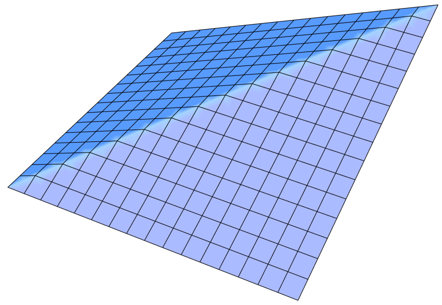


ACTSC 445: Quantitative Risk Management

Prof. David Saunders • Fall 2015 • University of Waterloo
TR 1:00–2:20PM in MC 2065

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

1 Enterprise Risk Management Foundations **3**
Risk Management and Enterprise Risk Management 4
Insurable Risks and Financial Risk Management 4
Actuarial Risk Management, Asset–Liability Management, Quantitative Risk Management, and Project Risk Management 6
Governance, Strategy, and Strategic Plans 7
Risk Management Process 9

Risk Treatment . . . 10
Financial Risks . . . 12
External Risks . . . 13
Nonfinancial and Operational Risks . . . 14

2 Risk Measures and Market Risk **16**
Risk Measures . . . 16
Value-at-Risk and Conditional Tail Expectation . . . 17
Coherence . . . 22
Estimating VaR and CTE . . . 27
VaR in Banking . . . 30
Delta-VaR . . . 32
Historical Simulation and Monte Carlo Methods . . . 32
Backtesting . . . 33

3 Frequency–Severity Analysis **34**
Aggregate Loss Models and Insurance . . . 35
Frequency Models . . . 35
The $(a, b, 0)$ and $(a, b, 1)$ Distributions . . . 37
Severity Models . . . 39
Estimating the Aggregate Loss Distribution . . . 41
Panjer Recursion . . . 44
Discrete Fourier Transforms . . . 45
Aggregate Loss Risk Measures . . . 46
Partial Insurance . . . 48

4 Extremes, Dependence, and Financial Return Models **52**
Maximum Losses . . . 52
GEV Distributions . . . 53
Fitting GEV Distributions by Block Maxima . . . 57
Threshold Exceedances and the GPD . . . 58
The Hill Method . . . 61

5 Copulas and Dependence **62**
Importance of Dependence . . . 62
Copulas and Sklar's Theorem . . . 63
Explicit and Implicit Bivariate Copulas . . . 67
Measures of Dependence . . . 69
Tail Dependence . . . 72
Archimedean Copulas . . . 73
Fitting Copulas . . . 75

Goodness of Fit and Simulation . . . 77

6 Models for Market Risk and Other Economic Risks **80**
Financial Return Stylized Facts . . . 80
Stock Price Models . . . 83
Economic Scenario Generators . . . 88
Model Risk and Model Governance . . . 88

7 Capital, Credit Risk, and Regulation **89**
Capital and Balance Sheets . . . 89
Bank Equity . . . 90
Economic Capital . . . 91
Regulatory Capital . . . 92
Trading Book Credit Risk . . . 93
Capital Allocation and Performance . . . 95
Credit Risk Models . . . 98
Credit Ratings and Recovery . . . 99
Merton's Structural Model . . . 101
Credit Migration and Threshold Models . . . 107
Vasicek's Single Factor Credit Model . . . 112
Basel II and Extensions . . . 116

Appendix: Lecture and Tutorial Index **123**

Editorial Note At the time of this offering, much of ACTSC 445's course content was based on *Quantitative Risk Management: Concepts, Techniques and Tools (Revised Edition)* by Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts (Princeton University Press, 2015), which is unquestionably the greatest textbook written on the subject.

Lecture 1 — Tuesday, September 15, 2015

1. Enterprise Risk Management Foundations

Risk Management and Enterprise Risk Management

All firms manage risk, but in financial services it is a core part of the business. Firms therefore need processes and frameworks for identifying, measuring, and managing risk; these should be **top-down**, with risk management integral to corporate governance rather than confined to an isolated technical function.

Enterprise risk management takes this top-down idea seriously. It implies a holistic approach in which risk management is incorporated into the firm's values, strategy, operations, and financial reporting.

Historically, risk management practice developed through several related areas: insurable risk management, financial risk management, actuarial risk management, asset–liability management, quantitative risk management, project risk management, and operational risk management. Enterprise risk management attempts to integrate these areas into a single enterprise-wide framework.

Insurable Risks and Financial Risk Management

The traditional management of insurable risks is studied in risk management and insurance. Such **pure risks** involve downside exposure rather than upside opportunity and can often be insured; the classical responses are transfer, retention, avoidance, and reduction.

Financial risk management begins with portfolio theory, especially the Markowitz–Sharpe mean–variance framework. Variance is used as a measure of risk, although it is a two-sided measure because it penalizes upside and downside deviations symmetrically. Mean–variance portfolio analysis balances risk and return, and it eliminates portfolios that are not on the efficient frontier.

The capital asset pricing model (CAPM) gives

$$\mathbb{E}\left[R_k\right] = r_f + \beta_k\big(\mathbb{E}\left[R_M\right] - r_f\big).$$

Here R_k is the return on asset k, r_f is the risk-free rate, R_M is the return on the market portfolio, and β_k is the asset's beta.

If we subtract the risk-free rate from both sides, we obtain

$$\mathbb{E}\left[R_k\right] - r_f = \beta_k\big(\mathbb{E}\left[R_M\right] - r_f\big),$$

so the expected excess return on the asset is its beta times the **market risk premium**. The coefficient β_k is the **systematic risk** of the asset, meaning the part of the asset's risk that is attributable to movements in the overall market and cannot be diversified away in the CAPM framework.

A European call option pays

$$\max(S_T - K, 0)$$

at time T, while a European put option pays

$$\max(K - S_T, 0).$$

Here K is the strike price and S_T is the underlying asset value at maturity. Writing these payoffs out by cases makes the economics clearer: the call pays $S_T - K$ when $S_T > K$ and otherwise pays 0, while the put pays $K - S_T$ when $S_T < K$ and otherwise pays 0. Black–Scholes–Merton analysis showed how to price and hedge such contingent claims through replication, and this became a powerful framework for financial risk management.

These classical tools also have important limitations. Standard deviation is not always a good measure of portfolio risk because it treats upside and downside deviations symmetrically, even though firms are usually more concerned with losses than gains. It may also fail to capture rare but high-impact events. In practice, means, standard deviations, and correlations are not known parameters; they must be estimated, often with substantial uncertainty, and different investors may have different views about them. From an enterprise perspective, beta is also too narrow because it studies assets or investments in isolation, ignores the interaction among different parts of the enterprise, and ignores mismatches between assets and liabilities. Likewise, Black–Scholes–Merton analysis depends on strong assumptions such as Gaussian log returns, no transaction costs or taxes, continuous trading, and constant volatility.

There is therefore both a contrast and a connection between insurable risk management and financial risk management. Insurable risk management is traditionally concerned with pure downside exposures and with the choice among transfer, retention, avoidance, and reduction. Financial risk management is more explicitly concerned with the joint determination of risk and return, with market prices, hedging, and valuation. Even so, both areas require measurement, modelling, and deliberate treatment of uncertainty, and both ultimately feed into enterprise-wide decision making.

Actuarial Risk Management, Asset–Liability Management, Quantitative Risk Management, and Project Risk Management

Actuarial risk management combines insurable and financial risks. Business risk has both upside and downside, and actuaries often manage solvency and, to some extent, profitability for insurers and pension plans through product design, pricing, valuation, and investment strategy. Dynamic financial analysis is one of the main actuarial frameworks here, and it makes use of scenario tests, stress tests, and stochastic simulation.

Asset–liability management is concerned with mismatching between assets and liabilities. Interest rate risk is a central issue, and duration and immunization analysis are standard tools. The same general framework also extends to exchange rate risk, credit spreads, and other sources of mismatch.

Quantitative risk management provides the technical toolkit used across these areas. Important examples include extreme value theory, dependency modelling, risk measures such as value-at-risk and conditional tail expectation, and econometric models of financial processes.

Project risk management is aimed at business projects with a definite beginning and end. It is multidisciplinary and must deal with complex budgeting and scheduling issues. In practice, this means that risk management methods must be adapted not only to financial and insurance exposures, but also to operational and organizational contexts.

The Committee of Sponsoring Organizations of the Treadway Commission (COSO) is a widely cited source for enterprise risk management. It describes enterprise risk management as "a process, effected by an entity's board of directors, management, and other personnel, applied in strategy setting and across the enterprise, designed to identify potential events that may affect the entity and to manage risk so that it remains within the entity's risk appetite, thereby providing reasonable assurance regarding the achievement of the entity's objectives." This definition emphasizes the process nature of enterprise risk management, its connection to strategy, its scope across the enterprise, and its focus on risk appetite and the achievement of objectives.

Meanwhile, the Risk Management Society describes enterprise risk management as "a strategic business discipline that supports the achievement of an organization's objectives by addressing the full spectrum of its risks and by managing the combined impact of those risks as an interrelated risk portfolio."

The Risk Management Society description can be unpacked in several ways. Enterprise risk management encompasses all major sources of organizational exposure, including financial, operational, reporting, compliance, governance, strategic, and reputational risks; it then prioritizes and manages them as an interrelated portfolio rather than as separate silos. That portfolio is evaluated in the context of the organization's environments, systems, circumstances, and stakeholders, with explicit recognition that interactions can make the combined exposure differ materially from the sum of its parts. The resulting process applies to both quantitative and qualitative risks, treats effective risk management as a possible competitive advantage, and embeds it in critical organizational decisions.

Under ISO 31000, enterprise risk management should create and protect value. To do so, it must be integral to processes and decisions, explicitly address uncertainty, and use the best available information in a systematic and timely way. The process should also be tailored to the organization: human and cultural factors matter, transparency and inclusion matter, and the framework must remain dynamic enough to support continuous improvement.

The ISO 31000 risk management process consists of establishing the context, identifying risks, analyzing risks, evaluating risks, and treating risks. Communication and consultation occur throughout the process, and monitoring and review also continue throughout rather than appearing only at the end. Risk management is therefore a cycle embedded in governance rather than a one-time calculation.

Governance, Strategy, and Strategic Plans

Enterprise risk management begins with corporate governance: the board of directors sets the firm's mission, vision, and values, which should incorporate its risk appetite. The board-level strategic plan then interprets those commitments in operational terms; it identifies stakeholders explicitly and links the resulting objectives to performance metrics, often through a balanced scorecard.

At the financial level, the ultimate objective is to maximize shareholder value. This is supported by objectives at the customer level, such as exceeding customer expectations and inspiring loyalty. Those customer outcomes, in turn, depend on internal business process objectives such as creating quality partnerships, maximizing operational effectiveness, and creating high-quality products. At the base are learning and growth objectives, namely recruiting quality staff and training employees. Strategy is read from the bottom upward: capabilities support processes, processes support customer outcomes, and customer outcomes support financial performance.

The Society of Actuaries (SOA) strategic plan provides a concrete example. Its mission is to advance actuarial knowledge and to enhance the ability of actuaries to provide expert advice and relevant solutions for financial, business, and societal problems involving uncertain future events. Its vision is that actuaries should be the leading professionals in the measurement and management of risk.

The stakeholder portion of the SOA plan identifies several constituencies and the value proposition for each: candidates want the skills and credentials needed to advance; members want opportunities to continue developing and applying those skills; employers and clients want highly qualified experts who remain relevant to their needs; and the public wants trust to be preserved and the common good served.

The strategic themes in the SOA plan are also explicit. The first theme is to develop knowledge by producing and supporting research that expands the boundaries of actuarial science, by promoting the development of intellectual capital, and by identifying opportunities for its application. The second theme is to transfer knowledge by offering a prestigious credential, maintaining its value through leading-edge educational systems and programs, and leveraging a variety of delivery channels to disseminate and facilitate knowledge transfer. The third theme is to cultivate opportunities by establishing and expanding the actuarial brand and marketing message, by creating and promoting new areas of practice, and by ensuring a robust stream of well-qualified new candidates and credentialed actuaries.

The same plan also emphasizes the development and promotion of professional networks. In particular, sections are to function as knowledge communities and networking facilitators, long-term partnerships are to be built with actuarial and other professional organizations around the world, and relationships are to be expanded across academia, think tanks, and policy influencers. Workforce and staff objectives include attracting, developing, and retaining talented people, focusing on outcomes and rewarding execution, identifying and developing leaders at all levels, and promoting a culture of commitment, innovation, service, and excellence. Financially, the plan seeks to generate margins in order to fund strategic activities, align the use of SOA financial resources with the strategic plan, and deliver cost-efficient support and services. The professional foundation of the plan is specialized knowledge, integrity, and public interest.

A strategic plan should not be separated from risk management. A **balanced scorecard** requires explicit performance metrics for each strategic element, and those metrics should include risk measures, triggers, and thresholds. The plan should identify whether the organization is operating within its risk appetite and should specify the risk management process that will be used when limits are approached or breached.

Risk Management Process

The risk management process has five steps: determine the firm's risk appetite and tolerance, identify its risks, analyze them, evaluate the results, and then manage or treat the exposures.

The firm's **risk appetite** is its high-level statement about the amount and type of risk it is willing to seek or accept in pursuit of its objectives. **Risk tolerance** is a more detailed and usually more quantitative articulation of that appetite. **Risk limits** are then set at the business unit or operational level so that the broad appetite is translated into day-to-day constraints.

One common interpretation is that the risk appetite specifies the target level of risk within the firm's risk corridor, while risk tolerance describes the acceptable range of deviation around that target. On that interpretation, risk limits are the concrete operational constraints that keep the firm from drifting too far away from its chosen risk appetite.

The board of directors is responsible for determining risk appetite, risk tolerance, and risk limits, whether quantitative or qualitative. These may incorporate external signals such as analyst or rating agency views, but they must be specific about the risks the firm is prepared to seek, avoid, or accept; they must also reflect the economic capital required to support those choices.

Economic capital is the capital required by a bank or insurer to limit the risk of insolvency to a specified level over a given time horizon. It is distinct from regulatory capital, since it is based on the firm's own internal standards and models rather than solely on external rules.

A good economic capital model promotes firm-wide consistency by supporting risk assessment, risk pricing, and the evaluation of return on risk capital. It is closely related to the Own Risk and Solvency Assessment (ORSA), the internal process through which an organization evaluates whether its capital is adequate for its risk profile.

> **Example 1.1** A firm might state that it wants to limit the probability of a 20% drop in earnings over the next two years to no more than 5%. A bank might say that it has a very low appetite for credit risk. A university might state that preserving its reputation is critical and therefore that it has a low appetite for activities that could place that reputation in jeopardy or lead to undue adverse publicity. A balance-sheet constraint such as ensuring that long-term borrowings never exceed 20% of net assets is another example of a risk limit expressed in operational terms.

Risk identification must be rigorous and regular, combining a bottom-up process that involves frontline staff with a top-down process led by senior officers. It should cover not only the risks and their types, but also changes and trends in known risks and new exposures as they emerge. Even the **unknown unknowns** (exposures not yet clearly conceptualized or measured) must remain in view; they are closely related to **Knightian uncertainty**, in which uncertainty is present but cannot be fully described by known probabilities.

Risk analysis then studies the potential financial consequences of adverse outcomes. It may begin with qualitative assessment and high-level tools such as heat maps, proceed through deterministic analysis, scenarios, and stress tests, and culminate in stochastic analysis; throughout, it should account for both the effect and the continuing maintenance of mitigation strategies.

Risk evaluation compares the analysis with the firm's targets and risk appetite. Aggregation matters at this stage: risks cannot be evaluated only one by one, because their interactions may increase or decrease the enterprise's total exposure.

A **top-down assessment** identifies and evaluates risk potential at the enterprise-wide level and supports strategic decisions about risk financing. It often relies on stress testing, scenario analysis, and risk mapping, together with surveys or interviews of senior managers and the use of internal or external data. Its weakness is that it may fail to identify operating issues and may not translate into action at the frontline.

By contrast, a **bottom-up assessment** identifies and evaluates risk potential at the transaction or business unit level, drawing on frontline staff, surveys, interviews, what-if analysis, self-assessment, reviews of new products, and internal data. It is particularly useful for process risks, although it may lack perspective or credibility, receive too little priority, or arrive too late. Both approaches are needed.

Risk Treatment

For insurable risks, the classical responses are avoidance, reduction or mitigation, transfer, and retention without mitigation. The broader ISO 31000 framework expands this menu: a firm may also pursue a risk consistent with its strategy, remove the risk source, alter the probability or severity of an event, share the exposure, or retain it by an informed decision. Treatment is not always about elimination; firms often take risk deliberately, but with governance, limits, and capital support in place.

Lecture 2 — Tuesday, September 22, 2015

Example 2.1 The Deepwater Horizon case involved an explosion on a drilling rig in the Gulf of Mexico. Eleven people were killed, sixteen were injured, and the resulting oil spill caused severe environmental damage.

Several causes were identified: defective cement, inadequate training in shutdown procedures, poor equipment maintenance, bypassed alarms and automatic shutdown processes, and an overall weak safety culture. Cheap materials and methods compounded these failures, while modelling outputs that indicated risk were ignored; the resulting losses included cleanup costs, legal liabilities, and severe reputational damage.

Example 2.2 In 2007, Manulife had roughly \$75 billion of exposure through segregated funds, which are mutual-fund-type products with repayment guarantees. The firm did not hedge the equity exposure and also appears to have gamed the capital requirements. When the 2008 equity crash arrived, the firm was forced into an emergency stock issue, a \$3 billion bank loan, a credit rating downgrade, class action lawsuits, and about a 75% drop in its share price.

The alleged causes included a strategic decision not to hedge equity exposure, an overly dominant chief executive officer, an underqualified board, thin-tailed models, the use of regulatory capital as a target rather than a baseline, poor understanding of finance by senior actuarial leadership, poor contract design, and aggressive marketing. The resulting losses included capital strain, liquidity risk, reputational damage, and legal liabilities.

A **risk taxonomy** is a classification system for organizing the major risks faced by an enterprise, not a device for forcing them into rigidly separate categories. Many risks belong naturally to more than one class, and one adverse event often generates further exposures elsewhere in the organization; nevertheless, a broad classification remains useful for structuring analysis and governance.

One common distinction separates external risks (including financial market, political, regulatory, macroeconomic, and environmental risks) from internal ones such as operational, strategic, and reputational risk. Even this division is imperfect, since some exposures have both internal and external components.

Financial Risks

Stock market risk is the risk arising from movements in equity market values. This may affect the asset side of the balance sheet, the liability side through embedded options, or the relationship between assets and liabilities. In some situations a firm can even be exposed to upward market movements, such as when liabilities increase as equity markets rise.

Interest rate risk is the risk arising from movements in borrowing and lending rates. Interest rate movements affect borrowing costs and the value of loans and bonds held by the firm. Recall that for a fixed-coupon bond with cash flows C_t and yield y, the price (as a function of y) is

$$P(y) = \sum_{t=1}^{n} \frac{C_t}{(1+y)^t}.$$

Assuming $y > -1$, all cash flows are nonnegative, and at least one cash flow is positive, differentiating with respect to y gives

$$\frac{\mathrm{d}P}{\mathrm{d}y} = -\sum_{t=1}^{n} \frac{tC_t}{(1+y)^{t+1}} < 0,$$

so bond prices fall when interest rates rise. Interest rate risk also includes **reinvestment risk**, since falling rates may prevent the firm from earning the return it had expected on reinvested cash flows. **Spread risk** is the risk arising from changes in the difference between short- and long-term rates or, more broadly, between different borrowing rates. **Asset–liability management risk** is a further aspect of interest rate risk and arises when the duration of assets is mismatched with the duration of liabilities.

Exchange rate risk (or **foreign exchange risk** or **currency risk**) is the risk arising from unexpected movements in relative currency values. **Transaction risk** arises from contracts denominated in different currencies. **Economic risk** refers to the broader exposure created by imports, exports, foreign investments, foreign subsidiaries, foreign competition, and supplier or customer relationships. **Translation risk** arises in financial reporting when foreign currency items must be converted into the reporting currency (it is not related to language translation).

Credit risk is the risk of default or deterioration in the credit quality of a bond issuer or other obligor. It includes **bond default risk**, **downgrade risk**, **credit spread risk**, and **sovereign risk**. Credit ratings such as AAA, BB, C, and D summarize perceived credit quality, with higher ratings generally corresponding to higher bond values because investors demand less compensation for default risk. A downgrade usually reduces the market value of a bond because its required yield increases.

Systemic risk is the risk arising from instability or failure in the financial system. This is often treated as a subset of market risk. It is particularly associated with domino or contagion effects in which the failure of one financial institution contributes to failures elsewhere in the system.

Liquidity risk is the risk that assets cannot be realized quickly enough to meet liabilities, without a severe adverse impact on price, even though the firm is technically solvent. It may be caused by illiquid assets, unanticipated cash flows, a reputational shock at the company level, or stressed market conditions. Note that liquidity is not the same as marketability! An asset may be marketable in normal times and still be difficult to sell quickly in the volume required. **Liquidation value risk** arises when unexpected timing or amounts of required cash force the firm to sell assets at depressed prices. **Affiliated company risk** arises when capital is tied up in a subsidiary and cannot be realized when needed. **Capital market risk** arises when the firm cannot raise money through the markets, for example by issuing equity or debt.

External Risks

Business cycle risk is the risk arising from economy-wide fluctuations in the economic environment. These fluctuations are difficult to predict in timing and duration. They are closely connected with financial market risks, and economic troughs are generally harmful to firms, although some firms may benefit from them. For example, a recession may reduce demand for luxury goods but increase demand for discount retailers.

Inflation risk is the reduction in real returns caused by unexpected inflation. Long-term fixed cash flows are especially vulnerable. The risk is fundamentally a cash flow mismatch problem. A firm that receives fixed nominal income while making inflation-linked payments is exposed to losses in real terms.

Political risk arises from adverse political changes in a firm's home country or in a foreign country where it has exposure. For example, a change in government or political instability can affect the firm's operations and profitability. Political instability can interrupt business operations, and a change in the political climate can directly harm a particular enterprise. Political risk is often connected with economic risk and exchange rate risk.

Regulatory risk is the risk that changes in law adversely affect the business. Examples include labour laws, currency transaction laws, financial reporting laws, licensing laws, regulatory capital laws, and tax laws.

Environmental risk consists of risks arising from environmental events and trends. Examples include changing patterns of natural disasters, soil degradation, drought, and the effects of climate change on markets, suppliers, and transportation costs. The term is also sometimes used for liability arising from environmental events, such as inadequate waste disposal or the failure of an oil tanker or drilling platform. In that second sense it overlaps with internal risk categories.

Nonfinancial and Operational Risks

The Basel Committee defines **operational risk** as the risk of loss arising from inadequate or failed internal processes, people, and systems, or from external events. Operational risk is often broken down into people risk, information technology risk, process risk, project risk, legal risk, pricing risk, and related categories.

People risk arises when employees or agents fail to follow the firm's rules, processes, or procedures. The failures may be deliberate, as in fraud, theft, sabotage, or bypassing systems to save time or increase remuneration. They may also be unintentional, as in error, carelessness, inadequate training, inadequate competence, poor resource allocation, or weak management controls. Human resource and management processes can also fail. This includes **key person risk** (where the loss of a key individual can significantly impact the firm's operations, revenue, or reputation), failure to recruit suitably qualified candidates, and inadequate processes for controlling fraud or expenses.

IT risk includes accidental loss or corruption of data, viruses, malware, denial-of-service attacks, unidentified program bugs, theft of data or intellectual property through security breaches, capacity failures, service outages, and failures by suppliers or consultants.

Project risk arises in project development and implementation. Common examples are scope risk, defect risk, schedule risk, and resource risk. **Scope risk** arises when the project scope is not well defined or controlled, leading to scope creep and cost overruns. **Defect risk** arises when the project delivers a product that does not meet quality standards or user requirements. **Schedule risk** arises when the project timeline is unrealistic or not adhered to, leading to delays. **Resource risk** arises when there are insufficient resources (such as personnel, budget, or equipment) to complete the project successfully.

Legal risk covers lawsuits against the enterprise, the risks of pursuing litigation, and legal defects in transactions. It includes liability arising from breaches of laws or regulations, negligence leading to civil liability, defective transactions, the costs of defending or opposing patents and copyright claims, and adverse changes in law.

Pricing risk is the risk of underpricing or undervaluing products. It is especially serious when there is a delay between contract formation and delivery of goods or services, or when currency risk is present. Model risk and adverse selection are important contributors to pricing risk.

Process risk is specific to the firm's business processes. Examples include health and safety failures caused by defective equipment, flawed procedures or human error, manufacturing and engineering failures caused by defective components, **model risk** (where a model is applied in a setting that violates its assumptions), **basis risk** (where a hedge or proxy does not move closely enough with the exposure it is meant to offset), and failures in maintenance processes.

Strategic risk is the risk arising from strategic directives and strategic decision making. All strategic decisions involve risk, including the decision not to act. Some strategic decisions fail and cause investments to be lost. Strategic risk is especially significant when decisions are made using flawed processes, weak planning, poor data, or insufficient expertise.

Reputational risk is the risk that the enterprise is harmed through damage to its reputation. Reputational risk often follows another risk event. The BP Deepwater Horizon disaster from Example 2.1 is a standard example. The **Enron scandal** is another standard example: Enron used accounting fraud and off-balance-sheet entities to conceal losses and overstate performance, and when the truth emerged the firm collapsed and the reputation of Arthur Andersen, their accounting firm and auditor, was effectively destroyed along with it.

The **Tylenol case** is a contrasting example. After several people died from cyanide-laced Tylenol capsules in 1982, Johnson & Johnson rapidly recalled the product, cooperated with authorities, communicated openly with the public, and later reintroduced Tylenol with tamper-resistant packaging. The case is therefore widely viewed as an example in which good crisis management strengthened reputation rather than destroying it. As Warren Buffett observed, it can take twenty years to build a reputation and five minutes to destroy it.

Example 2.3

The Deepwater Horizon case from Example 2.1 involves environmental risk, operational risk, legal risk, reputational risk, and strategic risk. The operational component includes defective equipment, poor maintenance, inadequate training, and failures in process and safety culture. The environmental losses then generated legal liabilities and severe reputational damage.

The Manulife segregated fund case from Example 2.2 involves stock market risk, liquidity risk, strategic risk, pricing and model risk, regulatory risk, and reputational risk. The unhedged guarantee exposure made the firm highly vulnerable to falling equity markets, while weak governance and poor model assumptions amplified the losses. Capital strain, emergency financing, and market reaction then created liquidity and reputational consequences.

Lecture 3 — Tuesday, September 29, 2015

2. Risk Measures and Market Risk

Risk Measures

For quantifiable risks, potential losses can often be modelled with a probability distribution. In simple settings we may use standard parametric families such as the normal, lognormal, Pareto, or compound Poisson distributions. In more complicated settings, such as operational risk losses, Monte Carlo methods are often used instead.

Definition 3.1 Fix a time horizon and a class of loss random variables. A **risk measure** is a functional ρ that maps each loss L in that class to a real number $\rho(L)$ intended to quantify the risk associated with L.

Risk measures are used in premium calculations, benchmarking, regulatory capital, and economic capital. They are sometimes described as estimates of the maximum possible loss, but that interpretation is too strong, since most risk measures summarize only part of the loss distribution.

There are several standard principles for calculating insurance premiums. The simplest is the **net premium principle**, which sets the premium equal to the expected loss, that is,

$$\rho(X) = \mathbb{E}[X].$$

Three other standard premium principles are

$$\rho(X) = (1+\alpha)\mathbb{E}[X], \quad \alpha \geq 0,$$

which is the **expected value principle**,

$$\rho(X) = \mathbb{E}[X] + \alpha\sqrt{\mathrm{Var}(X)}, \quad \alpha \geq 0,$$

which is the **standard deviation principle**, and

$$\rho(X) = \mathbb{E}[X] + \alpha\,\mathrm{Var}(X), \quad \alpha \geq 0,$$

which is the **variance principle**.

