-3}$</td>
<td>$4.96 \times 10^{-4}$</td>
<td>$5.42 \times 10^{-5}$</td>
</tr>
<tr>
<td>0.5</td>
<td>t8</td>
<td>1.20</td>
<td>1.65</td>
<td>1.94</td>
<td>3.01</td>
</tr>
<tr>
<td>0.5</td>
<td>t4</td>
<td>1.39</td>
<td>2.22</td>
<td>2.79</td>
<td>4.86</td>
</tr>
<tr>
<td>0.5</td>
<td>t3</td>
<td>1.50</td>
<td>2.55</td>
<td>3.26</td>
<td>5.83</td>
</tr>
<tr>
<td>0.7</td>
<td>N</td>
<td>$1.95 \times 10^{-2}$</td>
<td>$2.67 \times 10^{-3}$</td>
<td>$1.14 \times 10^{-3}$</td>
<td>$1.60 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.7</td>
<td>t8</td>
<td>1.11</td>
<td>1.33</td>
<td>1.46</td>
<td>1.86</td>
</tr>
<tr>
<td>0.7</td>
<td>t4</td>
<td>1.21</td>
<td>1.60</td>
<td>1.82</td>
<td>2.52</td>
</tr>
<tr>
<td>0.7</td>
<td>t3</td>
<td>1.27</td>
<td>1.74</td>
<td>2.01</td>
<td>2.83</td>
</tr>
</tbody>
</table>

For normal copula probability is given.
For $t$ copulas the factor by which Gaussian probability must be multiplied is given.
### Joint Tail Probabilities, $d \geq 2$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>C</th>
<th>Dimension $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>0.5</td>
<td>N</td>
<td>$1.29 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.5</td>
<td>t8</td>
<td>1.65</td>
</tr>
<tr>
<td>0.5</td>
<td>t4</td>
<td>2.22</td>
</tr>
<tr>
<td>0.5</td>
<td>t3</td>
<td>2.55</td>
</tr>
<tr>
<td>0.7</td>
<td>N</td>
<td>$2.67 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.7</td>
<td>t8</td>
<td>1.33</td>
</tr>
<tr>
<td>0.7</td>
<td>t4</td>
<td>1.60</td>
</tr>
<tr>
<td>0.7</td>
<td>t3</td>
<td>1.74</td>
</tr>
</tbody>
</table>

We consider only 99% quantile and case of equal correlations.
Financial Interpretation

Consider daily returns on five financial instruments and suppose that we believe that all correlations between returns are equal to 50%. However, we are unsure about the best multivariate model for these data.

If returns follow a multivariate Gaussian distribution then the probability that on any day all returns fall below their 1% quantiles is $7.48 \times 10^{-5}$. In the long run such an event will happen once every 13369 trading days on average, that is roughly once every 51.4 years (assuming 260 trading days in a year).

On the other hand, if returns follow a multivariate t distribution with four degrees of freedom then such an event will happen 7.68 times more often, that is roughly once every 6.7 years.
A6. Rank Correlation

Like coefficients of tail dependence, rank correlations are simple measures of dependence that depend only on the copula of the risk factors. They are mainly useful for the calibration of copulas to empirical data.
Rank Correlation

Spearman’s rho

\[ \rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = \rho(\text{copula}) \]

\[ \rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 \{C(u_1, u_2) - u_1 u_2\} du_1 du_2. \]

Kendall’s tau

Take an independent copy of \((X_1, X_2)\) denoted \((\tilde{X}_1, \tilde{X}_2)\).

\[ \rho_\tau(X_1, X_2) = 2P\left( (X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0 \right) - 1 \]

\[ \rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1. \]
Properties of Rank Correlation

The following statements are true for Spearman’s rho ($\rho_S$) or Kendall’s tau ($\rho_\tau$), but not for Pearson’s linear correlation ($\rho$).

- $\rho_S$ depends only on copula of $(X_1, X_2)$.

- $\rho_S$ is invariant under strictly increasing transformations of the random variables.

- $\rho_S(X_1, X_2) = 1 \iff X_1, X_2$ comonotonic.

- $\rho_S(X_1, X_2) = -1 \iff X_1, X_2$ countermonotonic.
Sample Rank Correlations

Consider iid bivariate data \( \{(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2})\} \). The standard estimator of \( \rho_\tau(X_1, X_2) \) is

\[
\frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sgn} \left[ (X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) \right],
\]

and the estimator of \( \rho_S(X_1, X_2) \) is

\[
\frac{12}{n(n^2 - 1)} \sum_{i=1}^{n} \left( \text{rank}(X_{i,1}) - \frac{n + 1}{2} \right) \left( \text{rank}(X_{i,2}) - \frac{n + 1}{2} \right).
\]
B. Modular Approach to Risk Integration

1. Aggregation and Diversification

2. The Modular Approach

3. The Modular Approach With Copulas
B1. Aggregation and Diversification

“Solvency I fails to recognise diversification benefits properly even though they are fundamental to value creation in the insurance industry and contribute to improved efficiency of insurance service provision, greater stability in financial performance which in turn contributes to policyholder protection, and a more efficient allocation of capital in the economy.” [Treasury and FSA, 2006]