Value-at-Risk and Conditional Tail Expectation

Definition 3.2 For $\alpha \in (0,1)$, the **α-Value-at-Risk** (or **α-VaR**) of the loss random variable L is

$$\mathrm{VaR}_\alpha(L) = Q_\alpha := \inf\{q : \mathbb{P}(L \le q) \ge \alpha\}.$$

This is the lower α-quantile, or generalized inverse, of the distribution function of L, which we all know and love. If the distribution function is continuous at Q_α, then

$$\mathbb{P}(L \le Q_\alpha) = \alpha \qquad \text{and} \qquad Q_\alpha = F_L^{-1}(\alpha).$$

Usually $\alpha \ge 0.95$ is used (typically $\alpha \in \{0.95, 0.99, 0.999\}$), so that the VaR is a high quantile of the loss distribution. For example, when the distribution is continuous at its 0.99-quantile, $\mathrm{VaR}_{0.99}(L)$ is exceeded with probability 1%. With an atom at the quantile, the exceedance probability can be smaller than $1-\alpha$. VaR is sometimes interpreted as a worst-case loss, but that interpretation is too strong, since it identifies a quantile and ignores the magnitude of losses beyond it.

Example 3.3 Suppose

$$\mathbb{P}(L = 100) = 0.005, \qquad \mathbb{P}(L = 50) = 0.045,$$
$$\mathbb{P}(L = 10) = 0.10, \qquad \text{and} \qquad \mathbb{P}(L = 0) = 0.85.$$

The distribution function therefore satisfies

$$F_L(0) = 0.85, \qquad F_L(10) = 0.95, \qquad F_L(50) = 0.995, \qquad \text{and} \qquad F_L(100) = 1.$$

It follows that

$$\mathrm{VaR}_{0.99}(L) = 50, \qquad \mathrm{VaR}_{0.95}(L) = 10, \qquad \text{and} \qquad \mathrm{VaR}_{0.80}(L) = 0.$$

Definition 3.4 For $\alpha \in (0,1)$ and an integrable loss L, the **conditional tail expectation** (**CTE**) (or **expected shortfall** or **Tail-VaR** or **CVaR**) is defined as

$$\mathrm{CTE}_\alpha(L) := \frac{1}{1-\alpha} \int_\alpha^1 Q_u \,\mathrm{d}u.$$

In words, this is the average loss in the worst $1-\alpha$ fraction of the distribution.

Proposition 3.5 If the distribution function of L is continuous at Q_α, then

$$\mathrm{CTE}_\alpha(L) = \mathbb{E}\left[L \mid L > Q_\alpha\right].$$

Proof. Let $U \sim \mathrm{Unif}(0,1)$ and set $X = Q_U$. The inverse-transform construction gives $X \stackrel{d}{=} L$. Continuity of F_L at Q_α implies $F_L(Q_\alpha) = \alpha$, and monotonicity of the quantile function then gives

$$X > Q_\alpha \quad \iff \quad U > \alpha \qquad \text{almost surely.}$$

Therefore,

$$\mathbb{E}\left[L \mid L > Q_\alpha\right] = \mathbb{E}\left[X \mid X > Q_\alpha\right] = \mathbb{E}\left[Q_U \mid U > \alpha\right] = \frac{1}{1-\alpha}\int_\alpha^1 Q_u \,\mathrm{d}u = \mathrm{CTE}_\alpha(L),$$

as required. $\square$

If there is mass at the quantile (i.e., there exists some $\varepsilon > 0$ such that $Q_\alpha = Q_{\alpha+\varepsilon}$), then the conditional expectation formula uses too little of the worst $1-\alpha$ tail and needs a correction:

Proposition 3.6 Let

$$\beta' := F_L(Q_\alpha).$$

If $\beta' < 1$, then

$$\mathrm{CTE}_\alpha(L) = \frac{(\beta'-\alpha)Q_\alpha + (1-\beta')\mathbb{E}\left[L \mid L > Q_\alpha\right]}{1-\alpha}.$$

Proof. Let $U \sim \mathrm{Unif}(0,1)$ and let $X = Q_U$. Then X has the same distribution as L by the inverse-transform construction. Since the quantile function is nondecreasing and β' is the largest level at which the quantile is still equal to Q_α, we have

$$Q_u = Q_\alpha \quad \text{for } \alpha \le u \le \beta', \qquad \text{and} \qquad Q_u > Q_\alpha \quad \text{for } u > \beta'.$$

Therefore,

$$\mathrm{CTE}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 Q_u \,\mathrm{d}u$$

$$= \frac{1}{1-\alpha}\left(\int_{\alpha}^{\beta'} Q_u \,\mathrm{d}u + \int_{\beta'}^{1} Q_u \,\mathrm{d}u\right)$$
$$= \frac{1}{1-\alpha}\left((\beta'-\alpha)Q_\alpha + \int_{\beta'}^{1} Q_u \,\mathrm{d}u\right).$$

Also, by the characterization above,

$$X > Q_\alpha \iff U > \beta'.$$

Since X and L have the same distribution,

$$\mathbb{E}\left[L \mid L > Q_\alpha\right] = \mathbb{E}\left[X \mid X > Q_\alpha\right] = \mathbb{E}\left[Q_U \mid U > \beta'\right] = \frac{1}{1-\beta'}\int_{\beta'}^{1} Q_u \,\mathrm{d}u,$$

because $U \sim \mathrm{Unif}(0,1)$. Substituting this into the previous expression gives

$$\mathrm{CTE}_\alpha(L) = \frac{(\beta'-\alpha)Q_\alpha + (1-\beta')\mathbb{E}\left[L \mid L > Q_\alpha\right]}{1-\alpha}.$$

□

Example 3.7 For the loss distribution in Example 3.3,

$$\mathrm{CTE}_{0.99}(L) = \frac{0.005 \cdot 50 + 0.005 \cdot 100}{0.01} = 75.$$

Also,

$$\mathrm{CTE}_{0.95}(L) = \mathbb{E}\left[L \mid L > 10\right] = \frac{100(0.005) + 50(0.045)}{0.05} = 55.$$

Finally,

$$\mathrm{CTE}_{0.80}(L) = \frac{0.05 \cdot 0 + 0.15 \cdot \mathbb{E}\left[L \mid L > 0\right]}{0.20} = \frac{100(0.005) + 50(0.045) + 10(0.10)}{0.20} = 18.75.$$

Several basic facts are immediate. We always have $\mathrm{CTE}_\alpha(L) \geq \mathrm{VaR}_\alpha(L)$, with equality only when the quantile is already the essential supremum of the loss. Extending the definition continuously to level zero gives

$$\mathrm{CTE}_0(L) = \mathbb{E}\left[L\right], \qquad \lim_{\alpha \downarrow 0} \mathrm{VaR}_\alpha(L) = \operatorname{ess\,inf} L, \qquad \text{and} \qquad \mathrm{VaR}_{0.5}(L) = \text{the lower median of } L.$$

VaR and CTE are available in closed form for a number of distributions.

Example 3.8 If $L \sim \mathcal{N}(\mu, \sigma^2)$ with $\sigma > 0$, then for $q \in (0,1)$,

$$\mathrm{VaR}_q(L) = \mu + \sigma \Phi^{-1}(q) \qquad \text{and} \qquad \mathrm{CTE}_q(L) = \mu + \sigma \frac{\varphi(\Phi^{-1}(q))}{1-q}.$$

Proof. Let $Z = (L - \mu)/\sigma$, so that $Z \sim \mathcal{N}(0,1)$. If $z_q = \Phi^{-1}(q)$, then

$$\mathbb{P}(L \le \mu + \sigma z_q) = \mathbb{P}(Z \le z_q) = \Phi(z_q) = q,$$

so

$$\mathrm{VaR}_q(L) = \mu + \sigma \Phi^{-1}(q).$$

Since the normal distribution is continuous, Proposition 3.5 gives

$$\mathrm{CTE}_q(L) = \mathbb{E}\left[L \mid L > \mathrm{VaR}_q(L)\right] = \mu + \sigma\, \mathbb{E}\left[Z \mid Z > z_q\right].$$

Now

$$\mathbb{E}\left[Z \mid Z > z_q\right] = \frac{1}{1-q} \int_{z_q}^{\infty} z\varphi(z)\,\mathrm{d}z = \frac{1}{1-q} \int_{z_q}^{\infty} -\varphi'(z)\,\mathrm{d}z = \frac{\varphi(z_q)}{1-q},$$

because $\varphi'(z) = -z\varphi(z)$ and $\varphi(z) \to 0$ as $z \to \infty$. Therefore,

$$\mathrm{CTE}_q(L) = \mu + \sigma \frac{\varphi(\Phi^{-1}(q))}{1-q}.$$

□

Example 3.9 If $L \sim \mathrm{Pareto}(\gamma, \theta)$ with $\gamma > 1$, then for $q \in (0,1)$,

$$\mathrm{VaR}_q(L) = \theta \left((1-q)^{-1/\gamma} - 1 \right) \qquad \text{and} \qquad \mathrm{CTE}_q(L) = \frac{\gamma\theta}{\gamma - 1}(1-q)^{-1/\gamma} - \theta.$$

Proof. From the distribution function

$$F_L(x) = 1 - \left(\frac{\theta}{x + \theta} \right)^{\gamma},$$

the q-quantile is obtained by solving

$$q = 1 - \left(\frac{\theta}{x+\theta}\right)^{\gamma}.$$

This gives

$$\frac{\theta}{x+\theta} = (1-q)^{1/\gamma} \qquad \text{and} \qquad x+\theta = \theta(1-q)^{-1/\gamma},$$

and hence

$$\mathrm{VaR}_q(L) = \theta\left((1-q)^{-1/\gamma} - 1\right).$$

Let $d = \mathrm{VaR}_q(L)$. Since the Pareto distribution is continuous, Proposition 3.5 gives

$$\mathrm{CTE}_q(L) = \mathbb{E}\left[L \mid L > d\right].$$

For $y \geq 0$,

$$\mathbb{P}(L - d > y \mid L > d) = \frac{\mathbb{P}(L > d+y)}{\mathbb{P}(L > d)} = \frac{\left(\frac{\theta}{d+y+\theta}\right)^{\gamma}}{\left(\frac{\theta}{d+\theta}\right)^{\gamma}} = \left(\frac{d+\theta}{d+y+\theta}\right)^{\gamma}.$$

Therefore $L - d \mid L > d \sim \mathrm{Pareto}(\gamma, d+\theta)$ in the same parameterization. Its mean is $(d+\theta)/(\gamma-1)$, so

$$\mathrm{CTE}_q(L) = d + \frac{d+\theta}{\gamma-1}.$$

Since $d + \theta = \theta(1-q)^{-1/\gamma}$, this becomes

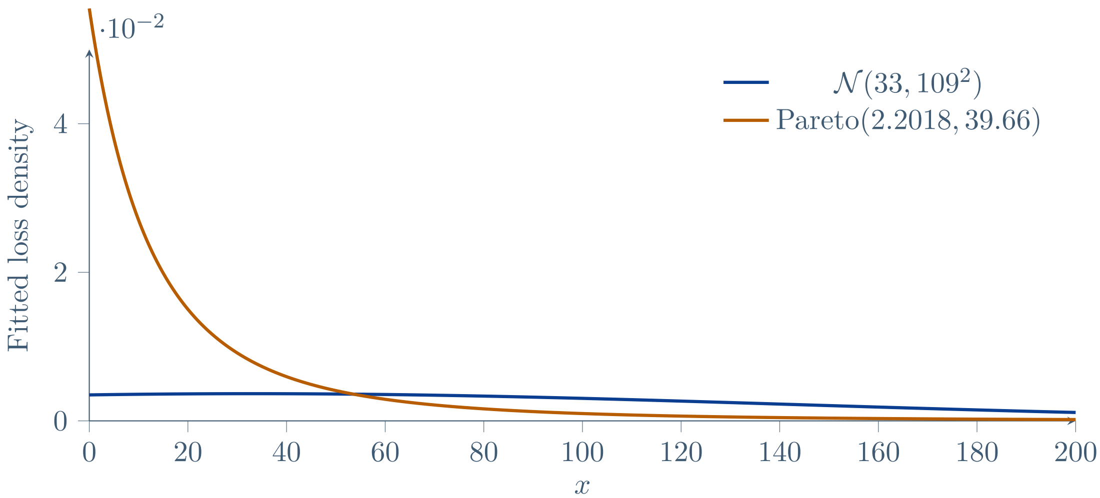
$$\mathrm{CTE}_q(L) = \theta\left((1-q)^{-1/\gamma} - 1\right) + \frac{\theta(1-q)^{-1/\gamma}}{\gamma-1} = \frac{\gamma\theta}{\gamma-1}(1-q)^{-1/\gamma} - \theta.$$

$\square$

The Pareto distribution is much heavier-tailed than the normal distribution, so its VaR and CTE are much larger at high quantiles. For example, suppose that we use empirical losses to estimate parameters by moment matching. If the sample mean is 33 and the sample standard deviation is 109, then the fitted normal distribution has $\mu = 33$ and $\sigma = 109$, while the fitted Pareto distribution has $\theta = 39.66$ and $\gamma = 2.2018$.

Matching only moments of low order does not determine tail risk. If $L \sim \mathcal{N}(33, 109^2)$, then $\mathrm{VaR}_{0.999}(L) \approx 370$ and $\mathrm{CTE}_{0.999}(L) \approx 400$, while if $L \sim \mathrm{Pareto}(2.2018, 39.66)$, then $\mathrm{VaR}_{0.999}(L) \approx$

874 and $\mathrm{CTE}_{0.999}(L) \approx 1635$. The fitted normal and Pareto distributions have similar central behaviour but very different extreme losses, so tail modelling matters far more than moment matching when capital at a high quantile is the objective.



VaR is widely used in banking and in Solvency II, while CTE is especially common in insurance practice in Canada and the United States. Basel regulation has also moved away from VaR towards CTE. Which risk measure is better?

Coherence

Definition 3.10 A risk measure ρ on a linear space of integrable losses is called **coherent** if it satisfies the following four properties:

1. **Translation invariance**: For any constant c, $\rho(L + c) = \rho(L) + c$.
2. **Positive homogeneity**: For any $\lambda > 0$, $\rho(\lambda L) = \lambda\rho(L)$.
3. **Monotonicity**: If $\mathbb{P}(L_1 \leq L_2) = 1$, then $\rho(L_1) \leq \rho(L_2)$.
4. **Subadditivity**: For any L_1 and L_2, $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$.

Each property has a natural interpretation. Translation invariance means that adding a certain amount to the loss increases the risk by the same amount. Positive homogeneity means that scaling the loss by a positive factor scales the risk by the same factor. Monotonicity means that if one loss is always less than or equal to another, then its risk should be less than or equal to the other's risk. Subadditivity means that diversification should not increase risk: the risk of a combined loss should be no greater than the sum of the risks of the individual losses.

Proposition 3.11 VaR satisfies translation invariance, positive homogeneity, and monotonicity.

Proof. Let

$$\mathrm{VaR}_\alpha(L) = \inf\{x : \mathbb{P}(L \le x) \ge \alpha\}.$$

For translation invariance, let c be a constant. Then

$$\begin{aligned}
\mathrm{VaR}_\alpha(L+c) &= \inf\{x : \mathbb{P}(L + c \le x) \ge \alpha\} \\
&= \inf\{x : \mathbb{P}(L \le x - c) \ge \alpha\} \\
&= c + \inf\{y : \mathbb{P}(L \le y) \ge \alpha\} \\
&= c + \mathrm{VaR}_\alpha(L).
\end{aligned}$$

For positive homogeneity, let $\lambda > 0$. Then

$$\begin{aligned}
\mathrm{VaR}_\alpha(\lambda L) &= \inf\{x : \mathbb{P}(\lambda L \le x) \ge \alpha\} \\
&= \inf\{x : \mathbb{P}(L \le x/\lambda) \ge \alpha\} \\
&= \lambda \inf\{y : \mathbb{P}(L \le y) \ge \alpha\} \\
&= \lambda \, \mathrm{VaR}_\alpha(L).
\end{aligned}$$

For monotonicity, suppose $L_1 \le L_2$ almost surely. Then, for every x,

$$\{L_2 \le x\} \subseteq \{L_1 \le x\},$$

so

$$\mathbb{P}(L_1 \le x) \ge \mathbb{P}(L_2 \le x).$$

If x belongs to the set

$$\{x : \mathbb{P}(L_2 \le x) \ge \alpha\},$$

then it also belongs to the set

$$\{x : \mathbb{P}(L_1 \le x) \ge \alpha\}.$$

Therefore the second set lies weakly to the left of the first, and hence

$$\mathrm{VaR}_\alpha(L_1) \leq \mathrm{VaR}_\alpha(L_2).$$

□

VaR is not subadditive in general, so it is not coherent in the sense of Definition 3.10. To see this, let X and Y be independent and identically distributed with

$$\mathbb{P}(X = 0) = 0.96 \qquad \text{and} \qquad \mathbb{P}(X = 1) = 0.04,$$

and take $\alpha = 0.95$. Then

$$\mathrm{VaR}_{0.95}(X) = \mathrm{VaR}_{0.95}(Y) = 0,$$

because

$$\mathbb{P}(X \leq 0) = 0.96 > 0.95.$$

However,

$$\mathbb{P}(X + Y = 0) = 0.96^2 = 0.9216,$$

so

$$\mathrm{VaR}_{0.95}(X + Y) = 1.$$

Thus

$$\mathrm{VaR}_{0.95}(X + Y) = 1 > 0 + 0 = \mathrm{VaR}_{0.95}(X) + \mathrm{VaR}_{0.95}(Y),$$

which violates subadditivity.

Proposition 3.12 CTE is coherent in the sense of Definition 3.10.

Proof. For translation invariance, let c be a constant. Since

$$\mathbb{P}(L + c \leq x) = \mathbb{P}(L \leq x - c),$$

the u-quantile of $L + c$ is $Q_u(L) + c$. Therefore

$$\mathrm{CTE}_\alpha(L + c) = \frac{1}{1-\alpha} \int_\alpha^1 Q_u(L + c)\,\mathrm{d}u = \frac{1}{1-\alpha} \int_\alpha^1 \big(Q_u(L) + c\big)\,\mathrm{d}u = \mathrm{CTE}_\alpha(L) + c.$$

For positive homogeneity, let $\lambda > 0$. Then the u-quantile of λL is $\lambda Q_u(L)$, so

$$\mathrm{CTE}_\alpha(\lambda L) = \frac{1}{1-\alpha}\int_\alpha^1 Q_u(\lambda L)\,\mathrm{d}u = \frac{1}{1-\alpha}\int_\alpha^1 \lambda Q_u(L)\,\mathrm{d}u = \lambda\,\mathrm{CTE}_\alpha(L).$$

For monotonicity, suppose $L_1 \le L_2$ almost surely. Then for every x,

$$\mathbb{P}(L_1 \le x) \ge \mathbb{P}(L_2 \le x),$$

so the quantiles satisfy $Q_u(L_1) \le Q_u(L_2)$ for every $u \in (0,1)$. Integrating gives

$$\mathrm{CTE}_\alpha(L_1) = \frac{1}{1-\alpha}\int_\alpha^1 Q_u(L_1)\,\mathrm{d}u \le \frac{1}{1-\alpha}\int_\alpha^1 Q_u(L_2)\,\mathrm{d}u = \mathrm{CTE}_\alpha(L_2).$$

It remains to prove subadditivity. Let X and Y be two integrable losses, and write $Z = X + Y$. We use an order-statistic representation of CTE: if $X_1, \dots, X_n$ is an independent sample from the distribution of X, with order statistics

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)},$$

then, almost surely,

$$\mathrm{CTE}_\alpha(X) = \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{j=n\alpha+1}^{n} X_{(j)},$$

which is an instance of a strong law of large numbers for upper order statistics. Its proof requires some care because the cutoff is an estimated quantile, and is beyond the scope of the course.

For notational simplicity, first consider sample sizes for which $n\alpha$ is an integer. For the remaining sample sizes, replacing $n\alpha$ by $\lfloor n\alpha \rfloor$ changes only the boundary order statistic and does not affect the limit. The same representation holds for Y and for Z. Now take an independent sample

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

from the joint distribution of (X, Y), and define

$$Z_j = X_j + Y_j, \quad j = 1, \dots, n.$$

Let

$$Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$$

be the order statistics of the sample $Z_1, \dots, Z_n$. Let A_n be the set of indices corresponding to the largest $n(1-\alpha)$ observations among the Z_j's. Then

$$\sum_{j=n\alpha+1}^{n} Z_{(j)} = \sum_{i\in A_n} Z_i = \sum_{i\in A_n} X_i + \sum_{i\in A_n} Y_i.$$

But the sum of any $n(1-\alpha)$ observations from the X_i's is at most the sum of the $n(1-\alpha)$ largest order statistics of the X_i's. Hence

$$\sum_{i\in A_n} X_i \le \sum_{j=n\alpha+1}^{n} X_{(j)}.$$

Similarly,

$$\sum_{i\in A_n} Y_i \le \sum_{j=n\alpha+1}^{n} Y_{(j)}.$$

Therefore,

$$\sum_{j=n\alpha+1}^{n} Z_{(j)} \le \sum_{j=n\alpha+1}^{n} X_{(j)} + \sum_{j=n\alpha+1}^{n} Y_{(j)}.$$

Dividing by $n(1-\alpha)$ and taking limits gives

$$\mathrm{CTE}_\alpha(Z) \le \mathrm{CTE}_\alpha(X) + \mathrm{CTE}_\alpha(Y).$$

Since $Z = X + Y$, this is precisely

$$\mathrm{CTE}_\alpha(X+Y) \le \mathrm{CTE}_\alpha(X) + \mathrm{CTE}_\alpha(Y).$$

Thus CTE is subadditive. Since all four properties hold, CTE is coherent. $\square$

But does subadditivity actually matter in practice, or is this just mathematical claptrap?

It can matter a great deal! In banking risk management, a VaR limit can sometimes be **gamed**: because VaR looks only at a single quantile and ignores the severity of losses beyond that quantile, a trader may be able to satisfy a VaR constraint while nevertheless shifting the portfolio toward positions with more severe tail losses. In that sense, a VaR restriction may actually incentivize riskier trading rather than safer trading. This concern is developed formally by Basak and Shapiro (2001), who show that risk constraints based on VaR can distort portfolio choice toward undesirable exposures in the tail.

There is also a broader concern across the market. If many institutions are managed using similar VaR limits, then those institutions may respond to changing market conditions in similar ways, for example by cutting positions simultaneously when measured risk rises. Such common reactions can amplify market stress, worsen liquidity conditions, and increase **systemic risk**. Danielsson et al. (2001) argue that this is one reason subadditivity is not merely an abstract mathematical nicety: the choice of risk measure can affect incentives, trading behaviour, and financial stability.

Estimating VaR and CTE

Risk measures may be estimated empirically from historical data, by Monte Carlo simulation, or by fitting a parametric model to the loss distribution. Each approach introduces estimation uncertainty.

If a Monte Carlo sample of size N is generated and arranged as

$$L_{(1)} \leq L_{(2)} \leq \cdots \leq L_{(N)},$$

then the empirical α-VaR is the generalized inverse of the empirical distribution function,

$$\widehat{Q}_\alpha = L_{(\lceil N\alpha \rceil)}.$$

Other interpolation conventions give slightly different finite-sample values, although they are asymptotically equivalent. For CTE, let $k = \lfloor N\alpha \rfloor$ and $\theta = N\alpha - k$. Integrating the empirical quantile function gives the precise estimator

$$\widehat{\mathrm{CTE}}_\alpha = \frac{(1-\theta)L_{(k+1)} + \displaystyle\sum_{j=k+2}^{N} L_{(j)}}{N(1-\alpha)}.$$

When $N\alpha$ is an integer, this reduces to the ordinary average of the largest $N(1-\alpha)$ observations.

For an iid sample, if the distribution has a density f that is continuous and strictly positive at the unique true quantile Q_α, then asymptotically

$$\mathrm{se}(\widehat{Q}_\alpha) = \frac{\sqrt{\alpha(1-\alpha)}}{\sqrt{N}\, f(Q_\alpha)}.$$

An approximate 95% confidence interval is therefore

$$\widehat{Q}_\alpha \pm 1.96\, \mathrm{se}(\widehat{Q}_\alpha).$$

In practice $f(Q_\alpha)$ must itself be estimated. This approximation can be unreliable in small samples or deep in the tail, and density estimation near the tail is often unstable.

A nonparametric alternative uses order statistics directly and avoids estimating the density at the quantile. Define indicators

$$\mathbb{1}_j = \begin{cases} 1, & L_j \le Q_\alpha, \\ 0, & L_j > Q_\alpha, \end{cases} \qquad j = 1, \dots, N,$$

and let

$$M = \sum_{j=1}^{N} \mathbb{1}_j.$$

Assume here that F_L is continuous at Q_α, so that $F_L(Q_\alpha) = \alpha$. Then $M \sim \mathrm{Bin}(N, \alpha)$. If integers $1 \le s < r \le N$ are chosen so that

$$\mathbb{P}(s \le M < r) = \sum_{j=s}^{r-1} \binom{N}{j} \alpha^j (1-\alpha)^{N-j} = q,$$

then this event is equivalent to

$$L_{(s)} \le Q_\alpha < L_{(r)},$$

so

$$\mathbb{P}\big(L_{(s)} \le Q_\alpha < L_{(r)}\big) = q.$$

This gives an exact nonparametric confidence interval for the true VaR in terms of order statistics. A convenient symmetric approximation chooses integer indices near

$$s \approx N\alpha - k \qquad \text{and} \qquad r \approx N\alpha + k.$$

Since

$$\mathbb{E}[M] = N\alpha \qquad \text{and} \qquad \mathrm{Var}(M) = N\alpha(1-\alpha),$$

the normal approximation to the binomial gives

$$k \approx z_{(1+q)/2} \sqrt{N\alpha(1-\alpha)}.$$

Hence an approximate symmetric q-confidence interval is

$$\big(L_{(\lfloor N\alpha - k \rfloor)}, L_{(\lceil N\alpha + k \rceil)}\big), \qquad \text{where} \qquad k = z_{(1+q)/2} \sqrt{N\alpha(1-\alpha)},$$

provided that the resulting indices lie between 1 and N and that both $N\alpha$ and $N(1-\alpha)$ are large enough for the normal approximation to be reliable.

Repeated simulation is another practical approach. We generate several independent Monte Carlo samples, compute the VaR estimate from each, and then use the sample standard deviation of those estimates as an estimated standard error.

For CTE, we can also study estimation uncertainty more explicitly. Assume that the sample is iid, that L has a finite second moment, and that its distribution is continuous at Q_α. A useful starting point is the variance decomposition

$$\mathrm{Var}\big(\widehat{\mathrm{CTE}}_\alpha\big) = \mathbb{E}\left[\mathrm{Var}\big(\widehat{\mathrm{CTE}}_\alpha \mid \widehat{Q}_\alpha\big)\right] + \mathrm{Var}\left(\mathbb{E}\big[\widehat{\mathrm{CTE}}_\alpha \mid \widehat{Q}_\alpha\big]\right).$$

Carrying this calculation through yields the large sample approximation

$$\mathrm{Var}\left(\widehat{\mathrm{CTE}}_\alpha\right) \approx \frac{\mathrm{Var}(L \mid L > Q_\alpha) + \alpha(\mathrm{CTE}_\alpha - Q_\alpha)^2}{N(1-\alpha)}.$$

The first term reflects the variability of losses inside the tail beyond the quantile. The second term reflects the additional uncertainty created by estimating the tail cutoff itself.

In practice, the unknown quantities are replaced by their empirical analogues. For simplicity, suppose here that $N(1-\alpha)$ is an integer of at least two. If

$$s^2 = \frac{1}{N(1-\alpha)-1} \sum_{j=N\alpha+1}^{N} \big(L_{(j)} - \widehat{\mathrm{CTE}}_\alpha\big)^2$$

is the sample variance of the losses lying in the estimated tail, then a practical estimator of the variance of $\widehat{\mathrm{CTE}}_\alpha$ is

$$\widehat{\mathrm{Var}}\big(\widehat{\mathrm{CTE}}_\alpha\big) = \frac{s^2 + \alpha\big(\widehat{\mathrm{CTE}}_\alpha - \widehat{Q}_\alpha\big)^2}{N(1-\alpha)}.$$

This makes it possible to report an estimated standard error for the CTE estimator in the same way as for VaR. With a noninteger tail count, the boundary observation receives the same fractional weight as in the empirical CTE formula above.

VaR and CTE are not the only risk measures used in practice. Variance and **semivariance** are also used as measures of risk. Semivariance focuses only on downside deviations below a benchmark such as the mean, so it is often more appealing than full variance when upside volatility is not viewed as harmful. These measures are useful in portfolio theory and in describing variability, especially downside variability in the case of semivariance. However, they do not directly provide capital requirements in the same way as measures based on tails, such as VaR and CTE, and they do not target the far tail as directly as measures of extreme quantiles do.