The pooling of risks across portfolios, business lines, organisations achieves diversification. The extent of the diversification benefit depends on the degree of dependence between the pooled risks. Aggregate solvency capital should reflect the diversification benefit.
Layers of Aggregation

In [Kuritzkes et al., 2002] three levels of aggregation are identified:

1. stand-alone risks within a single risk factor (e.g. underwriting risk in each contract of a domestic motor portfolio);

2. different risk factors within a single business line (e.g. combining asset, underwriting and operational risks in non-life or life insurance);

3. different business lines within an enterprise.
Mathematical Framework

An enterprise may be split into $d$ sub-units (business lines, risk factors by business line, contracts/investments). Each sub-unit generates a loss or a (negative) change-in-value $L_i$ over the time horizon of interest. The aggregate change-in-value distribution is given by

$$L = L_1 + \cdots + L_d.$$ 

**Ideal goal:**

Determination of risk capital should be based on a stochastic model for $(L_1, \ldots, L_d)$ that accurately reflects the dependence structure.
Diversification and Correlation

“Diversification benefits can be assessed by correlations between different risk categories. A correlation of +100% means that two variables will fall and rise in lock-step; any correlation below this indicates the potential for diversification benefits.” [Treasury and FSA, 2006]

The last statement is not true of ordinary linear (Pearson) correlation! But true of rank correlation.

Lock-step

The mathematical term for this is comonotonicity. It means all risks are increasing functions of a common underlying risk: 
\[(L_1, \ldots, L_d) = (v_1(Z), \ldots, v_d(Z)).\] Such risks would be considered undiversifiable.
Comonotonicity and Correlation

(linear correlation = 1) \Rightarrow \text{ comonotonicity}

comonotonicity \nRightarrow (linear correlation = 1)

• We can create models where individual risks move in lock-step (are undiversifiable), but have an arbitrarily small correlation.

• For two given distributions, \textit{attainable correlations} form a sub-interval of \([-1, 1]\).

• Upper bound corresponds to comonotonicity, lower to countermonotonicity (negative lock-step)

• Our intuition about linear correlation is in fact very faulty!
Individual risks (sub-units) are transformed into capital charges $EC_1, \ldots, EC_d$. These are then combined to calculate the overall solvency capital requirement $EC$. The combination operation generally involves a calculation of the following kind:

$$EC = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} EC_i EC_j}$$

where the $\rho_{ij}$ are the “correlations” between the risks.

The standard formula for the solvency capital requirement adopts a modular approach [CEIOPS-06, 2006], page 71, 98. Standard & Poors Risk-Adjusted Capital Framework also adopts such an approach.
Where is the Principle in This?

Suppose

• we use the Value-at-Risk measure to set capital, so that $EC_i = \text{VaR}_\alpha(L_i)$ and $EC = \text{VaR}_\alpha(L)$ where $\alpha > 0.5$;

• the risks $(L_1, \ldots, L_d)$ are jointly normal with zero mean and correlations given by $\rho_{ij}$.

More generally, we could consider any positive-homogeneous risk measure (such as expected shortfall) in first assumption and any centred elliptical distribution (such as multivariate Student t) in second.
Short Derivation of Aggregation Rule

\[ \text{sd}(L) = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \text{sd}(L_i) \text{sd}(L_j)} \]

Now \( \text{VaR}_\alpha(L) = \lambda_\alpha \text{sd}(L) \) and \( \text{VaR}_\alpha(L_i) = \lambda_\alpha \text{sd}(L_i) \) where \( \lambda_\alpha \) is the \( \alpha \)-quantile of standard normal. This yields

\[ \text{VaR}_\alpha(L) = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \text{VaR}_\alpha(L_i) \text{VaR}_\alpha(L_j)} \]

\[ \text{EC} = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \text{EC}_i \text{EC}_j} \]
Issues with this style of aggregation

• It is only underpinned by theoretical *principles* in a very specific and unrealistic model of the risk universe.

• It is dependent on the widely misunderstood concept of correlation.

• The kinds of risks where we have reliable empirical experience of typical values are in the minority (e.g. financial market risks, and even then only at shorter time horizons)

• Can we trust “experts” to deliver correlations in other cases? There are consistency requirements: every $\rho_{ij}$ should be compatible with the distribution of $L_i$ and $L_j$. The matrix $(\rho_{ij})$ must be positive semi-definite. It is quite easy to specify nonsensical correlation matrices.
"Further analysis is required to assess whether linear correlation, together with a simplified form of tail correlation may be a suitable technique to aggregate capita requirements for different risks." [CEIOPS-06, 2006] (page 75)

“When selecting correlation coefficients, allowance should be made for tail correlation. To allow for this, the correlations used should be higher than simple analysis of relevant data would indicate.” [CEIOPS-06, 2006] (page 142)
Is the sum of capital charges an upper bound?

Suppose again that we wish to employ Value-at-Risk to set capital, so that $EC_i = \text{VaR}_{\alpha}(L_i)$ and $EC = \text{VaR}_{\alpha}(L)$.

In the case where we have no diversification (comonotonic risks $L_i = u_i(Z), i = 1, \ldots, d$) we can compute that

$$EC = \sum_{i=1}^{d} EC_i$$

Fallacy:

“this is an upper bound for the solvency capital requirement under all dependence assumptions for $(L_1, \ldots, L_d)$.”
Superadditive Capital

Actually, it is possible to construct models for \( (L_1, \ldots, L_d) \) where superadditivity occurs for particular \( \alpha \) values:

\[
\text{VaR}_\alpha(L) > \sum_{i=1}^{d} \text{VaR}_\alpha(L_i).
\]

- To rectify this problem we would have to base risk measurement and capital charges on a subadditive risk measure (like expected shortfall).

- Many argue that the models leading to superadditivity are too implausible to consider, but they do undermine our principles!
B3. The Modular Approach with Copulas

• **Copulas** are a better theoretical tool for combining the individual capital charges. They avoid tricky consistency requirements imposed by working with linear correlations.

• Implicitly aggregation based on the **Gauss copula** has been used in insurance for years. For example @RISK by Palisade software implicitly uses the Gauss copula to perform Monte Carlo risk analysis.

• However, **calibration** remains a problem. Copula parameters are usually inferred from matrices of rank correlations, but are we expert enough to set these?

• Bottom-up approaches require the exogenous specification of parameters determining the dependence model.
Rank Correlations for Gauss and t Copula

Let \( X \) be a bivariate random vector with copula \( C^\text{Ga}_\rho \) and continuous margins. Then the rank correlations are

\[
\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \rho, \quad (18)
\]

\[
\rho_S(X_1, X_2) = \frac{6}{\pi} \arcsin \frac{\rho}{2}. \quad (19)
\]

Normal Variance Mixture Case

The formula (18) also holds when \( X \) has the copula of a normal variance mixture distribution with correlation parameter \( \rho \), for example the \( t \) copula \( C^t_{\nu, \rho} \).
Calibrating Gauss copula with Spearman’s rho

Suppose we assume a meta-Gaussian model for $X$ with copula $C^\text{Ga}_P$ and we wish to estimate the correlation matrix $P$. It follows from Theorem 5.36 in book that

$$\rho_S(X_i, X_j) = \frac{6}{\pi} \arcsin \frac{\rho_{ij}}{2} \approx \rho_{ij},$$

where the final approximation is very accurate. This suggests we estimate $P$ by the matrix of pairwise Spearman’s rank coefficients $R^S$. This is essentially the @RISK approach.
Calibrating t Copula with Kendall’s tau

Suppose we assume a meta t model for $\mathbf{X}$ with copula $C_{\nu, P}^t$ and we wish to calibrate the correlation matrix $P$. The theoretical relationship between Spearman’s rho and $P$ is not known in this case, but a relationship between Kendall’s tau and $P$ is known.

It follows from Proposition 5.37 in book that

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi} \arcsin \rho_{ij},$$

so that a possible estimator of $P$ is the matrix $R^*$ with components given by $r_{ij}^* = \sin(\pi r_{ij}^\tau / 2)$ This may not be positive definite, in which case $R^*$ can be transformed by the eigenvalue method (Algorithm 5.55 in MFE) to obtain a positive definite matrix that is close to $R^*$. 

©2008 (McNeil)
C. The Fully Integrated Approach

1. The Approach

2. Reducing Dimensionality With Factor Models

3. Rare Events and Importance Sampling
C1. The Approach

In a top-down approach the correlations are endogenous and result from specifying the mutual dependence of risks across the enterprise on common risk drivers or factors.

\[ L_i = f_i(\text{common factors, idiodyncratic factors}), \quad i = 1, \ldots, d. \]

These models are invariably handled by Monte Carlo, i.e. the random generation of scenarios for the factors.

They appeal because they are structural and explanatory.
Advantages

- Fully integrated models are much more “principles-based” than modular approaches.

- A natural framework for risk-based allocation of capital to business units which opens door to risk-based performance measurement (RORAC).

- A framework for actual computation of the diversification benefit and attribution of that benefit to sub-units.

- Framework for sensitivity analyses with respect to common factors and model risk studies with respect to model assumptions.

- Tail dependence may be studied in terms of extreme outcomes in key risk drivers.
The conclusions about capital adequacy and risk-based performance comparison are only as good as the underlying models, which need to be built by skilled craftsmen. The biggest issue is the sensitivity of the results to the model inputs, in particular the model components specifying the dependence of risks on common factors.

Seemingly innocuous assumptions about correlations can have large effects.

Consider again the following example from credit risk. By adding a common factor that induces a default correlation of 0.005 between every pair of counterparties, we inflate tail of loss distribution.
Impact of Dependence on Credit Loss Distribution

Comparison of the loss distribution of a homogeneous portfolio of 1000 loans with a default probability of 1% assuming (i) independent defaults and (ii) a default correlation of 0.5%.

©2008 (McNeil)
Technical Considerations

Fully integrated models may be built in all kinds of different ways. Decisions with respect to which factors to use and how to specify model components for particular risk types will vary between implementations.

However there are general issues that are likely to arise in any implementation:

- **Dimension reduction.** Among the universe of available risk factors, which smaller subset should we choose? It is impractical (and generally bad modelling practice) to try to include all possible factors. Parsimonious models are generally more useful tools.