Lecture 4 — Tuesday, October 06, 2015

In banking, market risk is often measured over very short horizons such as one day or ten days. If a portfolio has value V_t at time t and the VaR horizon is h, the loss random variable is

$$L_t = V_t - V_{t+h}.$$

We may also work with the centered loss

$$L_t^{\text{mean}} = \mathbb{E}_t\left[V_{t+h}\right] - V_{t+h}.$$

For small horizons it is often assumed that $\mathbb{E}_t\left[V_{t+h}\right] \approx V_t$, so the two definitions lead to essentially the same VaR.

If the portfolio return over $(t, t+h)$ is

$$R_V = \frac{V_{t+h}}{V_t} - 1,$$

then

$$V_{t+h} = V_t(1 + R_V),$$

and hence

$$L_t = V_t - V_t(1 + R_V) = -R_V V_t.$$

If the return distribution function is continuous and strictly increasing, quantiles reverse cleanly under a change of sign, and therefore

$$Q_\alpha(L_t) = Q_\alpha(-R_V)V_t = -Q_{1-\alpha}(R_V)V_t.$$

VaR in Banking

Assume a single stock price $S(t)$ has return

$$R = \frac{S(t+h) - S(t)}{S(t)},$$

and that

$$R \sim \mathcal{N}(0, h\sigma^2).$$

The loss on one share over the horizon is

$$L = -RS(t).$$

Since R is normal,

$$Q_{1-\alpha}(R) = z_{1-\alpha}\sigma\sqrt{h},$$

so

$$\mathrm{VaR}_\alpha(L) = -Q_{1-\alpha}(R)S(t) = -z_{1-\alpha}\sigma\sqrt{h}\,S(t) = z_\alpha\sigma\sqrt{h}\,S(t),$$

where $z_\alpha = \Phi^{-1}(\alpha)$.

Next, consider a portfolio with positions a_j in stocks $j = 1, \dots, n$. Let

$$R_j = \frac{S_j(t+h) - S_j(t)}{S_j(t)}$$

denote the return on stock j over the VaR horizon. Assume the return vector $\boldsymbol{R} = (R_1, \dots, R_n)$ is multivariate normal with

$$R_j \sim \mathcal{N}(0, \sigma_j^2 h) \qquad \text{and} \qquad \mathrm{Cov}(R_j, R_k) = \rho_{jk}\sigma_j\sigma_k h.$$

The portfolio value at time t is

$$V(t) = \sum_{j=1}^n a_j S_j(t),$$

and the loss is

$$L_V = V(t) - V(t+h) = -\sum_{j=1}^n a_j\big(S_j(t+h) - S_j(t)\big) = -\sum_{j=1}^n a_j S_j(t) R_j.$$

Because L_V is a linear combination of jointly normal random variables, it is normal. Its mean is zero, and its variance is

$$\mathrm{Var}(L_V) = h\sum_{j=1}^n\sum_{k=1}^n a_j S_j(t) a_k S_k(t)\rho_{jk}\sigma_j\sigma_k.$$

If

$$\sigma_L^2 = \sum_{j=1}^n\sum_{k=1}^n a_j S_j(t) a_k S_k(t)\rho_{jk}\sigma_j\sigma_k,$$

then

$$\mathrm{VaR}_\alpha(L_V) = z_\alpha\sigma_L\sqrt{h}.$$

The previous approach requires linear dependence on the underlying assets, which is not always appropriate, especially when the portfolio contains derivatives. The assumption of normality is also often unrealistic, and the estimation of many correlations can be difficult in any case. Instead, we can use a first-order Taylor expansion to approximate the loss, which leads to the **delta-VaR** approach.

Delta-VaR

Suppose the portfolio value depends on time and on risk factors $X_1, \dots, X_n$. A first-order Taylor expansion gives

$$V(t+h, X(t+h)) \approx V(t, X(t)) \; + \; \frac{\partial V}{\partial t} h \; + \; \sum_{j=1}^{n} \frac{\partial V}{\partial x_j} \big(X_j(t+h) - X_j(t) \big).$$

Hence the loss is approximately

$$L_V = V(t) - V(t+h) \approx -\frac{\partial V}{\partial t} h - \sum_{j=1}^{n} \frac{\partial V}{\partial x_j} \big(X_j(t+h) - X_j(t) \big).$$

If some factors are modelled as relative returns and others as absolute changes, this becomes

$$L_V \approx -\frac{\partial V}{\partial t} h - \sum_{j=1}^{m} X_j \frac{\partial V}{\partial x_j} R_j - \sum_{j=m+1}^{n} \frac{\partial V}{\partial x_j} \Delta X_j.$$

Under multivariate normality, the approximate loss is again normal and VaR can be computed in closed form.

Historical Simulation and Monte Carlo Methods

Historical simulation takes observed past changes in the risk factors,

$$\Delta X_{t-k} = X(t-k+1) - X(t-k), \quad k = 1, \dots, T,$$

and treats them as an empirical distribution with probabilities $1/T$. Revaluing the portfolio under each scenario gives losses

$$L_k = V(t, X_0) - V(t+h, X_0 + \Delta X_k), \quad k = 1, \dots, T.$$

These losses are then treated in the same way as simulated losses from a Monte Carlo simulation.

Historical simulation is intuitive, simple to implement, and nonparametric, so it can reflect skewness and heavy tails. Its weaknesses are that it must be modified to account for seasonal trends, it is sensitive to **window effects** (e.g., when a large historical loss drops out of the time window used for the simulation), it takes a long time to account for structural changes in the market, and it is not clear how to incorporate new information that has not yet been observed in the historical data.

Monte Carlo simulation is more flexible. It can incorporate arbitrary portfolios, large numbers of scenarios, heavy tails, skewness, volatility clustering, and autocorrelation. Its drawbacks are that it is slower, harder to implement, and capable of hiding strong modelling assumptions inside a complicated simulation engine.

Backtesting

To backtest a VaR model, define the **violation indicators**

$$\mathbb{1}_t = \begin{cases} 1, & L_t > \mathrm{VaR}_{\alpha,t}, \\ 0, & L_t \leq \mathrm{VaR}_{\alpha,t}\,. \end{cases}$$

Under correct conditional coverage, meaning that the VaR forecast is the conditional α-quantile given the information available at time $t-1$ and the conditional distribution has no atom at that quantile, the indicators behave like independent Bernoulli random variables with success probability

$$p = 1 - \alpha.$$

If T is the sample size and

$$T_1 = \sum_{t=1}^{T} \mathbb{1}_t$$

is the number of violations, then the usual MLE

$$\widehat{p} = \frac{T_1}{T}$$

estimates p, with

$$\mathbb{E}\left[\widehat{p}\right] = p \qquad \text{and} \qquad \mathrm{Var}(\widehat{p}) = \frac{p(1-p)}{T}.$$

An asymptotic mean test is based on

$$M_T = \frac{\sqrt{T}(\widehat{p} - p)}{\sqrt{p(1-p)}} \approx \mathcal{N}(0,1).$$

We can also use a likelihood ratio test. The Bernoulli likelihood is

$$L(\pi) = (1-\pi)^{T_0}\pi^{T_1},$$

where $T_0 = T - T_1$. The unrestricted maximizer is $\widehat{p} = T_1/T$, and the likelihood ratio statistic is

$$-2\log\left(\frac{L(p)}{L(\widehat{p})}\right) \approx \chi^2_1.$$

For small samples, critical values obtained by simulation may be preferable.

The violations should also show no time pattern. For example, if a model ignores volatility clustering, then VaR violations will often occur in clusters. This can be checked using the autocorrelation function

$$\gamma_k = \mathrm{Corr}(\mathbb{1}_t, \mathbb{1}_{t-k}).$$

Lecture 5 — Tuesday, October 13, 2015

3. Frequency–Severity Analysis

Many informal descriptions of quantitative risk say that

$$\text{Risk} = \text{Probability} \times \text{Severity}.$$

Under restrictive assumptions this leads to an expected loss, but expected value alone is only a limited summary of risk. A fuller analysis models both the number of loss events and their severities.

The basic assumption is that the number of losses and the severity of each loss can be analysed separately, an approach most appropriate for idiosyncratic, diversifiable, insurable losses such as many operational and environmental losses. Frequency is also usually assumed to be independent of individual loss severity; this last assumption strikes us as being incredibly dumb, but we press onward.

Aggregate Loss Models and Insurance

Let S be the aggregate loss in a given period, let N be the number of loss events, and let Y_j be the severity of the jth event. Then

$$S = \begin{cases} \sum_{j=1}^{N} Y_j, & N \geq 1, \\ 0, & N = 0. \end{cases}$$

The standard assumptions are that N is independent of the sequence $\{Y_j\}$ and that the severities are independent and identically distributed.

Two useful identities are

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid N]]$$

and

$$\operatorname{Var}(S) = \mathbb{E}[\operatorname{Var}(S \mid N)] + \operatorname{Var}(\mathbb{E}[S \mid N]).$$

Let X be a random variable. Recall that if $M_X(z) = \mathbb{E}\left[e^{zX}\right]$ is the moment generating function, $P_X(z) = \mathbb{E}\left[z^X\right]$ is the probability generating function, and $C_X(z) = \log M_X(z)$ is the cumulant generating function, then

$$M_S(z) = \mathbb{E}\left[e^{zS}\right] = P_N(M_Y(z)).$$

Frequency Models

The frequency random variable N takes values in $\{0, 1, 2, \dots\}$, so common models are counting distributions such as the binomial, Poisson, and negative binomial. Let us recall several facts about these distributions.

If $N \sim \operatorname{Bin}(m, q)$, then

$$\begin{aligned} \mathbb{P}(N = k) &= \binom{m}{k} q^k (1-q)^{m-k} \\ \mathbb{E}[N] &= mq \\ \operatorname{Var}(N) &= mq(1-q) \\ \mathbb{E}[(N - \mathbb{E}[N])^3] &= mq(1-q)(1-2q) \\ M_N(z) &= \mathbb{E}[e^{zN}] = (1 + q(e^z - 1))^m \\ P_N(t) &= \mathbb{E}[t^N] = (1 + q(t-1))^m. \end{aligned}$$

If $N \sim \text{Poisson}(\lambda)$, then

$$\begin{aligned}
\mathbb{P}(N = k) &= e^{-\lambda}\frac{\lambda^k}{k!} \\
\mathbb{E}[N] &= \lambda \\
\text{Var}(N) &= \lambda \\
\mathbb{E}[(N - \mathbb{E}[N])^3] &= \lambda \\
M_N(z) &= \mathbb{E}[e^{zN}] = \exp(\lambda(e^z - 1)) \\
P_N(t) &= \mathbb{E}[t^N] = e^{\lambda(t-1)}.
\end{aligned}$$

The Poisson model is very convenient analytically, particularly when aggregating or disaggregating risks. If independent Poisson random variables satisfy

$$N_j \sim \text{Poisson}(\lambda_j), \quad j = 1, \ldots, r,$$

then

$$\sum_{j=1}^{r} N_j \sim \text{Poisson}\left(\sum_{j=1}^{r} \lambda_j\right).$$

This result is often used to simulate or aggregate risks from different sources. Conversely, if $N \sim \text{Poisson}(\lambda)$ and each event is independently assigned to category j with probability q_j, then the number of events in category j is

$$N_j \sim \text{Poisson}(\lambda q_j),$$

which is useful for *disaggregating* risks.

If $N \sim \text{NegBin}(r, \beta)$, then

$$\begin{aligned}
\mathbb{P}(N = k) &= \frac{\Gamma(k + r)}{k!\Gamma(r)}\left(\frac{\beta}{1 + \beta}\right)^k\left(\frac{1}{1 + \beta}\right)^r \\
\mathbb{E}[N] &= r\beta \\
\text{Var}(N) &= r\beta(1 + \beta) \\
\mathbb{E}[(N - \mathbb{E}[N])^3] &= r\beta(1 + \beta)(1 + 2\beta) \\
M_N(z) &= \mathbb{E}[e^{zN}] = \left(\frac{1}{1 - \beta(e^z - 1)}\right)^r \\
P_N(t) &= \mathbb{E}[t^N] = \left(\frac{1}{1 - \beta(t - 1)}\right)^r.
\end{aligned}$$

The negative binomial distribution is flexible. For example, it allows overdispersion, meaning that the variance exceeds the mean.

The $(a, b, 0)$ and $(a, b, 1)$ Distributions

Definition 5.1 If N is a discrete random variable whose probabilities $p_k = \mathbb{P}(N = k)$ satisfy the **Panjer recursion formula**

$$p_{k+1} = \left(a + \frac{b}{k+1}\right) p_k, \quad k = 0, 1, 2, \ldots,$$

then N has an $\boldsymbol{(a, b, 0)}$ **distribution**.

Proposition 5.2 The Poisson, binomial, and negative binomial distributions are the only $(a, b, 0)$ distributions.

Proof. Since the recursion would otherwise force every probability to vanish, we must have $p_0 > 0$. We also have $a + b = p_1/p_0 \geq 0$. If $a + b = 0$, then $p_k = 0$ for every $k \geq 1$, so $N = 0$ almost surely; this is the degenerate limiting case of the Poisson or binomial family. We therefore assume $a + b > 0$ and analyze three cases.

Case $a = 0$. We have

$$p_1 = p_0 b, \qquad p_2 = p_1 \frac{b}{2} = \frac{p_0 b^2}{2!}, \qquad \text{and} \qquad p_3 = p_2 \frac{b}{3} = p_0 \frac{b^3}{3!},$$

and it follows by induction that $p_k = p_0 b^k/k!$, and for a valid probability distribution we must have $p_0 = e^{-b}$. Replacing b with λ gives the Poisson(λ) distribution.

Case $a > 0$. Let $r = (a + b)/a > 0$. Then

$$\frac{p_{k+1}}{p_k} = a\left(1 + \frac{r-1}{k+1}\right) = a\frac{k+r}{k+1}.$$

Iterating the recursion gives

$$p_k = p_0 a^k \prod_{j=0}^{k-1} \frac{j+r}{j+1} = p_0 a^k \frac{\Gamma(k+r)}{\Gamma(r)\, k!}, \quad k \geq 0.$$

For the probabilities to have a finite sum, we must have $a \in (0, 1)$. Using the generalized binomial expansion,

$$\sum_{k=0}^{\infty} \frac{\Gamma(k+r)}{\Gamma(r)\,k!} a^k = (1-a)^{-r},$$

so normalization requires $p_0 = (1-a)^r$. Therefore

$$p_k = (1-a)^r \frac{\Gamma(k+r)}{\Gamma(r)\,k!} a^k.$$

This is the negative binomial distribution with shape r and scale parameter

$$\beta = \frac{a}{1-a}.$$

Equivalently, $a = \beta/(1+\beta)$ and $b = (r-1)\beta/(1+\beta)$.

Case $a < 0$. Since $a+b > 0$, the recursion coefficients are initially positive. Because they decrease to $a < 0$, nonnegative probabilities are possible only if the support terminates at an integer $m \geq 1$ satisfying

$$a + \frac{b}{m+1} = 0.$$

Then $p_{m+1} = 0$, and the recursion gives $p_k = 0$ for every $k \geq m+1$. Solving for b gives $b = -a(m+1)$, so for $0 \leq k < m$ we have

$$\frac{p_{k+1}}{p_k} = a\left(1 - \frac{m+1}{k+1}\right) = (-a)\frac{m-k}{k+1}.$$

Let $q = -a/(1-a)$, so that $0 < q < 1$ and $-a = q/(1-q)$. Iterating the recursion gives

$$p_k = p_0 \binom{m}{k} \left(\frac{q}{1-q}\right)^k, \quad 0 \leq k \leq m.$$

Summing from $k = 0$ to m yields

$$1 = \sum_{k=0}^{m} p_k = p_0 \sum_{k=0}^{m} \binom{m}{k} \left(\frac{q}{1-q}\right)^k = p_0 \left(1 + \frac{q}{1-q}\right)^m = p_0(1-q)^{-m},$$

so $p_0 = (1-q)^m$ and hence

$$p_k = \binom{m}{k} q^k (1-q)^{m-k}, \quad 0 \leq k \leq m.$$

This is the $\mathrm{Bin}(m, q)$ distribution. $\square$

The proof above provides a and b for each of these three distributions:

$$\text{Poisson: } a = 0,\ b = \lambda,$$

$$\text{Binomial: } a = -\frac{q}{1-q},\ b = \frac{(m+1)q}{1-q},$$

$$\text{Negative binomial: } a = \frac{\beta}{1+\beta},\ b = \frac{(r-1)\beta}{1+\beta}.$$

Definition 5.3 If N is a discrete random variable whose probabilities $p_k = \mathbb{P}(N = k)$ satisfy

$$p_{k+1} = \left(a + \frac{b}{k+1}\right) p_k, \quad k = 1, 2, \ldots,$$

then N has an $\boldsymbol{(a, b, 1)}$ **distribution**.

The difference is that the recursion is imposed only for $k \geq 1$. This allows the probability of zero loss to be modified, including zero-truncated cases with $p_0 = 0$.

Severity Models

Severity is the loss amount conditional on a loss event having occurred. It is usually modelled by a continuous nonnegative distribution and is often positively skewed and heavy-tailed.

If $X \sim \text{Exp}(\theta)$ with $\theta > 0$, using the scale parameterization, then

$$\begin{aligned}
f(x) &= \frac{1}{\theta} e^{-x/\theta}, \quad x > 0, \\
F(x) &= 1 - e^{-x/\theta}, \\
\mathbb{E}[X] &= \theta, \\
\text{Var}(X) &= \theta^2, \\
M_X(z) = \mathbb{E}[e^{zX}] &= \frac{1}{1-\theta z}, \quad z < 1/\theta.
\end{aligned}$$

The exponential distribution is convenient and tractable. However, it has a constant hazard rate and relatively light tails, so it is often not a good fit for loss severity.

If $Y \sim \text{Exp}(1)$, then

$$X = \theta Y^{1/\tau} \sim \text{Weibull}(\theta, \tau), \quad \theta > 0 \qquad \text{and} \qquad \tau > 0,$$

with

$$
\begin{aligned}
f(x) &= \frac{\tau}{\theta}\left(\frac{x}{\theta}\right)^{\tau-1} e^{-(x/\theta)^\tau}, \quad x > 0,\\
F(x) &= 1 - e^{-(x/\theta)^\tau},\\
\mathbb{E}[X] &= \theta\Gamma\left(1+\frac{1}{\tau}\right).
\end{aligned}
$$

It has heavier tails than the exponential distribution when $\tau < 1$ and lighter tails when $\tau > 1$.

If $X \sim \text{Gamma}(\alpha, \beta)$ with $\alpha > 0$ and $\beta > 0$, then

$$
\begin{aligned}
f(x) &= \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}, \quad x > 0,\\
F(x) &= \frac{1}{\Gamma(\alpha)}\gamma\left(\alpha, \frac{x}{\beta}\right),\\
\mathbb{E}[X] &= \alpha\beta,\\
\text{Var}(X) &= \alpha\beta^2.
\end{aligned}
$$

The gamma distribution is flexible and tractable, but it has relatively light tails.

If $X \sim \text{Pareto}(\alpha, \theta)$ with $\alpha > 0$ and $\theta > 0$, then

$$
\begin{aligned}
f(x) &= \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, \quad x > 0,\\
F(x) &= 1 - \left(\frac{\theta}{x+\theta}\right)^\alpha,\\
\mathbb{E}[X] &= \frac{\theta}{\alpha-1}, \quad \alpha > 1,\\
\text{Var}(X) &= \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha > 2.
\end{aligned}
$$

The Pareto distribution is a standard model for heavy-tailed losses: its power-law tail decays like $x^{-\alpha}$ as $x \to \infty$, and smaller values of α produce heavier tails. In particular, the mean exists only for $\alpha > 1$ and the variance only for $\alpha > 2$.

If $X \sim \text{Lognormal}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$, then

$$
f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0,
$$

$$
\begin{aligned}
F(x) &= \Phi\left(\frac{\log x - \mu}{\sigma}\right), \\
\mathbb{E}\left[X^k\right] &= \exp\left(k\mu + \frac{k^2\sigma^2}{2}\right), \\
\operatorname{Var}(X) &= \exp\bigl(2\mu + \sigma^2\bigr)\left(e^{\sigma^2} - 1\right).
\end{aligned}
$$

Lognormal distributions are positively skewed and heavy-tailed, though not as heavy-tailed as Pareto distributions.

If $Y_1, \ldots, Y_N$ are independent and identically distributed with distribution F_Y, then the aggregate loss S has a **compound distribution** with frequency distribution F_N and severity distribution F_Y. The moment generating function of S is

$$
M_S(z) = \mathbb{E}\left[e^{zS}\right] = P_N(M_Y(z)).
$$

If $\mathbb{E}[Y] = \mu_Y$ and $\operatorname{Var}(Y) = \sigma_Y^2$, then

$$
\mathbb{E}[S \mid N] = N\mu_Y \qquad \text{and} \qquad \operatorname{Var}(S \mid N) = N\sigma_Y^2.
$$

Applying iterated expectation and variance yields

$$
\mathbb{E}[S] = \mu_Y \mathbb{E}[N]
$$

and

$$
\operatorname{Var}(S) = \sigma_Y^2 \mathbb{E}[N] + \mu_Y^2 \operatorname{Var}(N).
$$

In particular, if N is Poisson with parameter λ, then S has a **compound Poisson distribution** with rate λ, often written $S \sim \operatorname{CoPoi}(\lambda, F_Y)$. In that case

$$
\mathbb{E}[S] = \lambda\mathbb{E}[Y], \qquad \operatorname{Var}(S) = \lambda\mathbb{E}\left[Y^2\right], \qquad \text{and} \qquad \mathbb{E}\left[(S - \mathbb{E}[S])^3\right] = \lambda\mathbb{E}\left[Y^3\right].
$$

If $\mathbb{E}[Y^2] < \infty$, then as $\lambda \to \infty$ the central limit theorem suggests the approximation

$$
S \approx \mathcal{N}(\lambda\mu_Y, \lambda\mathbb{E}\left[Y^2\right]).
$$

Estimating the Aggregate Loss Distribution

We consider three main approaches: approximation by another distribution, **Panjer recursion**, and the **discrete Fourier transform**.

We use the following running example:

Example 5.4 Suppose that the frequency and severity distributions are

$$S \sim \text{CoPoi}(\lambda = 50, F) \qquad \text{and} \qquad Y \sim \text{Lognormal}(5, 1).$$

The severity moments are

$$\mathbb{E}[Y] = e^{5+1/2} = e^{11/2} \approx 244.692,$$

$$\mathbb{E}[Y^2] = e^{2\cdot 5+2} = e^{12} \approx 162754.791,$$

$$\text{and} \qquad \mathbb{E}[Y^3] = e^{3\cdot 5+9/2} = e^{39/2} \approx 294267566.042.$$

Hence

$$\mathbb{E}[S] = 50\mathbb{E}[Y] = 50e^{11/2} \approx 12234.6,$$

$$\text{Var}(S) = 50\mathbb{E}[Y^2] = 50e^{12} \approx 8137739.57 \qquad \text{and} \qquad \text{sd}(S) = \sqrt{8137739.57} \approx 2852.67,$$

and

$$\mathbb{E}\left[(S - \mathbb{E}[S])^3\right] = 50\mathbb{E}[Y^3] = 50e^{39/2} \approx 1.47134 \times 10^{10},$$

so the distribution is positively skewed.

Approximating S by a normal distribution with the same mean and variance works better when $\mathbb{E}[N]$ is large and we are interested in the center of the distribution, rather than the tails. The normal distribution has zero skewness, so it cannot capture the positive skewness of S. The normal approximation is therefore expected to underestimate tail probabilities. In this example, the normal approximation gives

$$S^* \sim \mathcal{N}(\mathbb{E}[S], \text{Var}(S)),$$

so

$$\mathbb{P}(S \le s) \approx \Phi\left(\frac{s - \mathbb{E}[S]}{\sqrt{\text{Var}(S)}}\right).$$

For $s = 20000$,

$$z = \frac{20000 - 12234.6}{2852.67} \approx 2.722,$$

and therefore

$$\mathbb{P}(S > 20000) \approx 1 - \Phi(2.722) \approx 0.00324.$$

For $s = 25000$,

$$z = \frac{25000 - 12234.6}{2852.67} \approx 4.475,$$

and therefore

$$\mathbb{P}(S > 25000) \approx 1 - \Phi(4.475) \approx 0.00000382.$$

These values are much too small in the tail because the normal approximation ignores positive skewness.

A **translated gamma approximation** writes

$$S \approx W + k \qquad \text{and} \qquad W \sim \text{Gamma}(\alpha, \beta),$$

and matches the first three central moments:

$$\alpha\beta + k = \mathbb{E}[S], \qquad \alpha\beta^2 = \text{Var}(S), \qquad \text{and} \qquad 2\alpha\beta^3 = \mathbb{E}\left[(S - \mathbb{E}[S])^3\right].$$

From the last two equations,

$$\beta = \frac{\mathbb{E}\left[(S - \mathbb{E}[S])^3\right]}{2\,\text{Var}(S)} \approx \frac{1.47134 \times 10^{10}}{2(8137739.57)} \approx 904.021,$$

and then

$$\alpha = \frac{\text{Var}(S)}{\beta^2} \approx \frac{8137739.57}{(904.021)^2} \approx 9.957.$$

Finally,

$$k = \mathbb{E}[S] - \alpha\beta \approx 12234.6 - (9.957)(904.021) \approx 3232.88.$$

Therefore

$$\mathbb{P}(S > s) \approx \mathbb{P}(W > s - k).$$

For $s = 20000$,

$$\mathbb{P}(S > 20000) \approx \mathbb{P}(W > 20000 - 3232.88) = \mathbb{P}(W > 16767.12) \approx 0.0110.$$

For $s = 25000$,

$$\mathbb{P}(S > 25000) \approx \mathbb{P}(W > 25000 - 3232.88) = \mathbb{P}(W > 21767.12) \approx 0.00039.$$

These values are much closer to the more accurate numerical methods.

The normal approximation is justified by the central limit theorem, but its tails may be too thin for Poisson and negative binomial frequency models (unless $\mathbb{E}[N]$ is large). The translated gamma approximation is more ad hoc; it is not justified by a limit theorem, but it can capture skewness and heavy tails. It often works quite well for the compound Poisson model, but it can perform poorly for other frequency models, such as the negative binomial. It has a positive skewness of $2/\sqrt{\alpha}$, so it cannot capture very heavy tails. However, the left tail of the translated gamma approximation is also often too thin, which can lead to underestimation of the probability of zero loss.

Panjer Recursion

Suppose the severity distribution is discrete with $f_k = \mathbb{P}(Y = k)$ and the aggregate loss probabilities are $g_s = \mathbb{P}(S = s)$. If N is an $(a, b, 0)$ frequency distribution, then

$$g_0 = P_N(f_0)$$

and, for $s \geq 1$,

$$g_s = \frac{1}{1 - af_0} \sum_{j=1}^{s} \left(a + \frac{bj}{s}\right) f_j g_{s-j}.$$

For a compound Poisson model this reduces to

$$g_0 = \exp\bigl(-\lambda(1 - f_0)\bigr)$$

and

$$g_s = \frac{\lambda}{s} \sum_{j=1}^{s} j f_j g_{s-j}, \quad s = 1, 2, \ldots.$$

For $(a, b, 1)$ distributions, the analogous recursion is

$$g_s = \frac{1}{1 - af_0} \left[\bigl(p_1 - (a + b)p_0\bigr) f_s + \sum_{j=1}^{s} \left(a + \frac{bj}{s}\right) f_j g_{s-j} \right].$$

The recursive structure makes these formulas computationally efficient: in practice, we stop at the largest aggregate loss of interest, and computing the sequence $g_1, \ldots, g_s$ requires $O(s^2)$ arithmetic operations rather than repeated convolutions. Panjer recursion is exact for discrete severity distributions, but it can also be used after discretizing a continuous severity distribution.

In the Poisson–lognormal running example, take $f_0 = 0$ and $g_0 = e^{-\lambda} = e^{-50}$. If we discretize the severity distribution in units from f_1 to $f_{2^{16}}$, then Panjer recursion gives

$$\mathbb{P}(S > 20000) \approx 0.0111 \qquad \text{and} \qquad \mathbb{P}(S > 25000) \approx 0.00070.$$

The computation takes mere seconds in R.

Discrete Fourier Transforms

The characteristic function approach uses

$$\phi_S(t) = \mathbb{E}\left[e^{itS}\right] = P_N(\phi_Y(t)).$$

After discretizing the severity distribution onto a grid $\{0, 1, \ldots, M-1\}$ with probabilities $f_0, f_1, \ldots, f_{M-1}$, define the discrete Fourier transform by

$$\widehat{f}(s) = \sum_{k=0}^{M-1} f_k e^{2\pi iks/M}, \quad s = 0, 1, \ldots, M-1.$$

The aggregate loss transform is then obtained pointwise from the probability generating function of the frequency distribution:

$$\widehat{g}(s) = P_N\big(\widehat{f}(s)\big), \quad s = 0, 1, \ldots, M-1,$$

where $\widehat{g}$ is the discrete transform of the aggregate loss probabilities $g_0, g_1, \ldots, g_{M-1}$. Finally, the aggregate loss probabilities are recovered from the inverse transform

$$g_s = \frac{1}{M} \sum_{k=0}^{M-1} \widehat{g}(k) e^{-2\pi isk/M}, \quad s = 0, 1, \ldots, M-1.$$

Thus the Fourier transform method reduces the compound distribution problem to three steps: transform the severity distribution, apply the frequency probability generating function pointwise, and invert. The **fast Fourier transform** makes these computations highly efficient.

Algorithm 1: FFT Method for Compound Distributions

Data: Discretized severity probabilities $f_0, f_1, \ldots, f_{M-1}$ on the grid $\{0, 1, \ldots, M-1\}$, where we typically choose $M = 2^m$.

Result: Aggregate loss probabilities $g_0, g_1, \ldots, g_{M-1}$ and cumulative probabilities $G(s) = \sum_{j=0}^{s} g_j$.

1 Form the vector $\boldsymbol{f} = (f_0, f_1, \ldots, f_{M-1})$ of discretized severity probabilities.

2 Compute the discrete Fourier transform $\widehat{f}(t)$ for $t = 0, 1, \ldots, M-1$ using the FFT.

3 Compute $\widehat{g}(t) = P_N\big(\widehat{f}(t)\big)$ pointwise for $t = 0, 1, \ldots, M-1$.

4 Apply the inverse FFT to $\widehat{g}$ to recover the aggregate loss probabilities g_s for $s = 0, 1, \ldots, M-1$.

5 Compute the cumulative sums $G(s) = \sum_{j=0}^{s} g_j$.

For a compound Poisson model, the corresponding R implementation is concise:

```
M <- length(f)
fhat <- fft(f, inverse = FALSE)
ghat <- exp(lambda * (fhat - 1))
g <- fft(ghat, inverse = TRUE) / M
g <- Re(g)
G <- cumsum(g)
```

The main numerical appeal is speed: direct convolution is slow, whereas the FFT computes these transforms in order $M \log M$ operations. We choose M large enough to capture the support of the discretized aggregate distribution, since otherwise wrap-around effects can contaminate the tail probabilities. A common improvement is **exponential tilting** before inversion, which stabilizes the far tail; it multiplies the severity probabilities by $e^{-\theta k}$ for some $\theta > 0$ before applying the FFT, then recovers the aggregate loss probabilities by multiplying by $e^{\theta s}$ after inversion.

For Example 5.4, the FFT gives

$$\mathbb{P}(S > 20000) \approx 0.0108 \qquad \text{and} \qquad \mathbb{P}(S > 25000) \approx 0.00068.$$

FFT is faster than Panjer recursion but may behave poorly in the far tail unless improvements such as exponential tilting are used.

Aggregate Loss Risk Measures

Under the normal approximation,

$$\mathrm{VaR}_\alpha(S) \approx \mathbb{E}[S] + z_\alpha \sqrt{\mathrm{Var}(S)}$$

and

$$\mathrm{CTE}_\alpha(S) \approx \mathbb{E}[S] \; + \; \sqrt{\mathrm{Var}(S)} \frac{\phi(z_\alpha)}{1-\alpha}.$$

For Example 5.4, this gives

$$\mathrm{VaR}_{0.95}(S) \approx 16927 \qquad \text{and} \qquad \mathrm{CTE}_{0.95}(S) \approx 18117.$$

Under the translated gamma approximation,

$$\mathrm{VaR}_q(S) \approx F_W^{-1}(q) + k$$

and

$$\mathrm{CTE}_q(S) \approx \mathrm{CTE}_q(W) + k,$$

where

$$\mathrm{CTE}_q(W) = \left(\frac{1 - F^*(Q_q)}{1-q}\right)\mathbb{E}[W]$$

and F^* is the distribution function of the size-biased distribution of W, which is $\Gamma(\alpha+1, \beta)$ in this case. The quantile function of the gamma distribution can be computed using numerical root-finding methods, and the CTE can be computed from the gamma distribution function and mean. For Example 5.4,

$$\mathrm{VaR}_{0.95}(S) \approx 17382 \qquad \text{and} \qquad \mathrm{CTE}_{0.95}(S) \approx 19096.$$

For Panjer recursion and the FFT, let Q_α be the smallest s such that $G(s) \geq \alpha$. Then

$$\mathrm{CTE}_\alpha(S) \approx \frac{Q_\alpha\big(G(Q_\alpha) - \alpha\big) + \displaystyle\sum_{s=Q_\alpha+1}^{M-1} s g_s}{1-\alpha}$$

for large M. This expression uses only the fraction of the probability mass at Q_α needed to form the upper $(1-\alpha)$ tail. For the example in (5.4), we have

$$\mathrm{VaR}_{0.95}(S) \approx 17342 \text{ and } 17286,$$

$$\mathrm{VaR}_{0.99}(S) \approx 20192 \text{ and } 20133,$$

$$\mathrm{CTE}_{0.95}(S) \approx 19101 \text{ and } 19079,$$

$$\mathrm{CTE}_{0.99}(S) \approx 21881 \text{ and } 21958.$$

The normal approximation understates the tail risk in this example because the aggregate distribution is positively skewed. The translated gamma approximation is materially better, and the numerical methods based on Panjer recursion and the FFT agree very closely. In practice, when Panjer and the FFT agree to this extent, we gain confidence that the discretization error is small.

Partial Insurance

Frequency–severity analysis is especially natural when losses are reasonably quantifiable, random from the policyholder's perspective, and sufficiently diversifiable for the insurer. Under **full insurance**, the insurer indemnifies the entire loss; under **partial insurance**, the policyholder retains a portion, reducing moral hazard and the expense of small claims while often making coverage more affordable.

Partial insurance can take various forms, including **proportional insurance** (or **coinsurance**) and **insurance with a deductible**. In proportional insurance, with retention rate $a \in (0,1)$, the policyholder retains aS and the insurer pays $(1-a)S$. Under a per-loss **deductible** D, the policyholder pays the first D dollars of each loss and the insurer pays the excess, possibly subject to a maximum claim amount. This resembles excess-of-loss coverage, whereas **stop-loss insurance** usually applies a retention to aggregate losses over a period rather than separately to every claim.

Example 5.5 A firm is interested in insuring losses from employee liability. Losses, measured in thousands of dollars, follow a compound Poisson model with frequency parameter $\lambda = 10$ and severity distribution

$$Y_j \sim \text{Pareto}(\alpha = 11, \theta = 1000).$$

The insurer uses the standard deviation premium principle,

$$\text{Premium} = \mathbb{E}[\text{Claim}] + \xi \ \text{sd}(\text{Claim}).$$

Full insurance is offered with loading $\xi = 0.5$. A 90% proportional insurance contract is also offered, under which the insurer covers 90% of each loss and the firm retains the remaining 10%; for this contract, $\xi = 0.4$. Finally, the insurer offers a deductible contract with deductible $D = 10$ and no upper limit on each loss, again with $\xi = 0.4$.

If the firm does not fully insure its losses, the regulator requires it to hold capital equal to the 95% CTE of the uninsured losses. This capital is treated as a loan from the shareholders, with an annual service charge equal to 10% of that CTE. Ignoring investment returns, the firm's expected annual outgo is therefore

$$\text{Premium} + \mathbb{E}[U] + 0.10 \ \text{CTE}_{0.95}(U),$$

where U denotes uninsured losses.

Compute the expected annual outgo under each of the following four options:

1. full insurance,
2. no insurance,
3. 90% proportional insurance,
4. deductible insurance with $D = 10$.

Use the FFT for the aggregate loss distribution where required.

Solution. We first compute the moments of the claim severity distribution. For a Pareto(α, θ) severity with $\alpha = 11$ and $\theta = 1000$,

$$\mathbb{E}[Y_j] = \frac{\theta}{\alpha - 1} = \frac{1000}{10} = 100,$$

and

$$\mathbb{E}[Y_j^2] = \frac{2\theta^2}{(\alpha-1)(\alpha-2)} = \frac{2 \cdot 1000^2}{10 \cdot 9} = 22222.2.$$

Since S is compound Poisson with parameter $\lambda = 10$,

$$\mathbb{E}[S] = \lambda \mathbb{E}[Y_j] = 10 \cdot 100 = 1000,$$

and

$$\mathrm{Var}(S) = \lambda \mathbb{E}[Y_j^2] = 10 \cdot 22222.2 = 222222,$$

so

$$\mathrm{sd}(S) = \sqrt{222222} \approx 471.4.$$

We now consider the four insurance choices.

(i) Full insurance. In this case the insurer covers the entire loss, so the premium is

$$\mathbb{E}[S] + 0.5\,\mathrm{sd}(S) = 1000 + 0.5(471.4) = 1235.7.$$

There is no uninsured loss, so $U = 0$ and the expected annual outgo is simply

$$1235.7.$$

(ii) No insurance. Here there is no premium, and the uninsured loss is $U = S$. Thus the expected uninsured loss is

$$\mathbb{E}[U] = \mathbb{E}[S] = 1000.$$

Using the FFT approach described above, the aggregate loss distribution gives

$$\mathrm{VaR}_{0.95}(S) = 1858 \qquad \text{and} \qquad \mathrm{CTE}_{0.95}(S) = 2167.$$

Therefore the expected annual outgo is

$$1000 + 0.10(2167) = 1216.7.$$

(iii) 90% proportional insurance. The uninsured loss is $U = 0.1S$ and the insured loss is $S^I = 0.9S$. The premium is therefore

$$\mathbb{E}[S^I] + 0.4\,\mathrm{sd}(S^I) = 0.9\mathbb{E}[S] + 0.4(0.9\,\mathrm{sd}(S)) = 900 + 0.36(471.4) = 1069.7.$$

The expected uninsured loss is

$$\mathbb{E}[U] = 0.1\mathbb{E}[S] = 100.$$

Since CTE is positively homogeneous,

$$\mathrm{CTE}_{0.95}(U) = 0.1\,\mathrm{CTE}_{0.95}(S) = 0.1(2167) = 216.7.$$

Hence the capital service charge is

$$0.10(216.7) = 21.7,$$

and the expected annual outgo is

$$1069.7 + 100 + 21.7 = 1191.4.$$

(iv) Deductible insurance with $D = 10$. Let Y_j^I and Y_j^U denote the insured and uninsured portions of a single loss. Then

$$Y_j^U = \min\{Y_j, D\} \qquad \text{and} \qquad Y_j^I = (Y_j - D)_+.$$

To compute the premium, we first analyze the insured severity. Let

$$Z = Y_j - D \mid Y_j > D.$$

Then, for $z \geq 0$,

$$\mathbb{P}(Z > z) = \mathbb{P}(Y_j - D > z \mid Y_j > D) = \frac{\mathbb{P}(Y_j > z + D)}{\mathbb{P}(Y_j > D)} = \left(\frac{\theta + D}{\theta + D + z}\right)^{\alpha},$$

which is the survival function of a Pareto distribution with parameters $\alpha = 11$ and $\theta + D = 1010$. Therefore

$$\mathbb{E}[Y_j^I] = \mathbb{E}[Y_j - D \mid Y_j > D]\mathbb{P}(Y_j > D) = \frac{\theta + D}{\alpha - 1}\left(\frac{\theta}{\theta + D}\right)^{\alpha} \approx 90.529,$$

and similarly

$$\mathbb{E}\big[(Y_j^I)^2\big] = \mathbb{E}\big[(Y_j - D)^2 \mid Y_j > D\big]\mathbb{P}(Y_j > D) = \frac{2(\theta + D)^2}{(\alpha - 1)(\alpha - 2)}\left(\frac{\theta}{\theta + D}\right)^{\alpha} \approx 20318.60.$$

Hence, for the aggregate insured loss S^I,

$$\mathbb{E}[S^I] = \lambda\mathbb{E}[Y_j^I] = 905.29 \qquad \text{and} \qquad \mathrm{sd}(S^I) = \sqrt{\lambda\mathbb{E}[(Y_j^I)^2]} = \sqrt{203186} \approx 450.76.$$

So the premium is

$$905.29 + 0.4(450.76) = 1085.59.$$

The expected uninsured loss is

$$\mathbb{E}[U] = \mathbb{E}[S] - \mathbb{E}[S^I] = 1000 - 905.29 = 94.71.$$

Using the FFT approach for the uninsured aggregate loss gives

$$\mathrm{CTE}_{0.95}(U) = 149.7.$$

Therefore the capital service charge is

$$0.10(149.7) = 15.0,$$

and the expected annual outgo is

$$1085.6 + 94.7 + 15.0 = 1195.3.$$

Thus the expected annual outgo under the four options is

$$\text{Full insurance } = 1235.7 \qquad \text{and} \qquad \text{No insurance } = 1216.7,$$

$$90\% \text{ proportional insurance } = 1191.4 \qquad \text{and} \qquad \text{Deductible insurance } = 1195.3.$$

The deductible and proportional contracts have very similar expected costs, but the deductible leaves the firm with less tail risk because its retained severity is capped at D, whereas proportional retention remains unbounded. $\square$

Lecture 6 — Tuesday, October 20, 2015

Extreme value theory studies rare, extreme outcomes. The focus here is on large (positive) losses. Two main approaches are used: modelling maxima and modelling exceedances over a high threshold.

4. Extremes, Dependence, and Financial Return Models

Maximum Losses

Let $X_1, \dots, X_n$ be independent and identically distributed with distribution function F, and let

$$M_n = \max\{X_1, \dots, X_n\}.$$

Then

$$\mathbb{P}(M_n \le m) = \mathbb{P}(X_1 \le m, \dots, X_n \le m) = [F(m)]^n.$$

As $n \to \infty$, this distribution either degenerates to 0 or 1, so some kind of normalization is required, which is analogous to the central limit theorem for sums or averages. In that setting we first center and rescale before taking a limit, for example

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right) \to \Phi(x).$$

Maxima also need centering and scaling. Without such normalization, the limit is typically trivial.

If the number of losses is random, say N, then, using the convention $M_0 = -\infty$, conditioning on N gives

$$\mathbb{P}(M_N \le x) = \sum_{n=0}^{\infty} \mathbb{P}(M_N \le x \mid N = n)\mathbb{P}(N = n) = \sum_{n=0}^{\infty} [F(x)]^n \mathbb{P}(N = n) = P_N(F(x)),$$

where P_N is the pgf of N.

Definition 6.1 Let M_n be the maximum of an independent and identically distributed sample from F. If there exist deterministic $c_n > 0$ and d_n such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - d_n}{c_n} \le x\right) = H(x),$$

then F is said to belong to the **maximum domain of attraction** of H, written

$$F \in \mathrm{MDA}(H).$$

GEV Distributions

There are three possible limit types for normalized maxima: Gumbel, Fréchet, and Weibull. In a location–scale parameterization, they can be written as follows.

Definition 6.2 The **Gumbel distribution** has distribution function

$$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \quad \sigma > 0.$$

The **Fréchet distribution** has distribution function

$$F(x) = \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right), \quad \alpha > 0, \quad x > \mu.$$

The **(negative) Weibull distribution** has distribution function

$$F(x) = \exp\left(-\left(-\frac{x-\mu}{\sigma}\right)^{\alpha}\right), \quad \alpha > 0, \quad x < \mu.$$

These three families are combined into the **generalized extreme value** (**GEV**) family

$$H_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), & \gamma \neq 0, \\ \exp(-e^{-x}), & \gamma = 0, \end{cases}$$

defined where $1 + \gamma x > 0$.

With location parameter μ and scale parameter $\sigma > 0$, the distribution becomes

$$H_{\gamma,\mu,\sigma}(x) = H_\gamma\left(\frac{x-\mu}{\sigma}\right).$$

If $\mu = 0$ and $\sigma = 1$, this reduces to the standard form H_γ. The shape parameter γ distinguishes the three cases: Fréchet corresponds to $\gamma > 0$, Gumbel to $\gamma = 0$, and Weibull to $\gamma < 0$. In the Fréchet case, the **tail index** is

$$\alpha = \frac{1}{\gamma}.$$

Theorem 6.3 (Fisher–Tippett–Gnedenko theorem) If $F \in \mathrm{MDA}(H)$ for some nondegenerate distribution H, then H is a GEV distribution; that is, $H = H_\gamma$ for some $\gamma \in \mathbb{R}$.

The Fisher–Tippett–Gnedenko theorem states that any nondegenerate limit law for normalized maxima must be a member of the GEV family. It is often convenient to express this convergence in terms of the **survival function** $S(x) = 1 - F(x)$ for x in the support of the limit law.

Theorem 6.4 $F \in \mathrm{MDA}(H_\gamma)$ if and only if there exist normalizing constants $c_n > 0$ and d_n such that

$$\lim_{n\to\infty} nS(c_n x + d_n) = -\log H_\gamma(x) = \begin{cases} (1+\gamma x)^{-1/\gamma}, & \gamma \neq 0, \\ e^{-x}, & \gamma = 0, \end{cases}.$$

This characterization is often easier to verify than the convergence of the full distribution of the normalized maxima.

Example 6.5 Determine the limiting distribution for a maximum of Pareto(α, λ) random variables, where $\alpha > 0$ and $\lambda > 0$, and find the normalizing constants c_n and d_n.

Solution. For the Pareto distribution,

$$S(x) = \mathbb{P}(X > x) = \left(\frac{\lambda}{\lambda + x}\right)^\alpha, \quad x > 0.$$

Hence

$$nS(c_n x + d_n) = n\left(\frac{\lambda}{\lambda + d_n + c_n x}\right)^\alpha = n\left(1 + \frac{d_n}{\lambda} + \frac{c_n}{\lambda}x\right)^{-\alpha}.$$

Rewriting this gives

$$nS(c_n x + d_n) = \left(n^{-1/\alpha} + \frac{n^{-1/\alpha} d_n}{\lambda} + \frac{n^{-1/\alpha} c_n}{\lambda} x\right)^{-\alpha}.$$

This has the Fréchet form $(k_1 + k_2 x)^{-\alpha}$, so the limit is in the Fréchet family with shape parameter

$$\gamma = \frac{1}{\alpha} > 0.$$

To match the standard Fréchet form

$$-\log H_\gamma(x) = (1 + \gamma x)^{-1/\gamma} = (1 + \gamma x)^{-\alpha},$$

we choose the constants so that, asymptotically,

$$n^{-1/\alpha} + \frac{n^{-1/\alpha} d_n}{\lambda} = 1 \qquad \text{and} \qquad \frac{n^{-1/\alpha} c_n}{\lambda} = \gamma = \frac{1}{\alpha}.$$

Therefore

$$d_n = (n^{1/\alpha} - 1)\lambda \qquad \text{and} \qquad c_n = \frac{\lambda n^{1/\alpha}}{\alpha}.$$

In summary, the Pareto distribution belongs to the maximum domain of attraction of the Fréchet distribution with shape parameter $\gamma = 1/\alpha$, and one suitable choice of normalizing constants is

$$c_n = \frac{\lambda n^{1/\alpha}}{\alpha} \qquad \text{and} \qquad d_n = (n^{1/\alpha} - 1)\lambda.$$

□

Let us discuss the main tail classes.

Fréchet limits arise from heavy-tailed distributions such as Pareto, inverse gamma, and Student's t. More generally, **Pareto tails**, which are of the form

$$cx^{-\alpha}$$

or, more generally,

$$L(x)x^{-\alpha},$$

where L is slowly varying at infinity, fall into the Fréchet domain. For a regularly varying tail with index $-\alpha$, moments of order below α are finite and moments of order above α are infinite; whether the moment of order exactly α exists depends on L. For example, a Pareto distribution with $\alpha = 3$ has finite mean and variance but infinite third moment, and it belongs to the Fréchet domain with $\gamma = 1/3$.

Definition 6.6 A function $L : (0, \infty) \to (0, \infty)$ is **slowly varying at infinity** if, for every $t > 0$,

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1.$$

A function $h : (0, \infty) \to (0, \infty)$ is **regularly varying at infinity** with index ρ if, for every $t > 0$,

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = t^{\rho}.$$

The Gumbel domain contains many unbounded distributions with tails lighter than a power law, including the normal, lognormal, exponential, gamma, and inverse Gaussian families. Exponential-type tails of the form $ke^{-x/\theta}$, or more generally $L(x)e^{-x/\theta}$ with L slowly varying at infinity, are standard examples. We can show that F is in the Gumbel MDA if and only if the survival function $S = 1 - F$ satisfies

$$\lim_{n \to \infty} nS(c_n x + d_n) = e^{-x}$$

for some normalizing constants $c_n > 0$ and d_n.

The Weibull maximum domain instead corresponds to distributions with finite upper endpoints. For example, the Unif$[0, 1]$ distribution belongs to the Weibull domain with $\gamma = -1$. More generally, if F has a finite upper endpoint x_F, then F is in the Weibull MDA if and only if the survival function S satisfies

$$\lim_{n \to \infty} nS(x_F - c_n x) = x^{-1/\gamma}, \quad x > 0,$$

for some normalizing constants $c_n > 0$. Because the tail is bounded, the Weibull domain is not usually relevant for unbounded loss severities, but it can be relevant when the variable has a natural upper endpoint.

Although many common loss models belong to one of these three maximum domains of attraction, not every distribution does. Classical counterexamples can be built from extremely slowly varying tails. For example,

$$F(x) = 1 - \frac{1}{\log x}, \quad x > e$$

and

$$F(x) = 1 - \frac{1}{\log \log x}, \quad x > e^e$$

have extremely heavy tails that do not belong to any of the three maximum domains of attraction. However, such examples are not common in practice.

Fitting GEV Distributions by Block Maxima

The **block maxima method** divides the data into blocks of size n, records the maximum in each block, and then treats those maxima as a sample from a three-parameter GEV distribution. If $m_1, \dots, m_k$ are the block maxima, then the GEV density is

$$h(x) = \frac{1}{\sigma}\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1-1/\gamma}\exp\left[-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right],$$

for $1+\gamma(x-\mu)/\sigma > 0$, and the log likelihood is

$$\ell(\gamma,\mu,\sigma) = \sum_{j=1}^{k} \log h(m_j).$$

For $\gamma \neq 0$, this can be written more explicitly as

$$\ell(\gamma,\mu,\sigma) = -k\log(\sigma) - \left(1+\frac{1}{\gamma}\right)\sum_{j=1}^{k}\log\left(1+\gamma\frac{m_j-\mu}{\sigma}\right) - \sum_{j=1}^{k}\left(1+\gamma\frac{m_j-\mu}{\sigma}\right)^{-1/\gamma},$$

with the Gumbel case obtained by continuity as $\gamma \to 0$. Maximum likelihood estimation is then used for γ, μ, and σ. We can test $\gamma = 0$ against the two-sided alternative $\gamma \neq 0$ using a likelihood ratio test; under the usual regularity conditions,

$$-2\big(\ell(0,\widehat{\mu}_0,\widehat{\sigma}_0) - \ell(\widehat{\gamma},\widehat{\mu},\widehat{\sigma})\big) \xrightarrow{d} \chi^2_1,$$

where $(\widehat{\gamma},\widehat{\mu},\widehat{\sigma})$ and $(0,\widehat{\mu}_0,\widehat{\sigma}_0)$ are the unconstrained and constrained maximum likelihood estimates, respectively. If the question is specifically whether $\gamma > 0$, the sign of $\widehat{\gamma}$ and an appropriate one-sided procedure also matter.

There is a tradeoff in block choice. Longer blocks give us more confidence that the maxima lie in the tail of the distribution, but they also reduce the number of block maxima available for estimation. Shorter blocks give more data but less confidence that the limiting approximation is accurate. The standard error of γ is often especially large, which matters because γ determines the tail type.

Threshold Exceedances and the GPD

The block maxima method discards most of the data, so an alternative is to study exceedances above a threshold d. Define the exceedance random variable

$$Y_d = X - d \mid X > d.$$

Its distribution function is

$$F_d(x) = \mathbb{P}(X - d \le x \mid X > d),$$

and its survival function is obtained from the parent tail. If $S_X(x) = 1 - F_X(x)$ is the survival function of X, then the survival function of the exceedance is

$$S_d(x) = \mathbb{P}(X - d > x \mid X > d) = \frac{S_X(x+d)}{S_X(d)}.$$

The **mean excess loss** is

$$e(d) = \mathbb{E}\left[X - d \mid X > d\right].$$

Definition 6.7 For $\beta > 0$, the **generalized Pareto distribution** is

$$G_{\gamma,\beta}(x) = \begin{cases} 1 - (1 + \gamma x/\beta)^{-1/\gamma}, & \gamma \ne 0, \\ 1 - e^{-x/\beta}, & \gamma = 0, \end{cases}$$

with $x \ge 0$ when $\gamma \ge 0$ and $0 \le x \le -\beta/\gamma$ when $\gamma < 0$.

When $\gamma > 0$, the GPD is Pareto; when $\gamma = 0$, it is exponential; and when $\gamma < 0$, it is a generalized beta distribution, which is bounded above.

If $Y \sim \text{GPD}(\gamma, \beta)$ in the sense of Definition 6.7, $\beta + \gamma d > 0$, and $Z = Y - d \mid Y > d$, then

$$Z \sim \text{GPD}(\gamma, \beta + \gamma d).$$

Therefore the mean excess function is linear:

$$e(d) = \frac{\beta + \gamma d}{1 - \gamma}, \quad \gamma < 1.$$

Theorem 6.8 (Pickands–Balkema–de Haan theorem) Let

$$x_F := \sup\{x : F(x) < 1\}.$$

Then $F \in \mathrm{MDA}(H_\gamma)$ if and only if there is a positive function $\beta(d)$ such that

$$\lim_{d \uparrow x_F} \sup_{\substack{x \geq 0 \\ x < x_F - d}} \left| F_d(x) - G_{\gamma,\beta(d)}(x) \right| = 0,$$

where the comparison is restricted to the common support of the two distributions.

One direction of the Pickands–Balkema–de Haan theorem states that if $F_X \in \mathrm{MDA}(H_\gamma)$ in the sense of Definition 6.1, then for high thresholds d the exceedance distribution is approximately generalized Pareto with the same shape parameter γ; see Definition 6.7. This makes threshold modelling extremely useful even when the full underlying distribution is unknown.

We now derive risk measures under a GPD tail approximation. Suppose the tail beyond threshold d is approximated by a GPD with parameters γ and β. Then, for $x > d$,

$$\mathbb{P}(X > x) = \mathbb{P}(X > d)\mathbb{P}(X > x \mid X > d) = S_X(d)\left(1 + \frac{\gamma(x-d)}{\beta}\right)^{-1/\gamma}.$$

Setting $x = Q_\alpha$ and solving for the quantile gives

$$1 - \alpha = S_X(d)\left(1 + \frac{\gamma(Q_\alpha - d)}{\beta}\right)^{-1/\gamma},$$

so

$$Q_\alpha = d + \frac{\beta}{\gamma}\left[\left(\frac{1-\alpha}{S_X(d)}\right)^{-\gamma} - 1\right].$$

In practice $S_X(d)$ is estimated empirically. If $d = x_{(n-j)}$ is the $(n-j)$th order statistic from a sample of size n, then

$$\widehat{S}_X(d) = \frac{j}{n}.$$

For the CTE, observe that

$$\mathrm{CTE}_\alpha = \mathbb{E}\left[X \mid X > Q_\alpha\right] = Q_\alpha + e(Q_\alpha).$$

Assuming that in the tail beyond $d < Q_\alpha$ the GPD approximation holds, we have that $X - Q_\alpha \mid$

$X > Q_\alpha$ is approximately GPD$(\gamma, \beta + \gamma(Q_\alpha - d))$. If $\gamma < 1$, then the mean of this distribution exists and is given by

$$e(Q_\alpha) = \frac{\beta + \gamma(Q_\alpha - d)}{1 - \gamma}.$$

Therefore

$$\mathrm{CTE}_\alpha = Q_\alpha + \frac{\beta + \gamma(Q_\alpha - d)}{1 - \gamma} = \frac{Q_\alpha + \beta - \gamma d}{1 - \gamma}.$$

The parameters γ and β may be estimated from exceedances above a threshold D by maximum likelihood. Confidence intervals follow from the asymptotic covariance matrix. We may test $\gamma = 0$ against $\gamma \neq 0$ by a likelihood ratio test under the usual regularity conditions; a specifically one-sided alternative $\gamma > 0$ requires an appropriate one-sided procedure.