- **Efficiency of Monte Carlo simulation.** How do we keep the simulation time to a minimum? The problem arises because we want to use the model to determine capital based on rare events.
C2. Dimension Reduction and Factor Models

**Idea:** Explain the variability in a $d$-dimensional vector $X$ in terms of a smaller set of common factors.

**Definition:** $X$ follows a $p$-factor model if

$$X = a + BF + \varepsilon,$$  \hspace{1cm} (20)

(i) $F = (F_1, \ldots, F_p)'$ is random vector of factors with $p < d,$
(ii) $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d)'$ is random vector of idiosyncratic error terms, which are uncorrelated and mean zero,
(iii) $B \in \mathbb{R}^{d \times p}$ is a matrix of constant factor loadings and $a \in \mathbb{R}^d$ a vector of constants,
(iv) $\text{cov}(F, \varepsilon) = E((F - E(F))\varepsilon') = 0.$
Remarks on Theory of Factor Models

• Factor model (20) implies that covariance matrix \( \Sigma = \text{cov}(X) \) satisfies \( \Sigma = BB' + \Psi \), where \( \Omega = \text{cov}(F) \) and \( \Psi = \text{cov}(\varepsilon) \) (diagonal matrix).

• Factors can always be transformed so that they are orthogonal:

\[
\Sigma = BB' + \Psi. \tag{21}
\]

• Conversely, if (21) holds for covariance matrix \( \Sigma \) of random vector \( X \), then \( X \) follows factor model (20) for some \( a, F \) and \( \varepsilon \).

• If, moreover, \( X \) is Gaussian then \( F \) and \( \varepsilon \) may be taken to be independent Gaussian vectors, so that \( \varepsilon \) has independent components.
Factor Models in Practice

We have multivariate financial return data $X_1, \ldots, X_n$ which are assumed to follow (20). Two situations to be distinguished:

1. Appropriate factor data $F_1, \ldots, F_n$ are also observed, for example returns on relevant indices. We have a multivariate regression problem; parameters ($\alpha$ and $B$) can be estimated by multivariate least squares.

2. Factor data are not directly observed. We assume data $X_1, \ldots, X_n$ identically distributed and calibrate factor model by one of two strategies: statistical factor analysis - we first estimate $B$ and $\Psi$ from (21) and use these to reconstruct $F_1, \ldots, F_n$; principal components - we fabricate $F_1, \ldots, F_n$ by PCA and estimate $B$ and $\alpha$ by regression.
In a generic Monte Carlo problem we have a random variable $X$ with density $f$ and we wish to compute an expected value of the form

$$\theta = E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx,$$

for some known function $h$. For VaR estimation we essentially consider $h$-functions of form $h(x) = 1\{x \geq c\}$ for some set $c \in \mathbb{R}$; for expected shortfall computation we consider $h(x) = x1\{x \geq c\}$.

Where the analytical evaluation of $\theta$ is difficult we can resort to a Monte Carlo (MC) approach:

1. Simulate $X_1, \ldots, X_n$ independently from density $f$.

2. Compute the standard MC estimate $\hat{\theta}_n^{MC} = \frac{1}{n} \sum_{i=1}^{n} h(X_i)$. 

C3. Rare Events and Importance Sampling
Rare Event Simulation

The MC estimator converges to $\theta$ by the strong law of large numbers, (SLLN) but the speed of convergence may not be particularly fast, particularly when we are dealing with rare event simulation.

If we consider estimating $\text{VaR}_{0.99}$, for example, then we have to consider a range of $c$ values such that only around 1\% of our standard Monte Carlo draws lead to a larger portfolio loss. For estimating $\text{ES}_{0.99}$ we have to first estimate $\text{VaR}_{0.99}$ and then compute averages of the rare Monte Carlo simulations that lead to losses larger than $\text{VaR}_{0.99}$. The standard MC estimators will be unstable and subject to high variability, unless the number of simulations is very large.

The technique of importance sampling is a way of reducing this variability and is well suited to problems of the kind we consider.
Importance Sampling

Importance sampling is based on an alternative representation of (22). We consider an importance sampling density \( g \) (whose support should contain that of \( f \)) and define the likelihood ratio \( r(x) \) by
\[
r(x) := f(x)/g(x) \quad \text{whenever } g(x) > 0 \text{ and } r(x) = 0 \quad \text{otherwise.}
\]
The integral may be written in terms of the likelihood ratio as
\[
\theta = \int_{-\infty}^{\infty} h(x) r(x) g(x) \, dx = E_g(h(X) r(X)),
\]
where \( E_g \) denotes expectation with respect to the density \( g \). Hence we can approximate the integral with the following algorithm.

1. Simulate \( X_1, \ldots, X_n \) independently from density \( g \).

2. Compute the IS estimate \( \hat{\theta}^{\text{IS}}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i) r(X_i) \).
Reducing the Variance

The art of importance sampling is in choosing $g$ such that for fixed $n$ the variance of the IS estimator is considerably smaller than that of the standard Monte Carlo estimator.

$$ \text{var}_g \left( \hat{\theta}^{\text{IS}}_n \right) = \frac{1}{n} \left( E_g \left( h(X)^2 r(X)^2 \right) - \theta^2 \right), $$

$$ \text{var} \left( \hat{\theta}^{\text{MC}}_n \right) = \frac{1}{n} \left( E \left( h(X)^2 \right) - \theta^2 \right). $$

The aim is to make $E_g(h(X)^2 r(X)^2)$ small compared to $E(h(X)^2)$.

Consider the case of estimating a tail probability where $h(x) = 1\{x \geq c\}$ for $c$ significantly larger than the mean of $X$. We try to choose $g$ so that the likelihood ratio $r(x) = f(x)/g(x)$ is small for $x \geq c$; in other words we make the event $\{X \geq c\}$ more likely under the IS density $g$ than it is under the original density $f$. 
Exponential Tilting

For $t \in \mathbb{R}$ we write $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$ for the moment generating function of $X$, which we assume is finite for $t \in \mathbb{R}$. It is not hard to check that we can define a density $g_t(x) := e^{tx} f(x)/M_X(t)$ which can be used for importance sampling when $X$ is light tailed.

Define $\mu_t$ to be the mean of $X$ with respect to the density $g_t$ i.e.

$$
\mu_t := E_{g_t}(X) = E(X \exp(tX)/M_X(t)).
$$

How can we choose $t$ optimally for a particular importance sampling problem? In the case of tail probability estimation theory suggests we should choose $t$ as the solution of $\mu_t = c$. 
Exponential Tilting For Normal Distribution

We illustrate the concept of exponential tilting in the simple case of a standard normal random variable. Suppose $X \sim N(0, 1)$ with density $\phi(x)$. Using exponential tilting we obtain the new density $g_t(x) = \exp(tx)\phi(x)/M_X(t)$. The moment generating function of $X$ is known to be $M_X(t) = \exp(t^2/2)$. Hence

$$g_t(x) = \frac{1}{\sqrt{2\pi}} \exp \left( tx - \frac{1}{2}(t^2 + x^2) \right) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{(x - t)^2}{2} \right),$$

so that under the tilted distribution, $X \sim N(t, 1)$. In particular, exponential tilting is a convenient way of shifting the mean of $X$. 
An Abstract View of Importance Sampling

To handle applications of IS in portfolio credit risk (where discrete Bernoulli distributions appear) we consider importance sampling from a slightly more general viewpoint. Given densities $f$ and $g$ we define probability measures $P$ and $Q$ by

$$P(A) = \int_A f(x)\,dx \quad \text{and} \quad Q(A) = \int_A g(x)\,dx, \; A \subset \mathbb{R}.$$ 

With this notation (23) becomes

$$\theta = E_P(h(X)) = E_Q(h(X)r(X)),$$

so that $r(X)$ equals $dP/dQ$, the (measure-theoretic) density of $P$ with respect to $Q$. 
More General View of Exponential Tilting

Using this more abstract view, exponential tilting can be applied in more general situations: given a rv $X$ on $(\Omega, \mathcal{F}, P)$ such that $M_X(t) = E_P(\exp(tX)) < \infty$, define the measure $Q_t$ on $(\Omega, \mathcal{F})$ by

$$
\frac{dQ_t}{dP} = \frac{\exp(tX)}{M_X(t)}, \quad \text{i.e.} \quad Q_t(A) = E_P\left(\frac{\exp(tX)}{M_X(t)}; A\right),
$$

and note that $(dQ_t/dP)^{-1} = M_X(t) \exp(-tX) = r_t(X)$.

The IS algorithm remains essentially unchanged: simulate independent realizations $X_i$ under the measure $Q_t$ and set $\hat{\theta}^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} X_i r_t(X_i)$ as before.
D. Capital Allocation

1. Allocation

2. Monte Carlo Approach to Shortfall Contributions

3. Diversification Scoring
D1. Capital Allocation

Consider an investor who can invest in a fixed set of $d$ investment possibilities with losses represented by the random variables $L_1, \ldots, L_d$. We have the following economic interpretations:

- **Performance measurement.** Here the investor is a financial institution and the $L_i$ represent the (negative of the) P&L of $d$ different lines of business.

- **Loan pricing.** Here the investor is a loan book manager responsible for a portfolio of $d$ loans.

- **General investment.** Here we consider either an individual or institutional investor and the standard interpretation that the $L_i$ are (negative) P&Ls corresponding to a set of investments in various assets.
The performance of the different business units or investments is usually measured using some sort of RORAC (return on risk-adjusted capital) approach, i.e. by considering a ratio of the form

\[
\frac{\text{expected return}}{\text{risk capital}}
\]

where we leave the precise definition of the terms vague; in many applications risk capital might correspond to economic capital — the capital derived by considering the fluctuation of the loss around the expected loss (the unexpected loss) rather than the absolute loss.
The General Procedure

Obviously the RORAC approach raises the question of what the appropriate risk capital for an individual investment opportunity might be. Thus the question of performance of the investment is intimately connected with the subject of risk measurement. A two-step procedure is used in practice:

1. Compute the overall risk capital $\varrho(L)$ where $L = \sum_{i=1}^{d} L_i$ and $\varrho$ is a particular risk measure such as VaR or ES; note that at this stage we are not stipulating that $\varrho$ must be coherent.