A standard diagnostic is the **empirical mean excess plot**. If the ordered data are

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)},$$

then the empirical mean excess function at threshold $x_{(j)}$ is

$$\widehat{e}(x_{(j)}) = \frac{1}{n-j} \sum_{i=j+1}^{n} \big(x_{(i)} - x_{(j)}\big).$$

We choose a threshold D above which this plot is approximately linear, balancing bias against variance. If the mean excess plot slopes upward, a Pareto-type tail is indicated. If it is roughly flat, an exponential tail is more plausible.

Once a threshold D has been selected, we fit the GPD to the exceedances

$$x_i - D, \quad x_i > D,$$

and then study the stability of the fitted parameters and their confidence intervals as the threshold varies. In a good threshold region, neighboring thresholds should lead to similar fitted tail distributions and similar estimates of extreme quantiles.

Estimation of tail quantities is often difficult because the most extreme part of the sample is volatile, the number of exceedances is small, and threshold choice can materially affect the fitted model. The **Hill estimator** provides another way to estimate the tail index and is taken up next.

Lecture 7 — Tuesday, October 27, 2015

The Hill Method

For distributions in the maximum domain of attraction of the Fréchet law, as in Definition 6.1, the survival function has a regularly varying form

$$S(x) = L(x)x^{-1/\gamma} = L(x)x^{-\alpha},$$

where L is slowly varying at infinity and $\alpha = 1/\gamma$; see Definition 6.6. This representation suggests using the tail behaviour of the logarithms of the losses to estimate the tail index.

Assume $X > 0$. The mean excess function of $\log X$ above threshold $\log u$ is

$$e_{\log X}(\log u) = \mathbb{E}\left[\log X - \log u \mid X > u\right].$$

Using the survival function, this can be written as

$$e_{\log X}(\log u) = \frac{1}{S_X(u)} \int_u^\infty (\log x - \log u)\, \mathrm{d}F_X(x) = \frac{1}{S_X(u)} \int_u^\infty \frac{S_X(x)}{x}\, \mathrm{d}x.$$

If $S_X(x) \approx L(x)x^{-\alpha}$ and u is large, then $L(x)$ is approximately constant over the relevant range, so

$$e_{\log X}(\log u) \approx \frac{L(u)}{S_X(u)} \int_u^\infty x^{-(\alpha+1)}\, \mathrm{d}x \approx \frac{L(u)}{S_X(u)} \cdot \frac{u^{-\alpha}}{\alpha} \approx \frac{1}{\alpha} = \gamma.$$

Thus the mean excess function of the logarithms of the losses approaches the extreme value parameter γ.

Definition 7.1 Let

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

be the ordered sample. Choosing a threshold at $d = x_{(n-j)}$, the **Hill estimator** of α is

$$\widehat{\alpha}_j^H = \left[\frac{1}{j}\sum_{k=1}^{j} \log\left(\frac{x_{(n-k+1)}}{x_{(n-j)}}\right)\right]^{-1}, \quad j = 1, \ldots, n-1.$$

The Hill estimator is a slight variant of the mean excess function of the logarithms of the losses, and it is consistent for α under suitable conditions. The choice of j is crucial. If j is too small, the estimate is based on very few data points and is highly variable. If j is too large, the estimate is biased because it includes data from the central part of the distribution rather than just the tail. Therefore, we typically examine a **Hill plot** over a range of upper order statistics and look for a stable region.

Once the threshold is fixed at $d = x_{(n-j)}$, the exceedance probability is estimated by

$$\mathbb{P}(X > d) \approx \frac{j}{n}.$$

Using the Pareto-type tail form, the Hill estimator of the survival function for $x > d$ becomes

$$\widehat{S}^H(x) = \frac{j}{n} \left(\frac{x}{x_{(n-j)}} \right)^{-\widehat{\alpha}_j^H}.$$

This provides a direct way to estimate extreme tail probabilities beyond the observed threshold.

Lecture 8 — Tuesday, November 03, 2015

5. Copulas and Dependence

Importance of Dependence

Portfolio loss distributions depend not only on the marginal behaviour of the individual risks, but also on their dependence structure. Expected loss, standard deviation, tail probabilities, VaR, and CTE can all change materially when dependence changes. Correlation alone does not capture this. It measures only linear dependence, so zero correlation does not imply independence, although independence does imply zero correlation. In the Vasicek illustration in Figure 1, the unconditional default probability is fixed at $p = 5\%$, but increasing the asset correlation ρ moves more mass toward small portfolio losses while leaving a heavier right tail.

If X and Y have finite, positive variances and their correlation is ± 1, then $X = aY + b$ almost surely, with the sign of a matching the sign of the correlation. Away from this extreme case, however, correlation can miss important dependence features. For example, $Y = X^2$ can exhibit strong deterministic nonlinear dependence even when the linear correlation is zero.

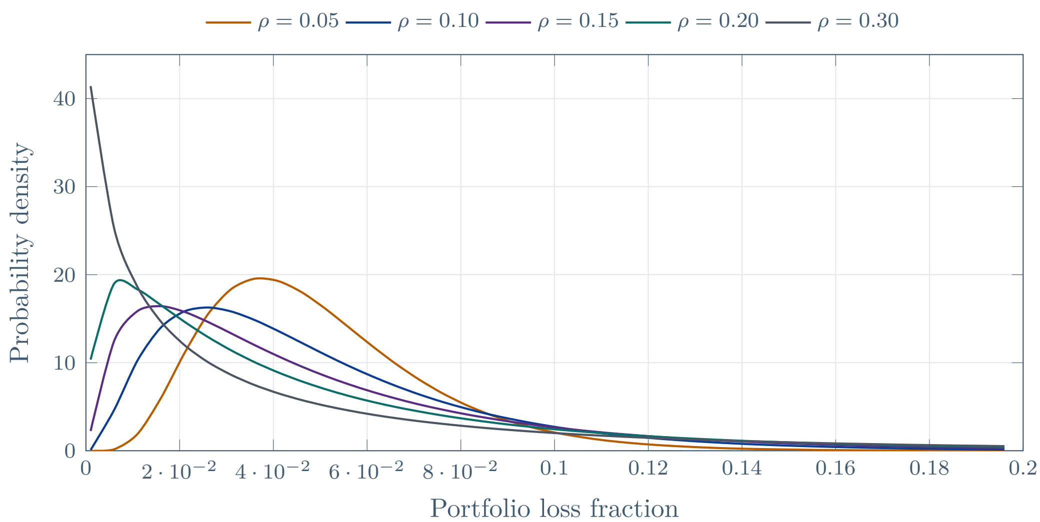


FIGURE 1

Copulas and Sklar's Theorem

Definition 8.1 A **copula** is a joint distribution function

$$C : [0,1]^d \to [0,1]$$

with uniform $(0,1)$ marginals.

By Sklar's theorem in Theorem 8.4, if $\boldsymbol{X} = (X_1, \ldots, X_d)$ has marginal distribution functions $F_1, \ldots, F_d$, then a copula links the marginals to the joint distribution by

$$F_{\boldsymbol{X}}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

This is useful because the marginals and the dependence can be modelled separately.

Example 8.2 Suppose

$$L_A \sim \mathcal{N}(1000, 300^2) \qquad \text{and} \qquad L_B \sim \Gamma(4, 250),$$

and consider the limits $x_A = 1500$ and $x_B = 1600$. Then

$$\mathbb{P}(L_A \leq 1500) = \Phi\left(\frac{1500 - 1000}{300}\right) = \Phi(1.6667) \approx 0.95221.$$

For the gamma risk,

$$\mathbb{P}(L_B \leq 1600) = F_{\Gamma(4,250)}(1600) \approx 0.88108.$$

Under independence,

$$\mathbb{P}(L_A \leq 1500,\ L_B \leq 1600) = 0.95221(0.88108) \approx 0.8389.$$

If instead

$$C(u, v) = (u^{-2} + v^{-2} - 1)^{-1/2},$$

then

$$\mathbb{P}(L_A \leq 1500,\ L_B \leq 1600) = C(0.95221, 0.88108) \approx 0.84787.$$

For the stronger dependence model,

$$C(u, v) = (u^{-50} + v^{-50} - 1)^{-1/50},$$

the corresponding joint probability becomes

$$C(0.95221, 0.88108) \approx 0.88075.$$

The largest possible value of

$$\mathbb{P}(L_A \leq x_A,\ L_B \leq x_B)$$

is

$$\min\{\mathbb{P}(L_A \leq x_A), \mathbb{P}(L_B \leq x_B)\} = \min\{0.95221, 0.88108\} = 0.88108,$$

which occurs under **comonotonicity**. The smallest possible value is

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1 = 0.95221 + 0.88108 - 1 = 0.83329,$$

which is attained in the bivariate case by **countermonotonicity**.

In the notation of the example, the upper bound is attained when the two losses are increasing transformations of the same underlying uniform variable. Equivalently,

$$L_A = F_A^{-1}(F_B(L_B)).$$

The lower Fréchet bound in the bivariate case is attained by making one loss an increasing transformation of the opposite tail of the other:

$$L_A = F_A^{-1}\big(1 - F_B(L_B)\big).$$

Proposition 8.3 To be a d-dimensional copula, a function must be grounded, have uniform marginals, and be ***d*-increasing**. Groundedness means

$$C(u_1, \ldots, u_d) = 0$$

whenever at least one coordinate is zero. The uniform marginal condition is

$$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$$

for each coordinate i. The d-increasing condition says that all rectangular probabilities are nonnegative. More explicitly, for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ satisfying $a_j < b_j$ for each j,

$$\mathbb{P}(a_1 < U_1 < b_1, \ldots, a_d < U_d < b_d) \geq 0,$$

which requires

$$\sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1+\cdots+j_d} C(u_{1j_1}, \ldots, u_{dj_d}) \geq 0, \quad u_{i1} = a_i \qquad \text{and} \qquad u_{i2} = b_i.$$

In the smooth case, a sufficient condition is

$$\frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \ldots, u_d) \geq 0$$

for all $0 < u_j < 1$.

Theorem 8.4 (Sklar's theorem) For any joint distribution function F with marginals $F_1, \ldots, F_d$, there exists a copula C such that

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

Conversely, any choice of marginals together with any copula produces a valid joint distribution. If the marginals are continuous, then the copula is unique.

The basic idea behind Sklar's theorem is the probability integral transform. If the marginals are continuous, then

$$F_j(X_j) \sim \text{Unif}(0,1),$$

and hence

$$\begin{aligned} F_{\boldsymbol{X}}(x_1,\dots,x_d) &= \mathbb{P}(X_1 \le x_1,\dots,X_d \le x_d) \\ &= \mathbb{P}(F_1(X_1) \le F_1(x_1),\dots,F_d(X_d) \le F_d(x_d)) \\ &= C(F_1(x_1),\dots,F_d(x_d)). \end{aligned}$$

In this continuous case the copula may be written as

$$C(u_1,\dots,u_d) = F_{\boldsymbol{X}}(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d)).$$

For simplicity, we focus on bivariate copulas going forward, but most of the ideas extend to higher dimensions.

Proposition 8.5 (Invariance under increasing transformations) Suppose X and Y have continuous distribution functions, so their copula is unique. If C is the copula of (X,Y), then it is also the copula of $(g(X),h(Y))$ whenever g and h are strictly increasing deterministic functions.

Proof. If C is the copula for (X,Y), then

$$C(u,v) = \mathbb{P}(X \le F_X^{-1}(u), Y \le F_Y^{-1}(v)).$$

If g and h are strictly increasing, then

$$\begin{aligned} C(u,v) &= \mathbb{P}(g(X) \le g(F_X^{-1}(u)), h(Y) \le h(F_Y^{-1}(v))) \\ &= \mathbb{P}(g(X) \le F_{g(X)}^{-1}(u), h(Y) \le F_{h(Y)}^{-1}(v)), \end{aligned}$$

which shows that the copula remains the same under strictly increasing transformations of the marginals. □

Explicit and Implicit Bivariate Copulas

Definition 8.6 Three fundamental bivariate copulas are the **independence copula**

$$C^{\perp}(u,v) = uv,$$

the **comonotonicity copula**

$$C^{+}(u,v) = \min(u,v),$$

and the **countermonotonicity copula**

$$C^{-}(u,v) = \max(u+v-1,0).$$

Important parametric copulas include the **Clayton copula**

$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \quad \theta > 0,$$

and the **Gumbel copula**

$$C_\theta(u,v) = \exp\left(-\left((-\log u)^\theta + (-\log v)^\theta\right)^{1/\theta}\right), \quad \theta \geq 1.$$

We can show that the Clayton copula tends to the independence copula as $\theta \downarrow 0$ and to the comonotonicity copula as $\theta \to \infty$. The Gumbel copula tends to the independence copula as $\theta \downarrow 1$ and to the comonotonicity copula as $\theta \to \infty$. The level curves in Figure 2 compare these families: Clayton concentrates lower-tail dependence, while Gumbel concentrates upper-tail dependence.

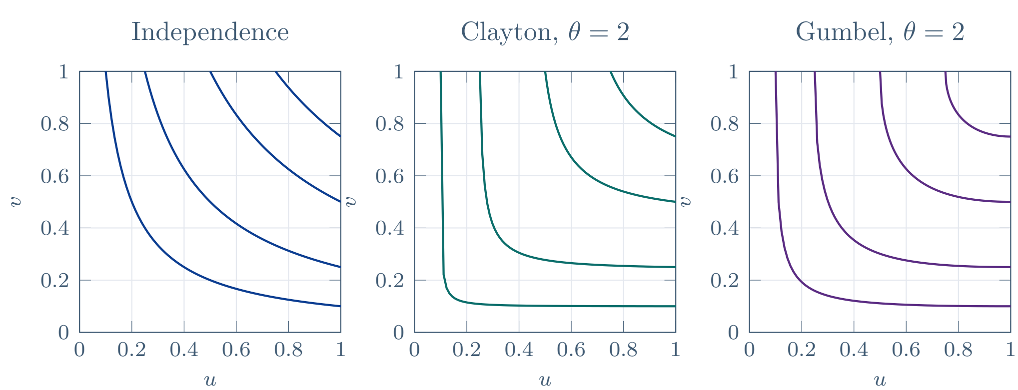


FIGURE 2

Definition 8.7 Implicit copulas arise from multivariate distributions. The **Gaussian copula** is

$$C(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), \quad -1 \le \rho \le 1,$$

and the **Student *t* copula** is

$$C(u,v) = F_{\nu,\rho}\big(F_\nu^{-1}(u), F_\nu^{-1}(v)\big).$$

Here Φ_ρ is the bivariate normal distribution function with correlation ρ, $F_{\nu,\rho}$ is the bivariate Student t distribution function with degrees of freedom $\nu > 0$ and dependence parameter $-1 \le \rho \le 1$, and F_ν is the ordinary univariate Student t distribution function. At $\rho = \pm 1$ the copula is degenerate. When $\nu \le 2$, ρ is not a Pearson correlation because the component variances are infinite.

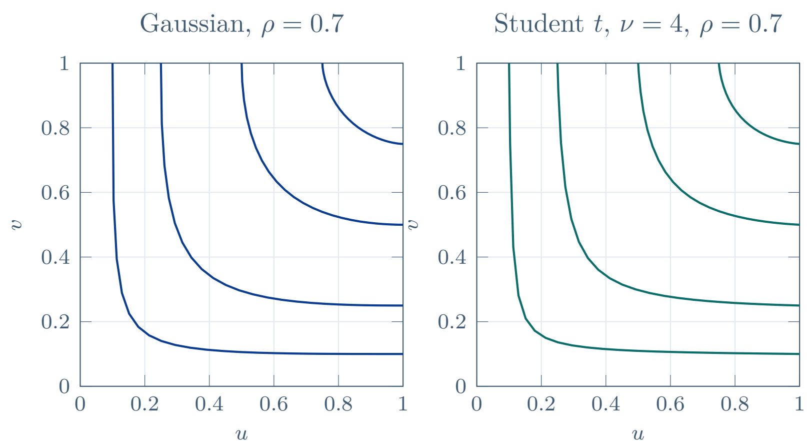


FIGURE 3

Figure 3 compares Gaussian and Student t copulas with the same correlation parameter. The Student t level curves bend more sharply into the corners, reflecting symmetric tail clustering. More generally, copulas with similar positive association can have very different tail behaviour: the Gaussian copula has no tail clustering in the limit, the Student t copula has symmetric tail clustering, the Gumbel copula emphasizes the upper tail, and the Clayton copula emphasizes the lower tail.

Definition 8.8 If C is the copula of (U,V), then the associated **joint survival function** is

$$\overline{C}(u,v) = \mathbb{P}(U > u, V > v) = 1 - u - v + C(u,v).$$

For losses X and Y with $F_X(x) = u$ and $F_Y(y) = v$, this is also

$$\mathbb{P}(X > x, Y > y) = \overline{C}(u, v).$$

When F_X and F_Y are continuous, the **survival copula**, namely the copula of $(1-F_X(X), 1-F_Y(Y))$, satisfies

$$\widehat{C}(u, v) = u + v - 1 + C(1-u, 1-v).$$

Proposition 8.9 (Fréchet–Hoeffding bounds) For any d-dimensional copula,

$$\max\{u_1 + \cdots + u_d - (d-1), 0\} \le C(u_1, \ldots, u_d) \le \min\{u_1, \ldots, u_d\}.$$

The upper bound is the comonotonicity copula. In dimensions larger than two, the lower bound is not a copula.

Definition 8.10 **Radial symmetry** means that U and $1-U$ have the same dependence structure, or equivalently that

$$C(u, v) = \widehat{C}(u, v) = u + v - 1 + C(1-u, 1-v).$$

Exchangeability means that the variables can be swapped without changing the copula:

$$C(u, v) = C(v, u).$$

Measures of Dependence

Pearson correlation measures linear association. It can be misleading when dependence is nonlinear, for example when $Y = X^2$, and some marginal distributions make certain Pearson correlations unattainable. Rank-based measures are often preferred because they are functions of the copula rather than of the marginal scales. Assume in the population formulas below that the marginals are continuous, and set

$$U = F_X(X) \qquad \text{and} \qquad V = F_Y(Y),$$

then the following two rank-based dependence measures are standard.

Definition 8.11 **Spearman's rho** is

$$\rho_S(X,Y) = \rho(U,V) = \frac{\mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V]}{\sqrt{\mathrm{Var}(U)\,\mathrm{Var}(V)}} = \frac{\mathbb{E}[UV] - 1/4}{1/12} = 12\mathbb{E}[UV] - 3.$$

Comonotonic risks have $\rho_S = 1$, while countermonotonic risks have $\rho_S = -1$.

Kendall's tau is

$$\tau_K = \mathbb{E}\left[\mathrm{sign}((X - X^*)(Y - Y^*))\right] = 4\mathbb{E}\left[C(U,V)\right] - 1,$$

where (X^*, Y^*) is an independent copy of (X,Y). It measures concordance: the product $(X - X^*)(Y - Y^*)$ tends to be positive when larger values of X are associated with larger values of Y, and tends to be negative when larger values of X are associated with smaller values of Y. Here

$$\mathrm{sign}(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0. \end{cases}$$

Spearman's rho and Kendall's tau both measure **concordance**, and are invariant under strictly increasing transformations of the marginals.

For data without ties, Spearman's rho can be estimated from ranks. Let $r_{x,i}$ and $r_{y,i}$ be the ranks of the ith observations of X and Y, using a consistent rank convention. With

$$\overline{r}_x = \overline{r}_y = \frac{n+1}{2} \qquad \text{and} \qquad s^2_{r_x} = s^2_{r_y} = \frac{n(n+1)}{12},$$

we have

$$r_S = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{r_{x,i} - \overline{r}_x}{s_{r_x}}\right)\left(\frac{r_{y,i} - \overline{r}_y}{s_{r_y}}\right) = 12\frac{\sum_{i=1}^n r_{x,i}r_{y,i}}{n(n^2-1)} - 3\frac{n+1}{n-1}.$$

Equivalently, if $d_i = r_{x,i} - r_{y,i}$, then

$$r_S = 1 - \frac{6\sum_{i=1}^n d_i^2}{n(n^2-1)}.$$

If there are ties, the tied observations are assigned their average rank. For example, if three observations would occupy ranks j, $j+1$, and $j+2$, each is assigned rank $j+1$.

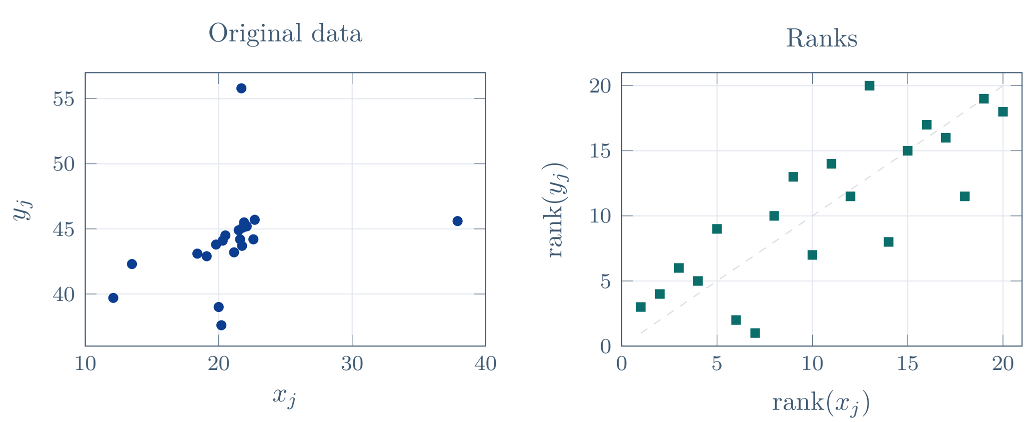
Kendall's tau is estimated by

$$t_K = \frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \operatorname{sign}\big((x_i - x_j)(y_i - y_j)\big),$$

assuming no ties. If there are ties, the formula can be modified.

Example 8.12 Consider the following data:

j	x_j	y_j	rank(x_j)	rank(y_j)
1	21.60	44.20	12	11.5
2	22.60	44.20	18	11.5
3	21.80	45.10	15	15
4	21.70	55.80	13	20
5	22.10	45.20	17	16
6	21.75	43.70	14	8
7	21.50	44.90	11	14
8	21.90	45.50	16	17
9	22.70	45.70	19	19
10	37.90	45.60	20	18
11	21.15	43.20	10	7
12	20.50	44.50	9	13
13	20.30	44.10	8	10
14	19.80	43.80	5	9
15	19.10	42.90	4	5
16	18.40	43.10	3	6
17	13.50	42.30	2	4
18	12.10	39.70	1	3
19	20.20	37.60	7	1
20	20.00	39.00	6	2



For these data,

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right) \approx 0.33, \qquad r_S \approx 0.81, \qquad \text{and} \qquad t_K \approx 0.62.$$

This shows that rank dependence can be much stronger than linear dependence.

Tail Dependence

Tail dependence is especially important in risk management. The practical questions are whether bad outcomes tend to occur independently, and whether dependence becomes stronger in extreme cases. For continuous marginals, tail dependence is defined using quantiles so that it depends on the copula rather than on the marginal distributions:

$$C(u,v) = \mathbb{P}(X \le Q_u(X),\, Y \le Q_v(Y)).$$

Definition 8.13 For the bivariate case, define

$$\lambda_U(q) = \mathbb{P}(X > Q_q(X) \mid Y > Q_q(Y)) \qquad \text{and} \qquad \lambda_L(q) = \mathbb{P}(X \le Q_q(X) \mid Y \le Q_q(Y)).$$

If the limits exist, the **upper** and **lower tail dependence coefficients** are

$$\lambda_U = \lim_{q\uparrow 1}\lambda_U(q) \qquad \text{and} \qquad \lambda_L = \lim_{q\downarrow 0}\lambda_L(q).$$

Proposition 8.14 In terms of the copula,

$$\lambda_U = \lim_{q\uparrow 1}\frac{1-2q+C(q,q)}{1-q} \qquad \text{and} \qquad \lambda_L = \lim_{q\downarrow 0}\frac{C(q,q)}{q}.$$

The upper-tail expression may also be written using the survival probability:

$$\lambda_U = \lim_{q\uparrow 1}\frac{\overline{C}(q,q)}{1-q} = \lim_{q\uparrow 1}\frac{1-2q+C(q,q)}{1-q}.$$

By L'Hopital's rule, when the relevant derivative limit exists,

$$\lambda_U = 2 - \lim_{q\uparrow 1}\frac{\mathrm{d}}{\mathrm{d}q}C(q,q) \qquad \text{and} \qquad \lambda_L = \lim_{q\downarrow 0}\frac{\mathrm{d}}{\mathrm{d}q}C(q,q).$$

The coefficients satisfy $0 \le \lambda_U, \lambda_L \le 1$. If $\lambda_U > 0$, then X and Y have upper-tail dependence; if $\lambda_L > 0$, then they have lower-tail dependence. A copula may have any combination of upper- and lower-tail dependence. Every nondegenerate Gaussian copula with $|\rho| < 1$ has zero tail dependence, while every nondegenerate Student t copula with finite ν and $-1 < \rho < 1$ has positive symmetric tail dependence.

Lecture 9 — Tuesday, November 10, 2015

Archimedean Copulas

Definition 9.1 An **Archimedean copula** is built from a **generator function**

$$\phi : [0,1] \to [0,\infty], \qquad \phi(1) = 0, \qquad \text{and} \qquad \phi(0) = \infty,$$

that is continuous and strictly decreasing. In the bivariate case, convexity of ϕ is also required (a more onerous condition called ***d*-monotonicity** is required for higher dimensions). The associated copula is

$$C(u_1, \ldots, u_n) = \phi^{-1}\big(\phi(u_1) + \cdots + \phi(u_n)\big).$$

Proposition 9.2 For Archimedean copulas with a differentiable generator for which the integral below exists, Kendall's tau can be expressed directly in terms of the generator:

$$\tau_K = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi'(u)} \, du.$$

Write $\psi = \phi^{-1}$. The upper and lower tail dependence coefficients can likewise be written as

$$\lambda_U = 2 - \lim_{t \downarrow 0} \frac{1 - \psi(2t)}{1 - \psi(t)} \qquad \text{and} \qquad \lambda_L = \lim_{t \to \infty} \frac{\psi(2t)}{\psi(t)},$$

whenever the limits exist. If the corresponding derivative limits exist and l'Hôpital's rule applies, these become

$$\lambda_U = 2 - 2 \lim_{t \downarrow 0} \frac{\psi'(2t)}{\psi'(t)} \qquad \text{and} \qquad \lambda_L = 2 \lim_{t \to \infty} \frac{\psi'(2t)}{\psi'(t)}.$$

Example 9.3 For the Clayton copula with parameter $\theta > 0$, the generator is

$$\phi(u) = \frac{u^{-\theta} - 1}{\theta}.$$

Compute ϕ^{-1}, τ_K, λ_U, and λ_L.