2. Allocate the capital $\varrho(L)$ to the individual investment possibilities according to some mathematical capital allocation principle such that, if $EC_i$ denotes the capital allocated to the investment with potential loss $L_i$, the sum of the allocated amounts corresponds to the overall risk capital $\varrho(L)$. 
We are interested in second step of the procedure; loosely speaking we require a mapping that takes as input the individual losses $L_1, \ldots, L_d$ and the risk measure $\varrho$ and yields as output the vector $(EC_1, \ldots, EC_d)$ such that

\[ \varrho(L) = \sum_{i=1}^{d} EC_i \]  

and such a mapping will be called a capital allocation principle. The relation (24) is sometimes called the full allocation property since all of the overall risk capital $\varrho(L)$ (not more, not less) is allocated to the investment possibilities; we consider this property to be an integral part of the definition of an allocation principle.
The Set-Up

For our discussion it will be useful to consider portfolios where the weights of the individual investment opportunities are varied with respect to our basic portfolio \((L_1, \ldots, L_d)\) which is regarded as a fixed random vector. That is, we consider an open set \(\Lambda \subset \mathbb{R}^d \setminus \{0\}\) of portfolio weights and define for \(\lambda \in \Lambda\) the loss \(L(\lambda) = \sum_{i=1}^{d} \lambda_i L_i\); the loss of our actual portfolio is of course \(L(1)\).

Let \(\varrho\) be some risk measure defined on a set \(\mathcal{M} \supseteq \{L(\lambda) : \lambda \in \Lambda\}\). We then define the associated risk measure function \(r_\varrho : \Lambda \to \mathbb{R}\) by \(r_\varrho(\lambda) = \varrho(L(\lambda))\). Thus \(r_\varrho(\lambda)\) is the required risk capital for a position \(\lambda\) in the set of investment possibilities.
Capital Allocation Principle: Definition

Let \( r_\varrho \) be a risk measure function on some set \( \Lambda \subset \mathbb{R}^d \setminus \{0\} \) such that \( 1 \in \Lambda \). A mapping \( \pi^{r_\varrho} : \Lambda \rightarrow \mathbb{R}^d \) is called a (per-unit) capital allocation principle associated with \( r_\varrho \) if for all \( \lambda \in \Lambda \) we have

\[
\sum_{i=1}^{d} \lambda_i \pi^{r_\varrho}_i(\lambda) = r_\varrho(\lambda). \tag{25}
\]

The interpretation of this definition is that \( \pi^{r_\varrho}_i \) gives the amount of capital allocated to one unit of \( L_i \), when the overall position has loss \( L(\lambda) \). The amount of capital allocated to the position \( \lambda_i L_i \) is thus \( \lambda_i \pi^{r_\varrho}_i \) and the equality (25) simply means that the overall risk capital \( r_\varrho(\lambda) \) is fully allocated to the individual portfolio positions.
Euler Principle

Now consider risk measures that are positive homogeneous (coherent risk measures, VaR, standard deviation) so that $r_\varrho(t\lambda) = tr_\varrho(\lambda)$ for all $t > 0, \lambda \in \Lambda$. In other words $r_\varrho : \Lambda \rightarrow \mathbb{R}$ is a positive homogeneous function of a vector argument.

Euler’s well-known rule which says that, if $r_\varrho$ is differentiable at $\lambda \in \Lambda$, we have

$$r_\varrho(\lambda) = \sum_{i=1}^{d} \lambda_i \frac{\partial r_\varrho}{\partial \lambda_i}(\lambda).$$

Comparison with (25) suggests we define the (per-unit) Euler capital allocation principle associated with $r_\varrho$ to be

$$\pi^{r_\varrho} : \Lambda \rightarrow \mathbb{R}^d, \quad \pi^{r_\varrho}_i(\lambda) = \frac{\partial r_\varrho}{\partial \lambda_i}(\lambda). \quad (26)$$

Obviously the Euler principle gives a full allocation of the risk capital.
Standard Deviation and the Covariance Principle

Suppose \( r_{SD}(\lambda) = \sqrt{\text{var}(L(\lambda))} \) and write \( \Sigma \) for the covariance matrix of \((L_1, \ldots, L_d)\). Then we have \( r_{SD}(\lambda) = (\lambda'\Sigma\lambda)^{1/2} \) from which it follows that

\[
\pi_i r_{SD}(\lambda) = \frac{\partial r_{SD}(\lambda)}{\partial \lambda_i} = \frac{(\Sigma\lambda)_i}{r_{SD}(\lambda)} = \frac{\sum_{j=1}^d \text{cov}(L_i, L_j)\lambda_j}{r_{SD}(\lambda)} = \frac{\text{cov}(L_i, L(\lambda))}{\sqrt{\text{var}(L(\lambda))}}.
\]

In particular, for the original portfolio of investment possibilities corresponding to \( \lambda = 1 \), the capital allocated to the \( i \)th investment possibility is

\[
\text{EC}_i = \pi_i r_{SD}(1) = \frac{\text{cov}(L_i, L)}{\sqrt{\text{var}(L)}}, \quad L := L(1).
\]

which is known as the covariance principle.
Suppose $r_{\text{VaR}}^\alpha(\lambda) = q_\alpha(L(\lambda))$. In that case it can be shown that, subject to technical conditions,

$$
\pi_i^{r_{\text{VaR}}^\alpha}(\lambda) = \frac{\partial r_{\text{VaR}}^\alpha}{\partial \lambda_i}(\lambda) = E (L_i \mid L(\lambda) = q_\alpha(L(\lambda))) , \quad 1 \leq i \leq d.
$$

A suitable technical condition is that $(L_1, \ldots, L_d)$ has a joint density, although this can be weakened.

We obtain a capital allocation of the form

$$
EC_i = E (L_i \mid L = \text{VaR}_\alpha(L)) , \quad L := L(1),
$$

where $EC_i$ is known as the VaR contribution of investment possibility (or line of business) $i$. 
Expected Shortfall and Shortfall Contributions

Now consider using the risk measure function

\[ r_{ES}^{\alpha}(\lambda) = E (L \mid L \geq q_{\alpha}(L(\lambda))) \]

corresponding to expected shortfall.

Subject to the same technical conditions, this gives a capital allocation of the form

\[ EC_i = E (L_i \mid L \geq \text{VaR}_{\alpha}(L)), \quad L := L(1), \tag{27} \]

where \( EC_i \) is known as the expected shortfall contribution of investment possibility (or line of business) \( i \). This is a popular allocation principle in practice, and is generally considered preferable to the covariance principle and the principle based on VaR contributions.
Example

In general the three risk measures (std. dev., VaR, ES) and their corresponding Euler allocations will lead to the allocation of different amounts of capital to the investment opportunities. This depends on the distribution of $L = (L_1, \ldots, L_d)'$.

Suppose $L \sim E_d(0, \Sigma, \psi)$ has a centred elliptical distribution, so that it really represents fluctuations of the loss around the expected loss. In this case the relative amounts of capital allocated using the three allocations rules (or any other allocation rule based on a positive-homogeneous risk measure) are the same and are determined by

$$\frac{EC_i}{EC_j} = \frac{\pi^r_i(1)}{\pi^r_j(1)} = \frac{\sum_{k=1}^d \Sigma_{ik}}{\sum_{k=1}^d \Sigma_{jk}}, \quad 1 \leq i, j \leq d.$$
First Justification of Euler Allocation

Let $r_\varrho$ be a risk measure function which is differentiable on $\Lambda$ and $\pi^{r_\varrho}$ an associated per-unit capital allocation principle. Then $\pi^{r_\varrho}$ is suitable for performance measurement if for all $\lambda \in \Lambda$ we have

$$\frac{\partial}{\partial \lambda_i} \frac{-E(L(\lambda))}{r_\varrho(\lambda)} > 0,$$

if

$$\frac{-E(L_i)}{\pi_i^{r_\varrho}(\lambda)} > \frac{-E(L(\lambda))}{r_\varrho(\lambda)},$$

$$\frac{\partial}{\partial \lambda_i} \frac{-E(L(\lambda))}{r_\varrho(\lambda)} < 0,$$

if

$$\frac{-E(L_i)}{\pi_i^{r_\varrho}(\lambda)} < \frac{-E(L(\lambda))}{r_\varrho(\lambda)}.$$

If the (per unit) return on investment opportunity $i$ divided by the (per unit) risk capital $\pi_i^{r_\varrho}$ is better (worse) than the performance of the overall portfolio, then increasing (decreasing) the weight $\lambda_i$ by a small amount improves the overall performance of the portfolio. Tasche proved that Euler is the only per-unit capital allocation principle suitable for performance measurement. [Tasche, 1999]
Justification of Euler Allocation for Coherent RM

If $r_\varrho$ derives from a coherent risk measure $\varrho$ then $\varrho(L) \leq \sum_{i=1}^{d} \varrho(L_i)$ and the overall risk capital required for the portfolio is smaller than the sum of the risk capital required for the business units on a stand-alone basis. **Fairness** suggests that each business unit should profit from this diversification benefit in the sense that $EC_i \leq \varrho(L_i)$. A slightly more general definition of fairness is that for all $\lambda \in \Lambda$ and all $\gamma \in [0, 1]^d$ the following inequality holds.

$$\sum_{i=1}^{d} \gamma_i \lambda_i \pi_i^r(\lambda) \leq r_\varrho(\gamma_1 \lambda_1, \ldots, \gamma_d \lambda_d).$$

[Denault, 2001] shows under technical conditions that the only fair allocation principle in this case is the Euler allocation.
D2. Monte Carlo Approach to Shortfall Contributions

Suppose we wish to measure portfolio risk with expected shortfall and to calculate a capital allocation based on expected shortfall contributions at the confidence level $\alpha$. We need to evaluate the conditional expectations

$$E(L \mid L \geq q_\alpha(L)) \quad \text{and} \quad E(L_i \mid L \geq q_\alpha(L)). \quad (28)$$

In standard Monte Carlo (MC) simulation the problem of rare event simulation arises, so importance sampling is advisable.
A Credit Portfolio Example

Consider a portfolio loss of the form $L = \sum_{i=1}^{m} e_i Y_i$, where the $e_i$ are deterministic, positive exposures and the $Y_i$ are default indicators with default probabilities $p_i$. $Y$ follows a Bernoulli mixture model with factor vector $\Psi$ and conditional default probabilities $p_i(\Psi)$.

We study the problem of estimating exceedance probabilities $\theta = P(L \geq c)$ for $c$ substantially larger than $E(L)$ using importance sampling.