Solution. We have

$$\phi^{-1}(t) = (1 + \theta t)^{-1/\theta}.$$

Then

$$\tau_K = 1 + 4\int_0^1 \frac{\phi(u)}{\phi'(u)}\,du = 1 + 4\int_0^1 \frac{(u^{-\theta} - 1)/\theta}{-u^{-\theta-1}}\,du = 1 - \frac{4}{\theta}\int_0^1 (u - u^{\theta+1})\,du.$$

Now

$$\int_0^1 (u - u^{\theta+1})\,du = \int_0^1 u\,du - \int_0^1 u^{\theta+1}\,du = \frac{1}{2} - \frac{1}{\theta+2} = \frac{\theta}{2(\theta+2)}.$$

Therefore

$$\tau_K = 1 - \frac{4}{\theta}\cdot\frac{\theta}{2(\theta+2)} = 1 - \frac{2}{\theta+2} = \frac{\theta}{\theta+2}.$$

Next,

$$\lambda_U = 2 - 2\lim_{t \downarrow 0} \frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)}$$

and since

$$(\phi^{-1})'(t) = -(1 + \theta t)^{-1/\theta - 1},$$

we obtain

$$\frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)} = \frac{-(1 + 2\theta t)^{-1/\theta-1}}{-(1 + \theta t)^{-1/\theta-1}} = \left(\frac{1 + \theta t}{1 + 2\theta t}\right)^{1/\theta+1}.$$

As $t \downarrow 0$, this ratio tends to 1, so

$$\lambda_U = 2 - 2 \cdot 1 = 0,$$

and

$$\lambda_L = 2\lim_{t \to \infty} \frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)} = 2\lim_{t \to \infty} \left(\frac{1 + \theta t}{1 + 2\theta t}\right)^{1/\theta+1}.$$

Dividing numerator and denominator inside the ratio by t gives

$$\lambda_L = 2\left(\frac{\theta}{2\theta}\right)^{1/\theta+1} = 2\left(\frac{1}{2}\right)^{1/\theta+1} = 2^{-1/\theta}.$$

□

Example 9.4 For the Gumbel copula with parameter $\theta > 1$, the generator is

$$\phi(u) = (-\log u)^{\theta}.$$

Compute ϕ^{-1}, τ_K, λ_U, and λ_L.

Solution. We have

$$\phi^{-1}(t) = \exp\left(-t^{1/\theta}\right).$$

Its derivative is

$$(\phi^{-1})'(t) = -\frac{1}{\theta}t^{1/\theta-1}e^{-t^{1/\theta}},$$

so

$$\frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)} = 2^{1/\theta-1}\exp\left[-\left(2^{1/\theta}-1\right)t^{1/\theta}\right].$$

Also,

$$\tau_K = 1 + 4\int_0^1 \frac{\phi(u)}{\phi'(u)}\,du = 1 + 4\int_0^1 \frac{(-\log u)^{\theta}}{-\theta(-\log u)^{\theta-1}/u}\,du = 1 - \frac{4}{\theta}\int_0^1 u(-\log u)\,du = 1 - \frac{1}{\theta}.$$

Next,

$$\lambda_U = 2 - 2\lim_{t\downarrow 0}\frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)} = 2 - 2^{1/\theta},$$

and

$$\lambda_L = 2\lim_{t\to\infty}\frac{(\phi^{-1})'(2t)}{(\phi^{-1})'(t)} = 0.$$

□

Fitting Copulas

One approach is full joint maximum likelihood, which maximizes simultaneously over the marginal and copula parameters. The inference-functions-for-margins approach estimates the marginals first and then estimates the copula parameters from the fitted marginal transforms. A semiparametric pseudo-likelihood instead uses rank-based pseudo-observations.

If C is an absolutely continuous bivariate copula, its density is

$$c(u,v) := \frac{\partial^2}{\partial u\,\partial v}C(u,v).$$

Proposition 9.5 If X and Y have marginal densities f_X and f_Y and their copula is absolutely continuous with density c, then the joint density of (X, Y) can be factorized, almost everywhere, as

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y)).$$

Proof. By Sklar's theorem (Theorem 8.4), the joint distribution function of (X, Y) can be written as

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$

Because X and Y are absolutely continuous, the marginal distribution functions are differentiable almost everywhere with derivatives f_X and f_Y. Since the copula is absolutely continuous, its mixed partial derivative exists almost everywhere and equals the copula density c. Differentiating first with respect to x gives

$$\frac{\partial}{\partial x} F_{X,Y}(x, y) = \frac{\partial C}{\partial u}(F_X(x), F_Y(y))\, f_X(x).$$

Now differentiate with respect to y. Since $f_X(x)$ does not depend on y, we obtain

$$f_{X,Y}(x, y) = \frac{\partial^2 C}{\partial v\, \partial u}(F_X(x), F_Y(y))\, f_Y(y)f_X(x).$$

By definition,

$$\frac{\partial^2 C}{\partial u\, \partial v}(u, v) = c(u, v),$$

so

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y)).$$

□

Given data $(x_1, y_1), \ldots, (x_n, y_n)$, the log likelihood of a bivariate copula model can be written as

$$\ell = \sum_{i=1}^{n} \log f_X(x_i) \; + \sum_{i=1}^{n} \log f_Y(y_i) \; + \sum_{i=1}^{n} \log c(F_X(x_i), F_Y(y_i)).$$

It is natural to decompose this as

$$\ell = \ell_M + \ell_C,$$

where ℓ_M is the marginal contribution and ℓ_C is the copula contribution.

For a full parametric model, we maximize the log likelihood over both the marginal and copula parameters. In practice this is usually done with software, and we must be careful about local maxima. If the main interest is the copula itself, a semiparametric approach replaces the unknown marginals by empirical estimates.

Definition 9.6 The **pseudo-observations** associated with data x_{ij} are

$$u_{ij} = \frac{\operatorname{rank}(x_{ij})}{n+1},$$

where the rank is computed within each margin. These transform the observed marginal scales to approximate uniform scores.

After forming pseudo-observations, we maximize the **copula log likelihood**

$$\ell_C = \sum_{j=1}^{n} \log c(u_{1j}, \dots, u_{dj}).$$

For one-parameter copulas, we may also estimate Kendall's tau first and then recover the copula parameter from the relationship $\theta = h(\tau_K)$ if h is known.

Goodness of Fit and Simulation

Chi-squared tests can be used in principle, but sparse cells often make them unreliable. Instead, we can use the following diagnostic based on the probability integral transform.

Proposition 9.7 Suppose the marginal distribution functions are continuous and the relevant conditional distributions are continuous for almost every conditioning value. If $U = F_X(X)$, $V = F_Y(Y)$, and $C(u, v)$ is the correct copula, then conditional transforms such as

$$F_{U|V}(U \mid V = v) \qquad \text{and} \qquad F_{V|U}(V \mid U = u)$$

have the $\operatorname{Unif}(0, 1)$ distribution.

Proof. Consider the first transform. For any fixed value v such that the conditional distribution of U given $V = v$ is continuous, the random variable $U \mid V = v$ has conditional distribution function $F_{U|V}(\cdot \mid v)$. Therefore, by the probability integral transform applied conditionally,

$$F_{U|V}(U \mid V = v) \mid V = v \sim \operatorname{Unif}(0, 1).$$

Equivalently, for $0 \le t \le 1$,

$$\mathbb{P}\big(F_{U|V}(U \mid V = v) \le t \mid V = v\big) = t.$$

Since this conditional distribution does not depend on v, averaging over V gives

$$\mathbb{P}\big(F_{U|V}(U \mid V) \le t\big) = \mathbb{E}\left[\mathbb{P}\big(F_{U|V}(U \mid V) \le t \mid V\big)\right] = \mathbb{E}[t] = t.$$

Hence $F_{U|V}(U \mid V)$ has the Unif$(0,1)$ distribution.

The proof for $F_{V|U}(V \mid U)$ is identical, with the roles of U and V reversed. Thus both conditional transforms are uniformly distributed when the copula model is correct. $\square$

From Proposition 9.7, we can check the fit of a copula model by plotting the conditional transforms against the uniform distribution, for example with a P-P plot or a Q-Q plot, or by performing a distributional hypothesis test such as the Kolmogorov–Smirnov test.

Simulation is used to generate samples from joint distributions, compute risk measures, and test strategies. We examine Gaussian copulas, Student t copulas, and Archimedean copulas.

To simulate a Gaussian copula with a positive semidefinite correlation matrix $\mathbf{\Sigma}$, first simulate

$$\boldsymbol{V} = (V_1, \dots, V_n), \qquad V_1, \dots, V_n \overset{\text{iid}}{\sim} \text{Unif}(0,1), \qquad \text{and} \qquad Z_k = \Phi^{-1}(V_k),$$

so that $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ is a standard normal random vector. Then choose a matrix $\mathbf{A}$ such that

$$\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma},$$

and set

$$\boldsymbol{X} = \mathbf{A}\boldsymbol{Z}.$$

The matrix $\mathbf{A}$ is often obtained by Cholesky decomposition, in which case $\mathbf{A}$ is lower triangular. Another approach is spectral decomposition: if

$$\mathbf{\Sigma} = \mathbf{O}\mathbf{D}\mathbf{O}^T$$

with $\mathbf{O}$ orthonormal and $\mathbf{D}$ diagonal with nonnegative entries, then we may take the symmetric square root

$$\mathbf{A} = \mathbf{O}\mathbf{D}^{1/2}\mathbf{O}^T,$$

so that

$$\mathbf{A}\mathbf{A}^T = \mathbf{O}\mathbf{D}^{1/2}\mathbf{O}^T\mathbf{O}\mathbf{D}^{1/2}\mathbf{O}^T = \mathbf{O}\mathbf{D}\mathbf{O}^T = \mathbf{\Sigma}.$$

Finally define

$$U_k = \Phi(X_k), \quad k = 1, \ldots, n.$$

Then $\boldsymbol{U} = (U_1, \ldots, U_n)$ has the desired Gaussian copula.

To simulate a Student t copula with dependence matrix $\boldsymbol{\Sigma}$ and degrees of freedom $\nu > 0$, first simulate a correlated normal vector $\boldsymbol{X}$ with correlation matrix $\boldsymbol{\Sigma}$. Next simulate

$$R \sim \chi^2_\nu$$

independently, for example by the probability integral transform, and set

$$\boldsymbol{Y} = \boldsymbol{X}\sqrt{\frac{\nu}{R}}.$$

Then $\boldsymbol{Y}$ has a multivariate Student t distribution with the required degrees of freedom and dependence matrix. When $\nu > 2$, this is also its correlation matrix. Define

$$U_k = F_\nu(Y_k), \quad k = 1, \ldots, n,$$

where F_ν is the univariate Student t distribution function. This gives the required t copula sample.

For an Archimedean copula, write $\psi = \phi^{-1}$ and suppose that ψ is completely monotone. Then ψ is the Laplace transform of a positive frailty variable W:

$$\mathbb{E}\left[e^{-tW}\right] = \psi(t).$$

Simulate W from this copula-specific frailty distribution and independently simulate

$$E_1, \ldots, E_n \overset{\text{iid}}{\sim} \operatorname{Exp}(1).$$

Then set

$$U_k = \psi(E_k/W), \quad k = 1, \ldots, n.$$

The vector $\boldsymbol{U} = (U_1, \ldots, U_n)$ has the required Archimedean copula. Equivalently, we may simulate independent uniforms V_k and take $E_k = -\log V_k$.

More generally, if we want a multivariate random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ with copula C and marginals $F_1, \ldots, F_n$, we first simulate

$$\boldsymbol{U} = (U_1, \ldots, U_n) \sim C$$

and then set

$$X_k = F_k^{-1}(U_k), \quad k = 1, \ldots, n.$$

Lecture 10 — Tuesday, November 17, 2015

6. Models for Market Risk and Other Economic Risks

Losses from different sources may be dependent because they are driven by common economic conditions such as inflation, stock prices, interest rates, unemployment, or broader macroeconomic trends. Econometric models are used to capture this structural dependence, but they bring their own risks, including model risk, basis risk, people risk from insufficient competency, and regulatory risk.

Common univariate econometric models include stock price models for market risk and embedded options, interest rate models for long-term liabilities and annuities, inflation models for deferred liabilities and projected future costs, and macroeconomic models for business cycles. These models for individual series become building blocks for multivariate economic scenario generators.

Deterministic simulations correspond to scenario or stress testing. They can be very detailed, but they do not supply a probabilistic interpretation. Stochastic simulations generate many sample paths for each variable of interest, treat each path as equally likely, and are used for risk measures and economic capital.

An **economic scenario generator** models the joint distribution of several economic series, typically including stock returns, interest rates, and inflation, and often also credit spreads, property indices, exchange rates, unemployment, or GDP growth. It is used to project asset and liability cash flows, assess profitability and risk, explore pricing, test risk transfer strategies, and evaluate the range of outcomes, including stress scenarios.

Financial Return Stylized Facts

Financial return series exhibit persistent empirical features often called **stylized facts**. The usual discussion concerns daily, weekly, or monthly returns; tick data have their own regularities, while lower-frequency patterns are harder to identify because observations are scarcer and stationarity is more difficult to defend.

The returns under discussion are usually log returns. Although they are not iid, their serial correlation is generally small:

$$\mathrm{Corr}(R_t, R_{t+1}) \approx 0.$$

Let $\mathcal{F}_{t-1} = \sigma(R_{t-1}, R_{t-2}, \dots)$ denote the information in past returns. Absolute or squared returns show strong serial correlation; by contrast, the short-horizon conditional mean is usually close to zero:

$$\mathbb{E}[R_t \mid \mathcal{F}_{t-1}] \approx 0,$$

while volatility changes over time and tends to cluster. Conditional standard deviations therefore change continually in a partly predictable way. Extreme returns appear in bunches, and return distributions are often heavy-tailed.

If R_t denotes a return with mean μ and standard deviation σ, then **kurtosis** is

$$\kappa = \frac{\mathbb{E}\left[(R_t - \mu)^4\right]}{\sigma^4},$$

and **excess kurtosis** is

$$\kappa_{\mathrm{ex}} = \kappa - 3.$$

Distributions with positive excess kurtosis are **leptokurtic**, while those with negative excess kurtosis are **platykurtic**; the normal distribution has zero excess kurtosis and is **mesokurtic**. Leptokurtosis often accompanies heavy empirical tails, but the two notions are not equivalent, and kurtosis is not finite for every heavy-tailed distribution. Return series tend to be leptokurtic, with sample excess kurtosis often in the range of 5 to 10 or even higher.

As the return horizon increases, for example from monthly to annual returns, these effects generally become less pronounced. There is less visible volatility clustering, the data look more like iid observations, and the tails look less heavy.

Multivariate return series also show stylized facts. Contemporaneous returns may be correlated, correlations vary over time, absolute returns show cross-correlation, and extreme returns in one series often coincide with extreme returns in others. The DAX daily log returns in Figure 4, computed from Yahoo Finance adjusted closes from 2000 through 2024, show the usual volatility clustering and bunching of large returns.

Figure 5 compares BMW and Siemens daily log returns over the same period. The scatterplot shows contemporaneous positive association and joint lower-tail observations; the three highlighted dates are March 12, 2020, October 15, 2008, and March 17, 2008.

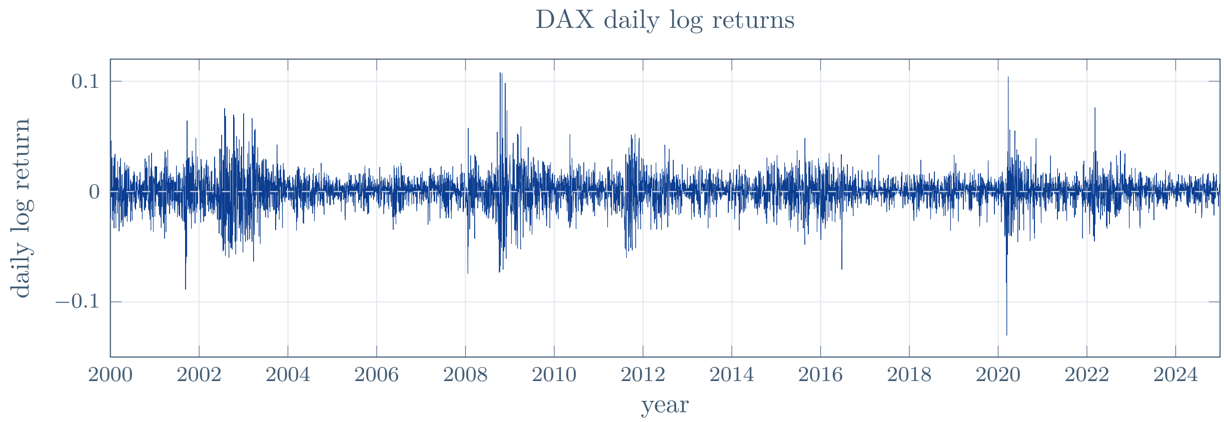


FIGURE 4

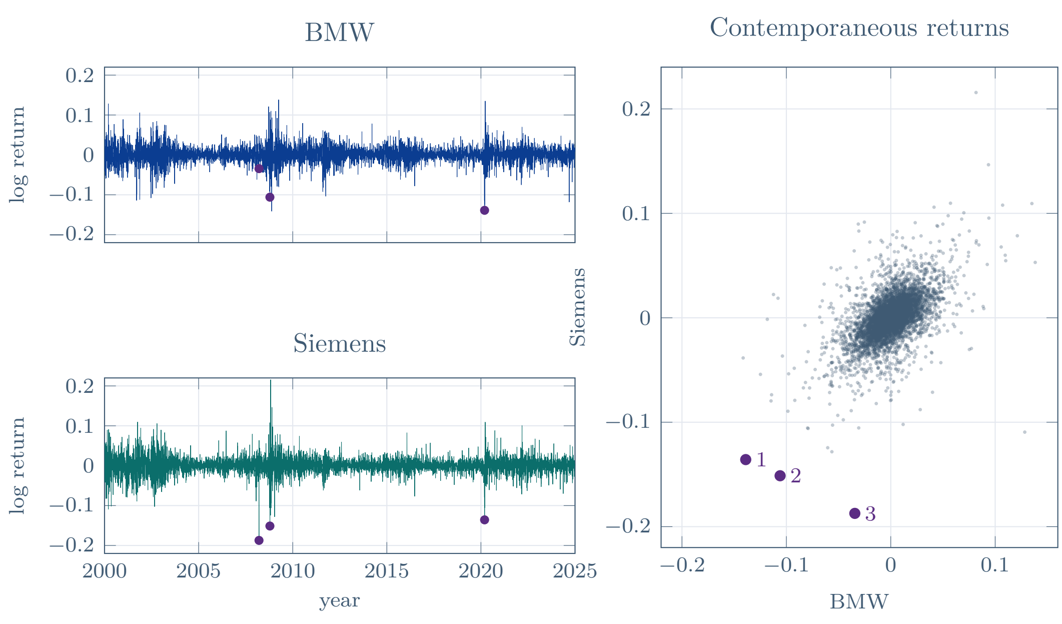


FIGURE 5

Editorial Note The course (and more specifically, the *Quantitative Risk Management* textbook) originally presented the stylized facts illustrated in Figure 4 and Figure 5 using return series from the late 1980s and early 1990s, capturing important events such as Black Monday, Black Friday, and the collapse of the Soviet Union. I have updated the data to show more recent returns, since we have seen several important events since then, including the dot-com bubble, the global financial crisis, and the COVID-19 pandemic. The stylized facts are still very much present in the data.

In many situations in finance, we assume that variables are normally distributed (for example when using the independent lognormal model for stock returns). It is therefore important to check whether this assumption is reasonable. The Jarque–Bera test is a commonly used test of normality that is based on the sample skewness and excess kurtosis.

Definition 10.1 The **Jarque–Bera statistic** for a sample of size n is

$$JB = \frac{n}{6}\left(g_1^2 + \frac{g_2^2}{4}\right),$$

where

$$g_1 = \frac{(1/n)\sum_{i=1}^n (X_i - \overline{X})^3}{\left((1/n)\sum_{i=1}^n (X_i - \overline{X})^2\right)^{3/2}}$$

is the sample skewness coefficient and

$$g_2 = \frac{(1/n)\sum_{i=1}^n (X_i - \overline{X})^4}{\left((1/n)\sum_{i=1}^n (X_i - \overline{X})^2\right)^2} - 3$$

is the sample excess kurtosis. Under the null hypothesis of normality, JB is asymptotically χ_2^2, so large values argue against normality.

Stock Price Models

For real-world loss modelling, we use real-world probabilities rather than risk-neutral probabilities. If S_t is the asset price, the log return is

$$Y_t = \log\left(\frac{S_{t+1}}{S_t}\right).$$

Risk-neutral probabilities are used for pricing and hedging, not for projecting real-world losses. Before choosing a Gaussian or Brownian model for $\{Y_t\}$, we should check whether it fits the historical data.

Figure 6 shows monthly log returns for the S&P 500 Total Return Index and a rolling 12-month annualized volatility estimate, computed from Yahoo Finance adjusted closes from 1988 through 2010. We use total return, so dividends are assumed to be reinvested. The important features are fat tails, especially in the left tail, nonconstant volatility, volatility clustering, more downside than upside during high-volatility periods, and some autocorrelation, but not much.

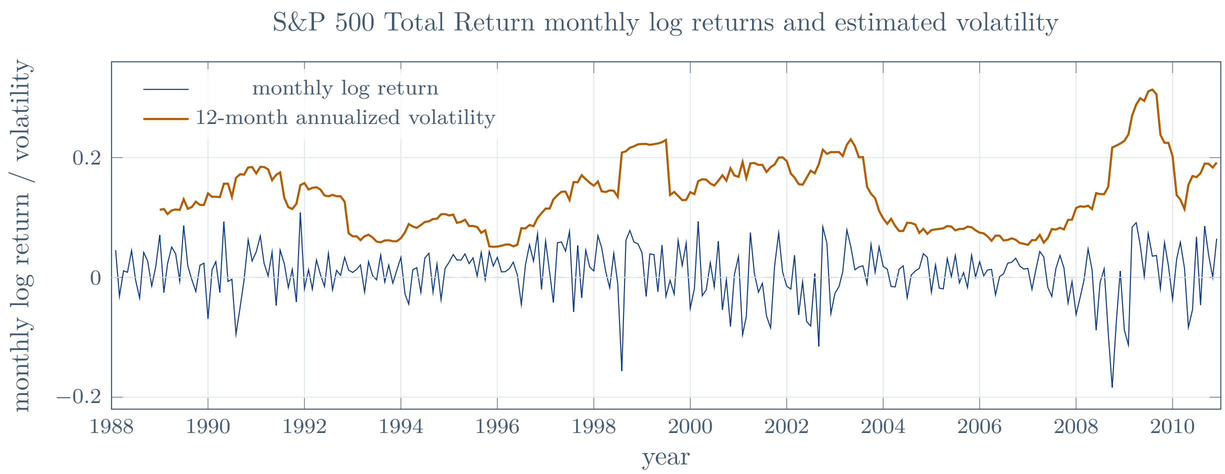


FIGURE 6

Definition 10.2 In the **independent lognormal model**, the log returns $\{Y_t\}$ are iid normal random variables with mean μ and variance σ^2:

$$Y_t \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2), \qquad Y_t = \mu + \sigma\varepsilon_t, \qquad \text{and} \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$

so

$$S_{t+1} = S_t e^{Y_t} = S_t e^{\mu + \sigma\varepsilon_t}.$$

This is the discrete-time version of geometric Brownian motion, with parameter vector

$$\boldsymbol{\theta} = (\mu, \sigma).$$

The same model generates lognormal annual, monthly, weekly, and daily data, and it is the model behind the Black–Scholes option pricing formula. It is analytically convenient and tractable, but it produces thin tails, fixed volatility, no volatility clustering, and essentially no autocorrelation structure. The Q-Q diagnostic in Figure 7 shows a much heavier empirical left tail than the normal model predicts.

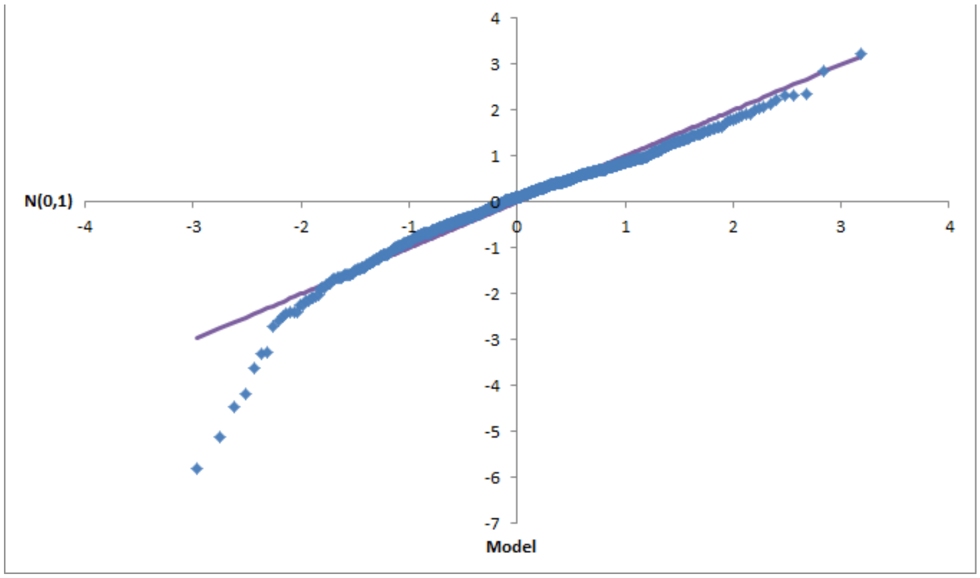


FIGURE 7

The **GARCH(1, 1) model** keeps a Gaussian innovation but allows time-varying conditional variance:

$$Y_t = \mu + \sigma_t \varepsilon_t \qquad \text{and} \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$

with

$$\sigma_t^2 = a_0 + a_1(Y_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2, \qquad a_0 > 0, \quad a_1, \beta \geq 0, \qquad \text{and} \qquad a_1 + \beta < 1.$$

The coefficient β drives volatility persistence, and the parameter set is

$$\Theta = \{\mu, a_0, a_1, \beta\}.$$

The model itself has no autocorrelation in returns, although it can be combined with an AR(1) mean equation. More generally, a GARCH model may include several lagged squared shocks and several lagged conditional variances; naming conventions for the two orders vary, so the recursion should always be stated explicitly. It remains a single factor model, capturing stochastic volatility, volatility clustering, and heavier unconditional tails than either the independent lognormal model or a basic ARCH model, but not the association between high volatility and low mean returns. The GARCH Q-Q plot in Figure 8 improves the centre of the distribution while leaving discrepancies in the tails.

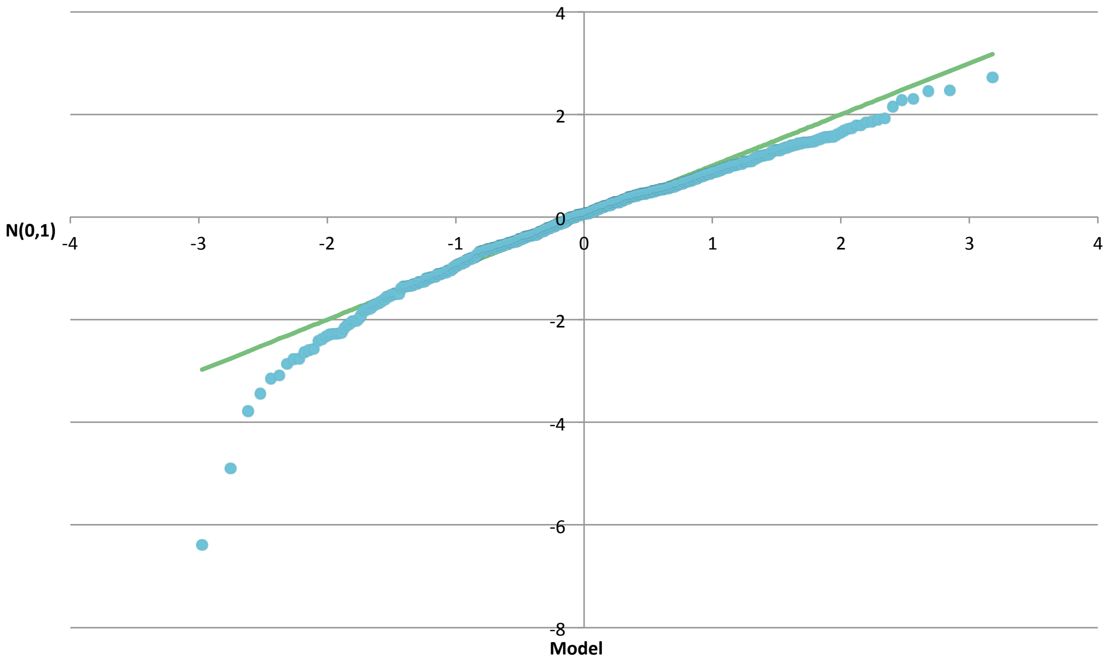


FIGURE 8

A simple **regime-switching lognormal model** assumes that the return distribution depends on an unobserved regime ρ_t. In regime k,

$$Y_t \mid \{\rho_t = k\} \sim \mathcal{N}(\mu_k, \sigma_k^2),$$

and the regime follows a Markov chain with transition probabilities

$$\mathbb{P}(\rho_t = k \mid \rho_{t-1} = j) = p_{jk}.$$

With two regimes, the parameter set is

$$\Theta = \{\mu_1, \sigma_1, \mu_2, \sigma_2, p_{12}, p_{21}\}.$$

The probabilities p_{11} and p_{22} are not listed separately because each row of the transition matrix sums to one, so $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$. Models with K regimes are possible, though two regimes are often enough; with three regimes there are twelve parameters. This model captures stochastic volatility in a simple way, volatility clustering, some autocorrelation, fat tails, and the tendency for high-volatility periods to be associated with lower returns. Its cost is greater complexity: it has two factors, since both returns and regimes evolve stochastically, and residuals are not clearly defined. The RSLN-2 residual Q-Q plot in Figure 9 is much closer to a straight line than the ILN diagnostic.

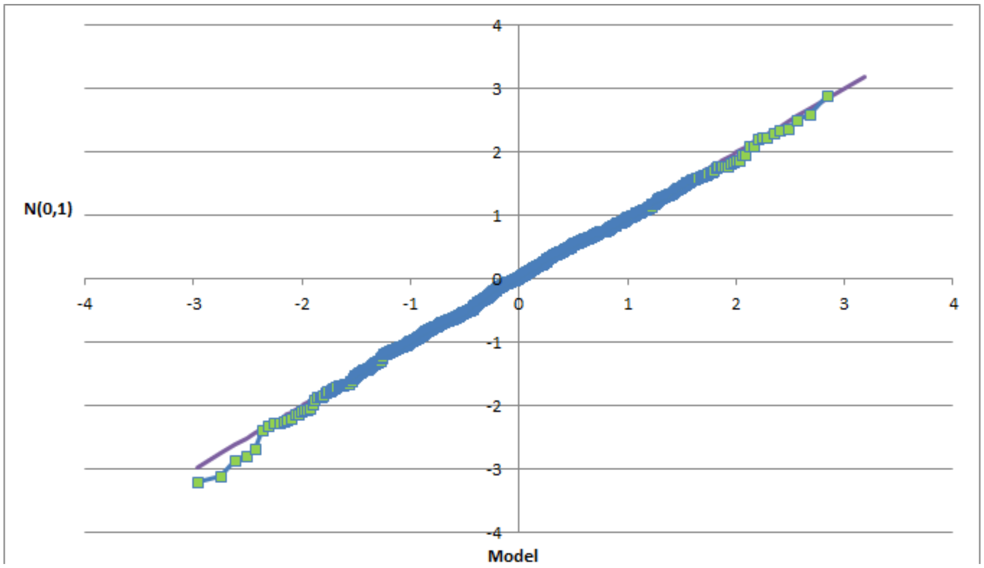


FIGURE 9

Other widely used models include regime-switching AR, ARCH, and GARCH models, as well as **stochastic volatility models** of the form

$$Y_t = \mu + \sigma_t \varepsilon_t^Y \qquad \text{and} \qquad \log \sigma_t = \nu_t = m + a(\nu_{t-1} - m) + \sigma_\nu \varepsilon_t^\nu,$$

where the innovation vector

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_t^Y \\ \varepsilon_t^\nu \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}) \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

For the usual stationary specification, $|a| < 1$, $\sigma_\nu > 0$, and $|\rho| \leq 1$. The correlation parameter allows return shocks and volatility shocks to move together.

Economic Scenario Generators

Economic scenario generators (**ESGs**) are commonly used in insurance, pensions, and banking. In insurance, they are important for Own Risk and Solvency Assessment (ORSA), including the US solvency framework and Solvency II. An economic scenario generator may be deterministic or stochastic, and design choices include the model frequency, projection horizon, included variables, multivariate structure, and whether the emphasis is on best estimate behaviour or tail risk.

Real-world models are used for projecting actual cash flows, profit testing, setting reserves, calculating risk measures, and testing risk mitigation strategies. Risk-neutral or market-consistent models are used for pricing and hedging financial guarantees and derivatives.

A good economic scenario generator should have sound economic properties, internal consistency, realistic ranges and relationships, an adequate fit to data, and acceptable computational efficiency. It should also be tractable enough to work as a module inside a larger asset–liability model. Sound economic properties include valid short-term relationships, such as no free lunches, the principle that extra expected return should come with extra risk, interest rate and inflation differentials affecting exchange rates, and dependence in extreme paths. They also include feasible long-term relationships, where economic theory gives some guidance about causes and effects over medium and long horizons. Calibration requires judgment as well as statistical and economic theory, especially because too little history gives unstable estimates and too much history may ignore structural change. Testing should include in-sample and out-of-sample validation, robustness across calibration periods, and sensitivity to parameters.

Model Risk and Model Governance

Every quantitative risk management calculation depends on a model, and every model is a simplification of reality. A model is a mathematical representation of some aspect of the real world, used to analyze a complex problem so that decisions can be made. Examples include a fixed interest assumption such as 6% per year, CAPM, Markov regime-switching models, compound Poisson loss models with lognormal severities, stock price dynamics such as

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t,$$

service tables for pension plans, and combinations of these components.

Some objects used with models are not themselves models. The Black–Scholes option price, VaR, CTE, and Monte Carlo simulation are outputs, risk measures, or computational methods rather than full representations of the underlying real-world mechanism.

The two common mistakes are to trust a model too much or to dismiss it too quickly. A sensible modelling process starts from the problem requirements, explores the data, formulates a modelling hypothesis, calibrates to appropriate data, tests goodness of fit, compares or adapts alternatives, and then selects the model that is fit for purpose. Implementation also requires testing and development protocols, review, calibration procedures, interface design, decisions about output formats, documentation, and controls. Minor models are often internal to the company, while major valuation models may be purchased from software firms. **Model risk** is the risk of loss from using a model that is not fit for purpose, and it can be mitigated by good model governance. **Model governance** includes policies and procedures for model development, testing, implementation, use, and review. It also includes controls such as independent validation, documentation standards, and regular review cycles.

Lecture 11 — Tuesday, November 24, 2015

7. Capital, Credit Risk, and Regulation

Capital and Balance Sheets

In a financial institution, **capital** is the amount held or required to be held to underpin the risk of loss in value of exposures, businesses, and other activities, so as to protect depositors and general creditors against loss. According to OSFI, it is a source of financial support that protects the institution against unexpected losses and therefore contributes to safety and soundness.

On the balance sheet, capital is the surplus of assets over liabilities. Assets include current assets, fixed assets, and other assets. Liabilities include current liabilities and fixed liabilities, and the residual claim belongs to owners' equity. Figure 10 compares this book balance-sheet view with economic capital measured using market or market-consistent values.

Economic capital is the amount of capital a financial institution thinks it should have. **Regulatory capital** is the amount of capital the regulator requires the institution to have.

Economic capital uses market or market-consistent valuation of assets and liabilities. It is natural to distinguish **available economic capital**, which is the current excess of the market value of

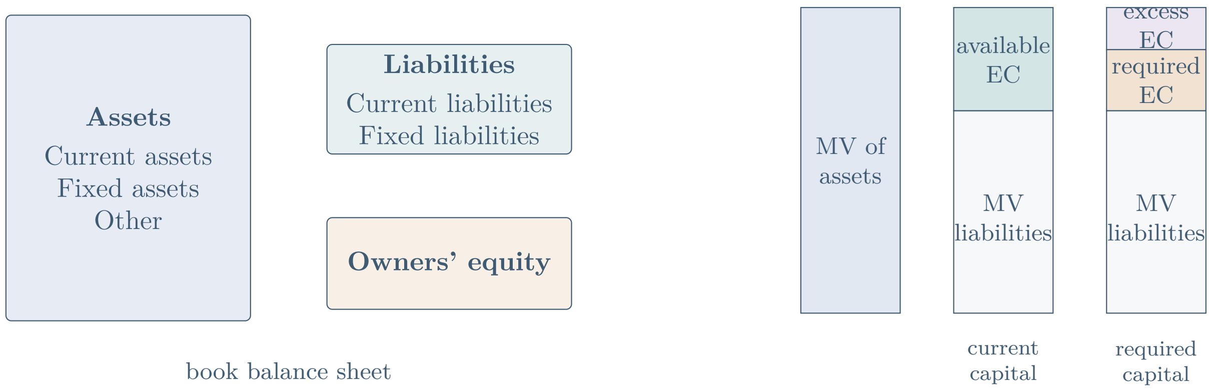


Figure 10

assets over the market value of liabilities, from **required economic capital**, which is the amount needed to support the firm's risks at a specified confidence level.

Capital is not the same thing as reserves. In banking, **reserves** are liquid, high-quality assets, such as deposits at a central bank or government bonds. A **reserve requirement** is a constraint on assets based on existing liabilities. Capital instead refers to the sources of funding on the right-hand side of the balance sheet. A **capital requirement** is a constraint on sources of funds based on the risks of existing assets.

Capital acts as a buffer against large unexpected losses, protects depositors and other claimants, and reassures external investors and rating agencies about the firm's financial health. **Capital management** is the continuing process of raising and maintaining enough capital to support planned operations, then allocating it in a way that recognizes the risk in each activity.

Bank Equity

High equity requirements are often criticized on the grounds that equity is expensive and that stronger capital requirements would affect credit markets adversely. The counterargument is that the required return on equity falls as leverage falls; tax subsidies, implicit guarantees, and other policies that favour debt are distortions; and the claim that high debt provides discipline has little empirical support.

Several common arguments confuse capital requirements with reserve requirements. Increased equity does not force a bank to set aside funds that could otherwise be lent, because lending can be funded by equity. Increased equity requirements do not mechanically raise funding costs merely because the cost of equity exceeds the cost of debt, since the required return on equity decreases when the debt-to-equity ratio decreases. A higher return on equity at high leverage reflects a higher risk premium, not free value creation.

Debt tax shields and subsidies can make debt privately attractive, but that is an argument about distortions rather than an argument for excessive leverage. Debt may create some market discipline for managers, but high leverage also gives incentives to take risk and relies on a fragile capital structure with substantial short-term debt. Discipline can also come from equity holders and regulation. A further objection is that higher capital requirements may push banking activities into less regulated markets, but many shadow banking activities still relied on backstops from regulated institutions.

A Modigliani–Miller style relation for the required return on equity is

$$R_S = R_0 + \frac{B}{S}(1 - T_C)(R_0 - R_B),$$

where R_S is the required return on equity, R_0 is the unlevered required return, B/S is the debt-to-equity ratio, T_C is the corporate tax rate, and R_B is the return on debt. The formula illustrates why required return on equity changes with leverage.

Economic Capital

Required **economic capital** for a firm as a whole is the notional amount of capital needed to withstand an extreme but rare loss event after taking into account all interactions among the risks in the portfolio. Economic capital may need to reflect interest rate risk, currency risk, credit risk, market risk, operational risk, and other material risks. It is often assessed using a top-down risk measure in the sense of Definition 3.1 applied to aggregate firm losses.

One common approach is to choose a very high confidence level, perhaps associated with a target credit rating, and then define economic capital as the VaR from Definition 3.2 at that level, possibly less expected loss. Historical annual survival rates for AA-rated companies have been approximately 99.98%. Setting VaR at this confidence level corresponds to a one-in-5000-year loss. Other approaches introduce subjective information through Bayesian methods or use a more realistic confidence level together with a safety factor, as in the Basel market risk charge.

Actual equity below required economic capital may signal that the firm is taking unwarranted or ill-advised risks; an ample surplus, by contrast, indicates greater resilience to shocks and more capacity to pursue growth opportunities. Economic capital therefore conveys important information to bondholders and shareholders, and its disclosure is connected to Pillar III under the Basel Accord.

Under Pillar II of Basel, banks are expected to maintain an **internal capital adequacy assessment process** (**ICAAP**). Principle 1 says that banks should have a process for assessing overall capital adequacy in relation to their risk profile and a strategy for maintaining capital levels. The ICAAP depends on the nature, size, and complexity of the bank. Advanced banks are expected to use a bottom-up approach with analytical tools.

The ICAAP components include an explicit capital policy approved by the board, including objectives, time horizon, and limits; a capital planning process with responsibilities assigned across the capital adequacy assessment process; a process to identify all risks inherent in the bank's portfolios; a system to measure all identified risks; and regular, adequate internal reporting to the board and senior management.

Regulatory Capital

Regulatory capital rules may be based on factors or formulas, as in United States risk-based capital requirements for insurance; based on scenarios, as in some Canadian life insurance requirements; or based on ideas from economic capital, as in Solvency II and Basel II/III. The goals of capital regulation include stability of the financial system, macroprudential regulation, adequate capital levels at individual institutions, transparency, comparability, implementation across a wide variety of institutions, and avoidance of procyclicality in capital requirements.

Procyclicality arises because losses in recessions erode bank capital while risk-based capital requirements, such as those in Basel II, often rise at the same time. If banks cannot quickly raise sufficient new capital, their lending capacity falls and a credit crunch may follow. Relaxing requirements in bad times may reduce the contractionary effect on credit supply, but it may also increase bank failure probabilities precisely when loan defaults are high. Basel III is a compromise: it reinforces the quality and quantity of minimum required capital while also creating mandatory buffers, including a capital conservation buffer and a countercyclical buffer, to be built up in good times and released in bad times.

For the trading book under Basel III, different risk measures use different confidence levels and horizons:

Risk	Risk measure	Confidence level and horizon
Market risk	VaR	99%, 10 days
Incremental risk charge	VaR	99.9%, 1 year
Credit risk and counterparty credit risk	VaR	99.9%, 1 year
Credit valuation adjustment	VaR	99%, 10 days.

The trading portfolio contains both market and credit risk. Its market exposures include equities, commodities, bonds and loans, over-the-counter derivatives, and credit derivatives such as credit default swaps and collateralized debt obligations; the 10-day, 99% portfolio VaR also includes spread risk (or specific risk). The stressed VaR component is instead calibrated from a period of significant financial stress, such as 2007–2008.

The market risk charge is split into ordinary and stressed components:

$$C = C_{MV} + C_{SV},$$

where

$$C_{MV} = \max\{\mathrm{VaR}_{t-1}, m_b\overline{\mathrm{VaR}}\}$$

and

$$C_{SV} = \max\{\mathrm{sVaR}_{t-1}, m_s\overline{\mathrm{VaR}}\}.$$

Here $\overline{\mathrm{VaR}}$ and $\overline{\mathrm{sVaR}}$ denote average ordinary and stressed VaR values used in the regulatory calculation, and the stressed component is calibrated to a period of significant historical stress.

Trading Book Credit Risk

The **incremental risk charge** (**IRC**) captures default and migration risks in trading book positions that are incremental to the risks captured by market VaR. It applies to trading book positions in unsecuritized credit products and is measured over a one-year capital horizon under an assumption of constant risk.

The revisions to the Basel II market risk framework finalized in July 2009 required banks to develop a methodology for calculating the incremental risk charge. The **capital horizon** is the time over which credit events are measured, usually one year. The **liquidity horizon** is the frequency at which the portfolio or transaction is rebalanced to a target level of risk, such as three months, six months, or one year.

Basel guidance requires liquidity horizons to reflect the time needed to sell a position or hedge all material risks in a stressed market without materially affecting prices. The floor is three months, with longer horizons expected for non-investment-grade or concentrated positions.

Under a **constant position assumption**, a deteriorated name remains in the portfolio for the whole capital horizon. Under a **constant risk assumption**, positions whose credit characteristics improve or deteriorate are replaced after the liquidity horizon so that the credit characteristics are equivalent to those at the start of the liquidity horizon. Default probabilities for the liquidity horizon can be estimated from ratings data, KMV or other market-implied sources, or roots of transition matrices.

Example 11.1 Suppose a name rated investment grade can remain investment grade, migrate to non-investment grade, or default after one period. The corresponding transition probabilities are 0.2, 0.7, and 0.1 from investment grade, and 0.1, 0.1, and 0.8 from non-investment grade. Over two periods, the default probability under constant position is

$$0.1(1) + 0.7(0.8) + 0.2(0.1) = 0.68.$$

Under constant risk, once the name survives the first period it is replaced by a fresh investment-grade exposure, so the two-period default probability is

$$0.1(1) + 0.9(0.1) = 0.19.$$

For constant risk calculations, the PD scaling factor is

$$\text{PD scaling factor} = \frac{\text{PD with rebalancing}}{\text{PD buy-and-hold}}.$$

For a one-month liquidity horizon and one-year capital horizon, the following scaling factors are used:

Rating	**Moody's**	**MKMV**
All investment grade	0.18	0.16
Ba	0.34	0.23
B	0.60	0.35
Caa–C	1.30	1.92.

Trading book credit risk can be divided into counterparty credit risk, issuer or borrower risk, and structured credit risk. **Counterparty credit risk** arises in derivatives, including credit derivatives, and is affected by collateral, guarantees, and other mitigation. **Issuer or borrower risk** arises in bonds, loans, and credit default swaps. **Structured credit** combines issuer or borrower risk with counterparty risk. Basel III capital for trading book credit risk includes CreditVaR for counterparty credit risk, the incremental risk charge for default and migration risk in bonds, loans, and credit default swaps, securitization and correlation trading capital, and CVA VaR.

For Basel credit risk capital, the regulatory formula has the same broad structure as the asymptotic single factor model, with heterogeneous exposures and default and migration risk:

$$\text{Capital} = \sum_j \mathrm{LGD}_j\, \mathrm{EAD}_j \left[\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_j) - \sqrt{\rho_j}\,\Phi^{-1}(0.001)}{\sqrt{1-\rho_j}}\right) - \mathrm{PD}_j\right] \mathrm{MF}(M_j, \mathrm{PD}_j).$$

The calculation of risk-weighted assets relies on four quantitative risk components: probability of default, exposure at default, loss given default, and maturity. The asset correlation parameter is another model parameter set by the Accord. We will discuss the Basel credit risk formula in more detail later in the course.

Capital Allocation and Performance

Suppose an insurer's aggregate loss from n business lines is

$$L = \sum_{i=1}^{n} L_i,$$

so the economic capital for the firm as a whole is

$$C = \rho(L),$$

where ρ is the chosen risk measure; see Definition 3.1. Economic capital can be allocated to individual business lines for profitability analysis, pricing, risk budgeting, strategic planning, and incentives.

Definition 11.2 For a portfolio with losses

$$L(\boldsymbol{w}) = \sum_{n=1}^{N} w_n L_n$$

and portfolio risk

$$R(\boldsymbol{w}) = \rho\left(\sum_{n=1}^{N} w_n L_n\right),$$

the **stand-alone contribution** of instrument or line n is

$$C_n^{SA} = \rho(w_n L_n),$$

the **incremental contribution** is

$$C_n^I = R(\boldsymbol{w}) - \rho\left(\sum_{k \neq n} w_k L_k\right),$$

and the **marginal contribution** (or **Euler contribution**) is

$$C_n^M = w_n \frac{\partial R(\boldsymbol{w})}{\partial w_n},$$

when the derivative exists.

Stand-alone contributions and incremental contributions are generally not additive. Marginal contributions are used because, under appropriate homogeneity and regularity assumptions, they are additive.

Definition 11.3 A function f is **positive homogeneous of degree $\boldsymbol{k}$** if

$$f(\alpha \boldsymbol{x}) = \alpha^k f(\boldsymbol{x})$$

for all $\alpha > 0$.

Theorem 11.4 (Euler's theorem) Let the domain of f be an open cone, so that $\alpha \boldsymbol{x}$ remains in the domain for every $\alpha > 0$. If f is continuously differentiable, then f is positive homogeneous of degree k if and only if

$$\boldsymbol{x} \cdot \nabla f(\boldsymbol{x}) = \sum_{n=1}^{N} x_n \frac{\partial f}{\partial x_n}(\boldsymbol{x}) = k f(\boldsymbol{x}).$$

Suppose the risk measure ρ is positive homogeneous. Set

$$f(\boldsymbol{w}) = \rho(L_{\boldsymbol{w}}) = \rho\left(\sum_{n=1}^{N} w_n L_n\right).$$

Then f is positive homogeneous of degree one, and Euler's theorem gives

$$\rho(L_{\boldsymbol{w}}) = f(\boldsymbol{w}) = \boldsymbol{w} \cdot \nabla f(\boldsymbol{w}) = \sum_{n=1}^{N} w_n \frac{\partial \rho(L_{\boldsymbol{w}})}{\partial w_n} = \sum_{n=1}^{N} C_n^M.$$

This additivity is one of the main reasons Euler allocation is used in practice. Applications include risk management, hedging, risk-adjusted performance measurement, and macroprudential financial regulation.

Under the differentiability and density conditions that justify conditioning at a portfolio quantile, and assuming L has a continuous distribution and $\sigma(L) > 0$, important marginal contributions take the explicit forms

$$C_n^{\mathrm{VaR}} = w_n \mathbb{E}\left[L_n \mid L = \mathrm{VaR}_\alpha(L)\right],$$
$$C_n^{\mathrm{CVaR}} = w_n \mathbb{E}\left[L_n \mid L > \mathrm{VaR}_\alpha(L)\right],$$

and

$$C_n^{\sigma} = w_n \frac{\mathrm{Cov}(L_n, L)}{\sigma(L)}.$$

Traditional performance measures such as return on assets and return on equity ignore the amount of risk taken to achieve the return. For example, if assets are $A = 100$, liabilities are 84, equity is $\mathrm{Eq} = 16$, and net income is $I = 4.0$, then

$$\mathrm{ROA} = \frac{4.0}{100} = 0.04 \qquad \text{and} \qquad \mathrm{ROE} = \frac{4.0}{16} = 0.25.$$

The return looks attractive, but the calculation does not say how much risk was taken.

Risk-adjusted performance measurement either adjusts the numerator, for example by deducting expected losses or focusing only on variation arising from random factors, or adjusts the denominator by comparing return to required economic capital. A generic performance measure has the form

$$\text{Performance} = \frac{\text{reward}}{\text{risk}}.$$

For example, the Sharpe ratio is

$$\mathrm{SR} = \frac{\mu - r}{\sigma}.$$

Using capital in the denominator allows comparison across business units inside a financial institution.

A general **risk-adjusted return on capital** (**RAROC**) style measure is

$$\text{RAROC} = \frac{\text{revenues} - \text{costs} - \text{expected losses}}{\text{capital}}.$$

RAROC can be used to assess capital requirements by business line, product, or customer; measure risk consistently across business units; promote fair and consistent risk-adjusted performance measures; and improve the relation between risk and reward. Expected losses should capture losses arising from all sources, including credit, market, and operational risk. The denominator may be stand-alone capital or the unit's contribution to overall firm capital. When RAROC is measured in advance, expected revenues, losses, and costs are used. For performance evaluation, realized values may be used in the numerator.

Example 11.5 Suppose assets are 100, liabilities are 84, so equity and available economic capital are 16. If net income over the year is 4.0, required economic capital is 12, and risk-adjusted income is $4.0 - 1.5 = 2.5$, then

$$\text{RAROC} = \frac{2.5}{16} = 0.15625 \approx 15.6\%,$$

$$\text{RARORAC} = \frac{2.5}{12} = 0.2083 \approx 20.8\%,$$

and

$$\text{RORAC} = \frac{4.0}{12} = 0.3333 \approx 33.3\%.$$

RAROC measures return against the capital actually supplied by shareholders, while RARORAC measures return against required capital and is therefore useful for internal risk budgeting and comparison across business units.

Lecture 12 — Thursday, December 03, 2015

Credit Risk Models

Credit risk is the risk that the value of a portfolio changes because of unexpected changes in the credit quality of issuers or trading counterparties. It includes losses caused by default and losses caused by changes in credit quality before default. **Issuer risk** is the risk arising from the credit quality of issuers of securities such as bonds. **Counterparty risk** is the risk that the counterparty to a derivative contract or similar agreement defaults.

Credit risk models have two main uses: managing risk by determining portfolio loss distributions, computing risk measures, and allocating capital; and analyzing or pricing credit-sensitive securities such as credit default swaps.

A **static credit model** builds a future loss distribution from today's market factors using a finite set of random variables, whereas a **dynamic credit model** describes the credit market's evolution, usually through continuous-time stochastic processes.

A **structural model**, also called a **firm value model**, models default as a function of the value of the firm's assets and liabilities. A **reduced-form model** specifies a mathematical model for default times without modelling the firm value mechanism that produces default.

Credit risk modelling is difficult for three reasons. First, public information and data are often limited, making calibration hard and creating problems of asymmetric information such as moral hazard and adverse selection. Second, credit loss distributions are highly skewed: small gains occur frequently, but occasional large losses determine the tail. Third, default events are dependent, and simultaneous defaults have a major effect on portfolio tail risk. As the Vasicek loss distributions in Figure 1 illustrate, stronger default dependence can produce a much heavier right tail even when the individual default probabilities are unchanged.

Credit Ratings and Recovery

Credit ratings are assessments of credit quality produced by rating agencies such as Moody's and S&P. They provide information about default probability and potential recovery, and they affect who is able or willing to hold a bond issue.

Issuers pay agencies to rate their issues, making the bonds more attractive to traders and accessible to investors restricted to securities above a given grade. Bonds from the same issuer usually receive the same rating. These assessments exhibit momentum because agencies tend to rate **through the cycle**: defaults may become more likely during a downturn, but ratings often do not move immediately, so bonds with the same rating can show higher historical default frequencies in recessions than in expansions. This stability is valuable to issuers and traders, for whom a rating change (especially a downgrade) is a major event.

Moody's	S&P	Category
Aaa	AAA	Investment grade
Aa	AA	Investment grade
A	A	Investment grade
Baa	BBB	Investment grade
Ba	BB	Speculative, high yield, or junk
B	B	Speculative, high yield, or junk
Caa	CCC	Speculative, high yield, or junk
D	D	In default

Many bank borrowers are unrated externally, so banks also maintain internal rating systems and estimate both probability of default and recovery rate themselves.

If an obligor defaults, the lender usually recovers part of the exposure through a debt restructuring, through new securities, or through liquidation. The **recovery rate** RR is the proportion of the contract's ideal market value that remains after default. The **loss given default** is

$$\mathrm{LGD} = 1 - \mathrm{RR}\,.$$

Historical recovery rates vary substantially by seniority. A Moody's study of US defaults found the following average recovery rates by class:

Class	Average recovery rate
Senior secured	57.4%
Senior unsecured	44.9%
Senior subordinated	39.1%
Subordinated	32.0%
Junior subordinated	28.9%

Moody's also found that recovery rates and default rates are negatively related. One empirical fit (based on historical data from 1983–2004) is

$$\text{Average Recovery Rate} = 0.52 - 6.9 \cdot \text{Average Default Rate}.$$

Thus recoveries tend to be smaller precisely when defaults are more common. This is bad for banks: when firms are doing badly, they are more likely to default, and the collateral posted by those firms is often worth less at the same time.

Merton's Structural Model

Merton's model applies the Black–Scholes argument to credit risk. A firm has a single liability with face value L due at time T. The firm pays no dividends, does not issue new debt after time 0, and has asset value A_t satisfying the geometric Brownian motion

$$\mathrm{d}A_t = \mu A_t \,\mathrm{d}t + \sigma A_t \,\mathrm{d}Z_t,$$

so

$$A_t = A_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma Z_t\right).$$

Default can occur only at time T. When $A_T \geq L$, the debt is repaid: bondholders receive L and shareholders receive $A_T - L$. When $A_T < L$, the firm is liquidated instead, leaving bondholders with A_T and shareholders with nothing. Therefore

$$E_T = \max(A_T - L, 0).$$

Thus equity is a call option on the firm's assets with strike L and maturity T, while the debt payoff is

$$D_T = A_T - \max(A_T - L, 0) = \min(A_T, L) = L - \max(L - A_T, 0),$$

so bondholders own the firm but have sold a call option to shareholders; equivalently, risky debt is a default-free bond minus a put option on the firm's assets. The threshold mechanism in Figure 11 makes the conclusion immediate: default occurs at maturity precisely when $A_T < L$.