We consider first the situation where the default indicators $Y_1, \ldots, Y_m$ are independent and discuss subsequently the extension to the case of conditionally independent default indicators. Our exposition is based on [Glasserman and Li, 2003].
Independent Default Indicators

Here we use the more general IS approach and set \( \Omega = \{0, 1\}^m \), the state space of \( Y \). The probability measure \( P \) is given by

\[
P(\{y\}) = \prod_{i=1}^{m} p_i^{y_i} (1 - p_i)^{1 - y_i}, \quad y \in \{0, 1\}^m.
\]

The moment generating function of \( L \) is

\[
M_L(t) = E\left( \exp\left( t \sum_{i=1}^{m} e_i Y_i \right) \right) = \prod_{i=1}^{m} E\left( e^{t e_i Y_i} \right) = \prod_{i=1}^{m} \left( e^{t e_i p_i} + 1 - p_i \right).
\]

The measure \( Q_t \) is given by \( Q_t(\{y\}) = E_P(\exp(t L) / M_L(t); Y = y) \) and hence

\[
Q_t(\{y\}) = \frac{\exp\left( t \sum_{i=1}^{m} e_i y_i \right)}{M_L(t)} P(\{y\}).
\]
We obtain

\[
Q_t(\{y\}) = \prod_{i=1}^m \frac{\exp(te_i y_i)}{\exp(te_i) p_i + 1 - p_i} p_i^{y_i} (1 - p_i)^{1-y_i}.
\]

and it follows that

\[
Q_t(\{y\}) = \prod_{i=1}^m q_{t,i}^{y_i} (1 - q_{t,i})^{1-y_i}
\]

where \(q_{t,i} := \exp(te_i)p_i/(\exp(te_i) p_i + 1 - p_i)\). The default indicators remain independent but with new default probability \(q_{t,i}\).

The optimal value of \(t\) is chosen such that \(E_{Q_t}(L) = c\), leading to the equation \(\sum_{i=1}^m e_i q_{t,i} = c\).
Conditionally Independent Default Indicators

The first step in the extension to conditionally independent defaults is obvious: given a realization $\psi$ of the economic factors, the conditional exceedance probability $\theta(\psi) := P(L \geq c \mid \Psi = \psi)$ is estimated using the approach for independent default indicators. This yields Algorithm 8.26 in MFE and gives an estimate $\hat{\theta}^{IS,1}(\psi)$ where $n_1$ is the number of random draws of $(Y_1, \ldots, Y_m)$.

Our ultimate aim is to estimate $\theta = P(L \geq c)$. In a naive approach we could generate realizations of $\Psi$ and estimate $\theta$ by calculating the average $\frac{1}{n} \sum_{i=1}^{n} \hat{\theta}^{IS,1}(\Psi_i)$.

However a dramatic improvement can be obtained by also applying importance sampling to the distribution of $\Psi$. This is the idea behind Algorithm 8.27 in MFE.
One-Factor KMV/CreditMetrics Model

Consider a model with conditional default probabilities \( p_i(\Psi) \) where \( \Psi \sim N(0, 1) \). Instead of generating \( \Psi_1, \ldots, \Psi_n \) from a standard normal \( N(0, 1) \) distribution we should use exponential tilting to generate them from a \( N(\mu, 1) \) distribution for some sensibly chosen value of \( \mu \). An approach to determining \( \mu \) and references to literature on this subject are given in MFE (p.373).

We obtain the algorithm:

1. Generate \( \Psi_1, \ldots, \Psi_n \sim N(\mu, 1) \) independently.

2. For each \( \Psi_i \) calculate \( \hat{\theta}_{IS,1}^{(n)}(\Psi_i) \) by importance sampling.

3. Determine the full IS estimator: \( \hat{\theta}_{IS} = \frac{1}{n} \sum_{i=1}^{n} r_\mu(\Psi_i) \hat{\theta}_{IS,1}^{(n)}(\Psi_i) \) where
   \( r_\mu(\psi) = \exp(-\mu \psi + \frac{1}{2} \mu^2) \).
Example

We consider an exchangeable portfolio of 100 firms with identical unit exposures, default probabilities 0.05 and asset correlations (i.e. values of $\beta$) 0.05. Aim is to calculate the tail probability $P(L \geq 20)$ by IS. (In such a simple model it can in fact be calculated analytically to be 0.00112.) We compare

1. Naive Monte Carlo ($n = 10000$).

2. IS for factor distribution ($n = 10000$).

3. Naive Monte Carlo for factor ($n = 10000$) and IS for conditional default distribution ($n1 = 50$).

4. IS for factor distribution ($n = 10000$) and conditional default distribution ($n1 = 50$).
Results
D3. Diversification Scoring

Tasche [Tasche, 2006] defines diversification factors as follows:

\[
DF = \frac{SRC}{\sum_{i=1}^{d} SRC_i}
\]

\[
DF_i = \frac{EC_i}{SRC_i}
\]

The former measures portfolio diversification - overall benefit in terms of reduction in solvency capital that the business units achieve by being together within the enterprise.

The latter measures effect of diversification for unit \( i \) - the benefit to business unit \( i \) in terms of reduction in solvency capital achieved by belonging to enterprise.
Bibliography