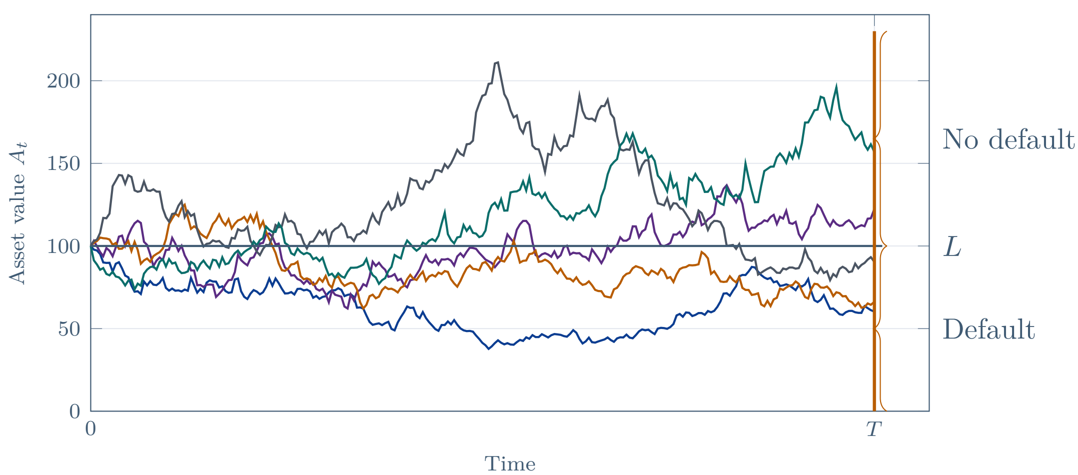


Figure 11

Proposition 12.1 (Pricing in Merton's model) Suppose continuous trading is possible, the risk-free rate is r, the asset value A_t is independent of the face value L and of how the firm is financed, and A_t can be traded. A security with an integrable payoff $h(A_T)$ at time T has time-t price

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}h(A_T) \mid \mathcal{F}_t\right],$$

where under $\mathbb{Q}$ the drift μ is replaced by r:

$$A_T = A_t \exp\left(\left(r - \frac{\sigma^2}{2}\right)(T-t) + \sigma(W_T - W_t)\right).$$

The equity value is the Black–Scholes call value

$$E_t = A_t\Phi(d_1) - e^{-r(T-t)}L\Phi(d_2),$$

where

$$d_1 = \frac{\log(A_t/L) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \qquad \text{and} \qquad d_2 = d_1 - \sigma\sqrt{T-t}.$$

This risk-neutral pricing argument is mathematically convenient, but it depends on the strong assumption that the firm's assets can be traded in a way that supports replication. The key structural ideas are that the firm's assets and liabilities are modelled explicitly, and default occurs when assets fall below a liability threshold.

Example 12.2 (Credit spread in Merton's model) Let $P_1(T)$ be the price of a zero-coupon bond with face value 1 issued by the firm. If $c(0,T)$ is the credit spread over $[0,T]$, then

$$P_1(T) = e^{-(r+c(0,T))T} \qquad \text{and} \qquad c(0,T) = -\frac{\log P_1(T)}{T} - r.$$

For a bond with face value L, the time-0 debt value is

$$D_0 = A_0 - E_0.$$

Using the put representation of risky debt,

$$D_0 = Le^{-rT}\Phi(d_2) + A_0\Phi(-d_1),$$

where

$$d_1 = \frac{\log(A_0/L) + (r+\sigma^2/2)T}{\sigma\sqrt{T}} \qquad \text{and} \qquad d_2 = d_1 - \sigma\sqrt{T}.$$

Therefore
$$P_1(T) = \frac{D_0}{L}$$
and the spread $c(0,T)$ increases with volatility σ. Its dependence on maturity T need not be monotone, because the spread annualizes a maturity-dependent debt value.

Proposition 12.3 Under Merton's model, the real-world default probability is

$$\mathbb{P}(A_T \leq L) = \Phi\left(\frac{\log(L/A_0) - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}\right),$$

while the risk-neutral default probability replaces μ by r:

$$\mathbb{Q}(A_T \leq L) = \Phi\left(\frac{\log(L/A_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right).$$

Default probability therefore increases with leverage and decreases with asset value and drift. For usual parameter values, it also increases with volatility when the firm initially lies above the default boundary, although that monotonicity is not universal.

Example 12.4 A firm's asset value is modelled using Merton's model. Suppose $V_0 =$ 1,000,000, the asset return rate is $\mu_V = 0.1$, and the firm has issued a three-year zero-coupon bond with face value $B = 500{,}000$.

1. Find the volatility σ_V of the firm's asset return if its probability of default is 0.06.
2. Determine the expected fraction of money recovered by the lender when the firm defaults, conditional on default.

Use the fact that if $X \sim \mathcal{N}(a,b)$, then

$$\mathbb{E}\left[e^X \mathbb{1}_{\{e^X < c\}}\right] = \exp(a + b/2)\,\Phi\left(\frac{\log(c) - a - b}{\sqrt{b}}\right).$$

Solution. Let A_3 denote the asset value at time 3. Under Merton's model,

$$A_3 = V_0 \exp\left(\left(\mu_V - \frac{\sigma_V^2}{2}\right) 3 + \sigma_V \sqrt{3}\, Z\right), \quad Z \sim \mathcal{N}(0,1).$$

For (1), default occurs when $A_3 < B$, so

$$0.06 = \mathbb{P}(A_3 < B) = \Phi\left(\frac{\log(B/V_0) - \left(\mu_V - \sigma_V^2/2\right) 3}{\sigma_V\sqrt{3}}\right).$$

Let $z_{0.06} = \Phi^{-1}(0.06) \approx -1.5548$. Then σ_V solves

$$\frac{\log(500000/1000000) - \left(0.1 - \sigma_V^2/2\right) 3}{\sigma_V\sqrt{3}} = z_{0.06},$$

that is,

$$1.5\sigma_V^2 + 2.693\sigma_V - 0.9931 = 0.$$

Taking the positive root gives

$$\sigma_V \approx 0.3139.$$

For (2), the lender receives A_3 in default, so the conditional recovery fraction is

$$\mathbb{E}\left[\frac{A_3}{B}\middle| A_3 < B\right] = \frac{1}{B\mathbb{P}(A_3 < B)}\mathbb{E}\left[A_3 \mathbb{1}_{\{A_3<B\}}\right].$$

Set $X = \log A_3$. Then $X \sim \mathcal{N}(a, b)$ with

$$a = \log V_0 + \left(\mu_V - \frac{\sigma_V^2}{2}\right) 3 \qquad \text{and} \qquad b = 3\sigma_V^2.$$

Using the given identity with $c = B$, we obtain

$$\mathbb{E}\left[A_3 \mathbb{1}_{\{A_3<B\}}\right] = \exp(a + b/2)\Phi\left(\frac{\log B - a - b}{\sqrt{b}}\right).$$

Therefore

$$\mathbb{E}\left[\frac{A_3}{B}\middle| A_3 < B\right] = \frac{\exp(a + b/2)}{500000 \cdot 0.06}\Phi\left(\frac{\log(500000) - a - b}{\sqrt{b}}\right).$$

Substituting $\sigma_V \approx 0.3139$ gives

$$\mathbb{E}\left[\frac{A_3}{B}\middle| A_3 < B\right] \approx 0.8068.$$

So the expected recovery fraction conditional on default is about 80.7%. □

Example 12.5 A firm's asset value is modelled using Merton's model. Suppose $V_0 = 750{,}000$, $\mu_V = 0.08$, $\sigma_V = 0.25$, and the firm issues a two-year bond with face value 500,000 paying an annual coupon of 6%. The risk-free rate is 4% annually.

1. What is the probability of default at time 1?
2. Given that the firm's value at time 1 is 775,000, what is the probability of default at time 2?
3. What is the value of the bond at time 0?
4. What is the firm's credit spread?

Solution. We use the natural extension of Merton's model to discrete payments: default can occur at coupon dates, and at time 1 shareholders continue only if the continuation value of equity is at least the coupon due. The coupon is

$$C = 0.06 \times 500000 = 30000,$$

so after the time-1 coupon is paid, the remaining one-year liability is

$$K_2 = 500000 + 30000 = 530000.$$

At time 1, if the firm continues, equity is a one-year call option on the asset value with strike 530000. Hence the continuation equity is

$$E_1^{\text{cont}}(a) = a\Phi(\widetilde{d}_1(a)) - 530000e^{-0.04}\Phi(\widetilde{d}_2(a)),$$

where

$$\widetilde{d}_1(a) = \frac{\log(a/530000) + (0.04 + 0.25^2/2)}{0.25} \qquad \text{and} \qquad \widetilde{d}_2(a) = \widetilde{d}_1(a) - 0.25.$$

Default at time 1 occurs when $E_1^{\text{cont}}(A_1) < 30000$. Solving

$$E_1^{\text{cont}}(a^*) = 30000$$

gives the default threshold

$$a^* \approx 466246.$$

For (1), under the physical measure,

$$\log A_1 \sim \mathcal{N}\left(\log 750000 + \left(0.08 - \frac{0.25^2}{2}\right),\ 0.25^2\right),$$

so

$$\mathbb{P}(\tau = 1) = \mathbb{P}(A_1 < a^*) = \Phi\left(\frac{\log(a^*/750000) - \left(0.08 - 0.25^2/2\right)}{0.25}\right) \approx 0.0180.$$

For (2), since $775000 > a^*$, the firm survives to time 1. Default at time 2 then occurs iff $A_2 < 530000$. Conditional on $A_1 = 775000$,

$$\log A_2 \mid A_1 = 775000 \sim \mathcal{N}\left(\log 775000 + \left(0.08 - \frac{0.25^2}{2}\right),\ 0.25^2\right),$$

so

$$\mathbb{P}(\tau = 2 \mid A_1 = 775000) = \Phi\left(\frac{\log(530000/775000) - \left(0.08 - 0.25^2/2\right)}{0.25}\right) \approx 0.0432.$$

For (3), under the risk-neutral measure we replace μ_V by $r = 0.04$. Let

$$D_1(a) = 530000e^{-0.04}\Phi(\widetilde{d}_2(a)) + a\Phi(-\widetilde{d}_1(a))$$

be the time-1 value of the remaining risky zero-coupon debt when the asset value is a. Therefore

$$D_0 = e^{-0.04}\mathbb{E}^{\mathbb{Q}}\left[A_1 \mathbb{1}_{\{A_1 < a^*\}} + \left(30000 + D_1(A_1)\right)\mathbb{1}_{\{A_1 \geq a^*\}}\right].$$

Evaluating this expectation numerically gives

$$D_0 \approx 506076.$$

For (4), if c is the constant credit spread over the flat risk-free curve, then the bond price also satisfies

$$506076 = 30000e^{-(0.04+c)} + 530000e^{-2(0.04+c)}.$$

Solving for c gives

$$c \approx 0.0121.$$

So the credit spread is about 1.21% per year, or about 121 basis points. □

Two common extensions of Merton's model are first-passage models and richer asset processes. In a first-passage model, default occurs the first time the asset value falls below a liability boundary rather than only at time T. Richer stochastic processes can allow jumps in asset prices, producing surprise defaults.

The **KMV model** builds on Merton's framework through the **expected default frequency** EDF, its key quantity and the one-year default probability. It is modelled as

$$\text{EDF} = f(\text{DtD}),$$

where f is estimated empirically and the **distance to default** is

$$\text{DtD} = \frac{\log(A_0/L) + (\mu - \sigma^2/2)T}{\sigma\sqrt{T}},$$

with L replaced by a firm-specific liability threshold.

Because the firm's asset value A_0 is not directly observable, it is inferred iteratively from equity value. KMV then replaces the normal tail expression in Merton's default probability with an empirically estimated decreasing function; the resulting model has been widely used in industry.

Credit Migration and Threshold Models

A **credit migration model** assumes that there is a finite number of credit states, such as AAA, AA, A, BBB, BB, B, CCC, and default. Each firm starts in one state and may move between states over the horizon. The model requires a transition matrix $\mathbf{P}$ whose entry p_{ij} gives the probability of moving from rating i to rating j over the horizon. Markov chain methods can then be used to compute default probabilities over longer horizons.

Structural default can be written in threshold form. From Merton's model, default occurs when $A_T < L$, equivalently when

$$Z < \frac{\log(L/A_0) - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}},$$

where Z is a standard normal random variable. If a default probability PD is estimated directly, the threshold can instead be calibrated by

$$Z < \Phi^{-1}(\text{PD}).$$

So far, we have analyzed the default probability of a single firm, but a portfolio depends on more than its individual probabilities: the *dependence* among defaults is vitally important. A **general**

threshold model provides a natural framework, allowing multiple credit states and linking firms through a copula.

Consider m obligors. Let $\boldsymbol{X} = (X_1, \ldots, X_m)$ be an m-dimensional vector of creditworthiness variables. For obligor i, let

$$d_{i,1} < d_{i,2} < \cdots < d_{i,n}$$

be thresholds, collected in an $m \times n$ matrix $\mathbf{D}$. Set $d_{i,0} = -\infty$ and $d_{i,n+1} = \infty$. The state S_i of firm i is defined by

$$S_i = j \iff d_{i,j} < X_i \leq d_{i,j+1}, \quad j = 0, \ldots, n.$$

The state $S_i = 0$ represents default. The default indicator is

$$Y_i = \mathbb{1}_{\{X_i \leq d_{i,1}\}},$$

and the marginal default probability is

$$\bar{p}_i = \mathbb{P}(X_i \leq d_{i,1}) = F_i(d_{i,1}) = \mathbb{P}(Y_i = 1).$$

The number of defaults and the portfolio loss are

$$M = \sum_{i=1}^{m} Y_i \qquad \text{and} \qquad L = \sum_{i=1}^{m} \delta_i e_i Y_i,$$

where e_i is the exposure for firm i and δ_i is the fraction lost in default. Risk measures such as VaR and CTE are then computed from the distributions of M and L.

Definition 12.6 For two default indicators Y_i and Y_j with marginal default probabilities $0 < \bar{p}_i, \bar{p}_j < 1$, the **default correlation** is

$$\rho(Y_i, Y_j) = \frac{\mathbb{E}[Y_i Y_j] - \bar{p}_i \bar{p}_j}{\sqrt{(\bar{p}_i - \bar{p}_i^2)(\bar{p}_j - \bar{p}_j^2)}}.$$

Example 12.7 Suppose there are two obligors, $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_2 = 1) = 0.1$, and $\delta_1 e_1 = \delta_2 e_2 = 10^6$. If Y_1 and Y_2 are independent, then

$$\mathbb{P}(L = 2 \cdot 10^6) = 0.01, \qquad \mathbb{P}(L = 10^6) = 2(0.1)(0.9) = 0.18, \qquad \text{and} \qquad \mathbb{P}(L = 0) = 0.81.$$

If instead

$$\mathbb{P}(Y_1 = 1, Y_2 = 1) = 0.09 \qquad \text{and} \qquad \mathbb{P}(Y_1 = 1, Y_2 = 0) = \mathbb{P}(Y_1 = 0, Y_2 = 1) = 0.01,$$

and

$$\mathbb{P}(Y_1 = 0, Y_2 = 0) = 0.89,$$

then the marginal default probabilities remain 0.1, but

$$\mathbb{P}(L = 2 \cdot 10^6) = 0.09.$$

Thus the same marginal default probabilities can imply very different portfolio tail risk once dependence changes.

Copulas are natural in threshold models because marginal default probabilities are often easier to estimate than the joint dependence structure. By Sklar's theorem (Theorem 8.4), the marginal cdfs and the copula can be chosen separately. In credit modelling, this means the default or migration thresholds are calibrated to match marginal probabilities, while the copula specifies the dependence among creditworthiness variables.

In **Li's Gaussian copula model**, X_i is the default time of firm i and the default threshold is $d_i = T$ for all i. It assumes that the default times are exponentially distributed with rate parameters λ_i:

$$X_i \sim \text{Exp}(\lambda_i) \qquad \text{and} \qquad F_i(t) = 1 - e^{-\lambda_i t}.$$

A Gaussian copula $C_{\boldsymbol{\Sigma}}$ is used for the joint distribution:

$$\mathbb{P}(X_1 \le t_1, \ldots, X_m \le t_m) = C_{\boldsymbol{\Sigma}}\left(1 - e^{-\lambda_1 t_1}, \ldots, 1 - e^{-\lambda_m t_m}\right).$$

If $\bar{p}_i = \mathbb{P}(X_i \le T)$, then

$$\bar{p}_i = 1 - e^{-\lambda_i T} \iff \lambda_i = -\frac{\log(1 - \bar{p}_i)}{T}.$$

For a more general threshold model with marginal cdfs $F_1, \ldots, F_m$ and copula C, default probabilities involving the indicators require only the copula and the marginal default probabilities. For example, if $m = 4$, then

$$\mathbb{P}(Y_1 = 1, Y_2 = 1) = \mathbb{P}(X_1 \le d_1, X_2 \le d_2, X_3 < \infty, X_4 < \infty) = C(\bar{p}_1, \bar{p}_2, 1, 1).$$

Exact enumeration becomes impossible for large portfolios: with $m = 1000$, the vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ has 2^{1000} possible default configurations.

Example 12.8 Consider Li's model for two firms with one-year default probabilities 0.1 and 0.15.

1. Determine λ_1 and λ_2 such that $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$.
2. If X_1 and X_2 are instead modelled as standard normal critical variables, determine the thresholds d_1 and d_2.
3. Suppose $\boldsymbol{X} = (X_1, X_2)$ is modelled using a Gaussian copula with correlation $\rho = 0.5$. Give expressions for $\mathbb{P}(Y_1 = 1, Y_2 = 1)$ and $\mathbb{P}(Y_1 = 0, Y_2 = 0)$ in terms of the bivariate normal CDF Φ_ρ.
4. Repeat the previous part using the Frank copula

$$C(u_1, u_2) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right),$$

   with $\theta = 2$, and compute the probabilities.
5. Let $L = 500{,}000Y_1 + 750{,}000Y_2$. Compute $q_{0.95}(L)$, the smallest q such that $\mathbb{P}(L > q) \leq 0.05$, and repeat for $q_{0.99}(L)$.

Solution. For (1), with a one-year horizon $T = 1$, Li's model gives

$$\bar{p}_i = 1 - e^{-\lambda_i} \Longrightarrow \lambda_i = -\log(1 - \bar{p}_i).$$

Therefore

$$\lambda_1 = -\log(0.9) \approx 0.1054 \qquad \text{and} \qquad \lambda_2 = -\log(0.85) \approx 0.1625.$$

For (2), if the critical variables are standard normal, the thresholds satisfy

$$\Phi(d_1) = 0.1 \qquad \text{and} \qquad \Phi(d_2) = 0.15,$$

so

$$d_1 = \Phi^{-1}(0.1) \approx -1.2816 \qquad \text{and} \qquad d_2 = \Phi^{-1}(0.15) \approx -1.0364.$$

For (3), under the Gaussian copula with correlation $\rho = 0.5$,

$$\mathbb{P}(Y_1 = 1, Y_2 = 1) = \Phi_{0.5}(d_1, d_2),$$

and by inclusion-exclusion,

$$\mathbb{P}(Y_1 = 0, Y_2 = 0) = 1 - 0.1 - 0.15 + \Phi_{0.5}(d_1, d_2).$$

Numerically,

$$\Phi_{0.5}(d_1, d_2) \approx 0.0428 \qquad \text{and} \qquad \mathbb{P}(Y_1 = 0, Y_2 = 0) \approx 0.7928.$$

For (4), with the Frank copula and $\theta = 2$,

$$\mathbb{P}(Y_1 = 1, Y_2 = 1) = C(0.1, 0.15) = -\frac{1}{2}\log\left(1 + \frac{(e^{-0.2} - 1)(e^{-0.3} - 1)}{e^{-2} - 1}\right) \approx 0.0279.$$

Again by inclusion-exclusion,

$$\mathbb{P}(Y_1 = 0, Y_2 = 0) = 1 - 0.1 - 0.15 + C(0.1, 0.15) \approx 0.7779.$$

For (5), the possible loss values are

$$0, \qquad 500000, \qquad 750000, \qquad \text{and} \qquad 1250000.$$

Under the Gaussian copula,

$$\begin{aligned}
\mathbb{P}(L = 1250000) &= \mathbb{P}(Y_1 = 1, Y_2 = 1) = \Phi_{0.5}(d_1, d_2) \approx 0.0428,\\
\mathbb{P}(L = 750000) &= \mathbb{P}(Y_1 = 0, Y_2 = 1) = \mathbb{P}(Y_2 = 1) - \mathbb{P}(Y_1 = 1, Y_2 = 1) = 0.15 - 0.0428 = 0.1072,\\
\mathbb{P}(L = 500000) &= \mathbb{P}(Y_1 = 1, Y_2 = 0) = \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_1 = 1, Y_2 = 1) = 0.1 - 0.0428 = 0.0572.
\end{aligned}$$

Thus

$$\mathbb{P}(L > 500000) = 0.1500 > 0.05 \qquad \text{and} \qquad \mathbb{P}(L > 750000) = 0.0428 < 0.05,$$

so

$$q_{0.95}(L) = 750000.$$

Also,

$$\mathbb{P}(L > 750000) = 0.0428 > 0.01 \qquad \text{and} \qquad \mathbb{P}(L > 1250000) = 0,$$

so

$$q_{0.99}(L) = 1250000.$$

Under the Frank copula, the joint default probability is smaller,

$$\mathbb{P}(L = 1250000) = 0.0279,$$

but the same quantiles result:

$$q_{0.95}(L) = 750000 \qquad \text{and} \qquad q_{0.99}(L) = 1250000.$$

□

Figure 12 shows how thresholds partition a continuous creditworthiness variable into default and rating states.

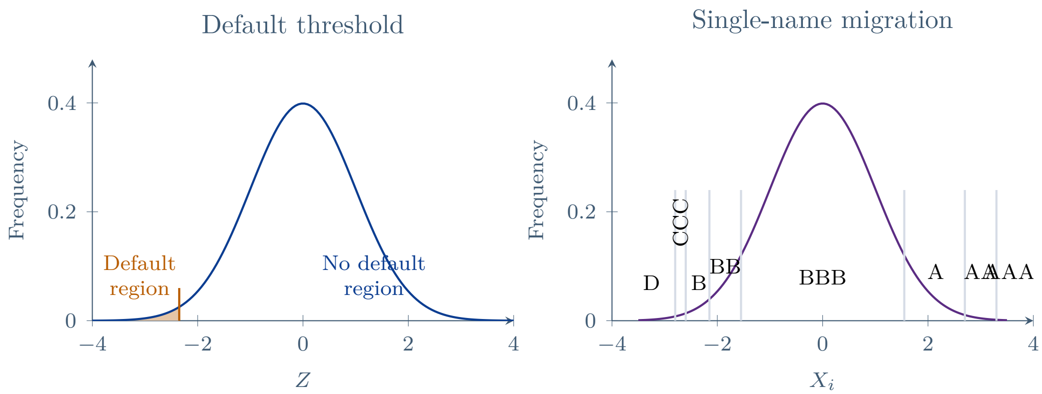


FIGURE 12

Vasicek's Single Factor Credit Model

For a name n, define a latent creditworthiness index

$$Y_n = \beta_n Z + \sqrt{1-\beta_n^2}\,\varepsilon_n,$$

where Z and $\varepsilon_1, \ldots, \varepsilon_N$ are iid standard normal random variables and $0 \le \beta_n < 1$. The factor Z is the **systematic component**, ε_n is the **idiosyncratic component**, and β_n is the **factor correlation** of name n. Default occurs when

$$Y_n \le H_n \qquad \text{and} \qquad H_n = \Phi^{-1}(\mathrm{PD}_n).$$

This model is based on Merton's structural threshold idea. It is called the **Vasicek model** (for credit risk, not interest rates) and the **single factor Gaussian copula model**, and it forms the basis for the Bank for International Settlements credit risk capital calculations. The construction makes Y_n standard normal:

$$\mathrm{Var}(Y_n) = \beta_n^2 + (1-\beta_n^2) = 1.$$

The default indicator

$$\mathbb{1}_{\{Y_n \le H_n\}}$$

is Bernoulli with mean PD_n and variance $\mathrm{PD}_n(1-\mathrm{PD}_n)$.

The analogy with CAPM is direct. CAPM writes excess return as a systematic component plus idiosyncratic noise:

$$R_i = r_f + \beta_i(R_M - r_f) + \varepsilon_i.$$

Taking expectations gives the security market line

$$\mathbb{E}[R_i] = r_f + \beta_i(\mathbb{E}[R_M] - r_f),$$

where

$$\beta_i = \frac{\mathrm{cov}(R_i, R_M)}{\sigma_M^2}.$$

The regression form is

$$R_i = \alpha + \beta R_M + \varepsilon.$$

Investors can diversify away idiosyncratic risk, so they demand compensation for systematic risk. In the credit model, a large diversified portfolio behaves similarly: idiosyncratic credit risk diversifies away, while systematic credit risk remains, as summarized in Figure 13.

The correlation between the creditworthiness indices of names n and m is $\beta_n\beta_m$. The correlation between their default indicators is

$$\rho_{nm} = \frac{\mathbb{E}[\mathbb{1}_{\{Y_n \le H_n\}}\mathbb{1}_{\{Y_m \le H_m\}}] - \mathrm{PD}_n\,\mathrm{PD}_m}{\sqrt{\mathrm{PD}_n(1-\mathrm{PD}_n)\,\mathrm{PD}_m(1-\mathrm{PD}_m)}} = \frac{\Phi_2(H_n, H_m; \beta_n\beta_m) - \mathrm{PD}_n\,\mathrm{PD}_m}{\sqrt{\mathrm{PD}_n(1-\mathrm{PD}_n)\,\mathrm{PD}_m(1-\mathrm{PD}_m)}},$$

where $\Phi_2(x, y; \rho)$ is the usual cumulative bivariate normal distribution function with correlation ρ:

$$\Phi_2(x, y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{x}\int_{-\infty}^{y} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right) \mathrm{d}v\,\mathrm{d}u.$$

If exposures at default are EAD_n and recovery rates are RR_n, the total portfolio loss is

$$L_T = \sum_{n=1}^{N} \mathrm{EAD}_n(1-\mathrm{RR}_n)\mathbb{1}_{\{Y_n \le H_n\}}.$$

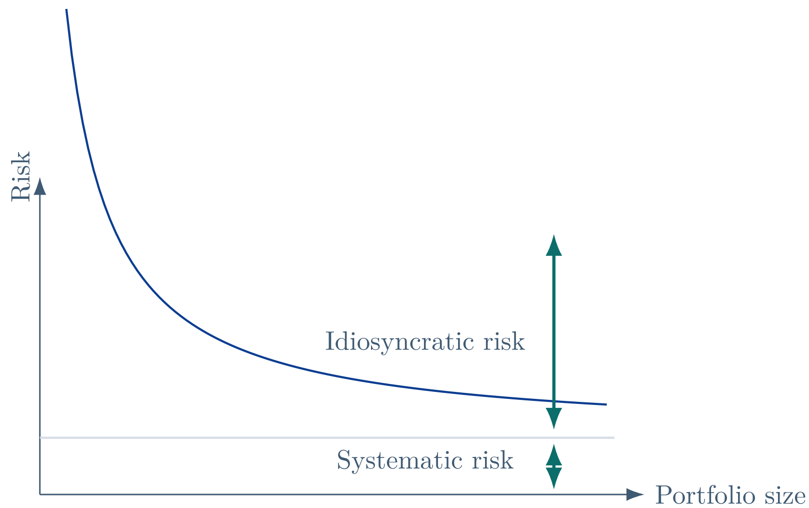


FIGURE 13

Proposition 12.9 For a sequence of increasingly granular portfolios, write

$$c_n = \mathrm{EAD}_n(1 - \mathrm{RR}_n) \qquad \text{and} \qquad C_N = \sum_{n=1}^{N} c_n,$$

where the coefficients are understood as a triangular array indexed by N, although that index is suppressed. Assume

$$\sum_{n=1}^{N} \left(\frac{c_n}{C_N}\right)^2 \to 0 \quad \text{as } N \to \infty,$$

so that no finite collection of names retains a material share of total exposure. Then idiosyncratic loss is negligible relative to total exposure:

$$\frac{L_T - \mathbb{E}[L_T \mid Z]}{C_N} \longrightarrow 0 \quad \text{in } L^2.$$

We call the conditional expectation the **systematic loss**:

$$L_S = \mathbb{E}[L_T \mid Z] = \sum_{n=1}^{N} \mathrm{EAD}_n(1 - \mathrm{RR}_n)\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \beta_n Z}{\sqrt{1 - \beta_n^2}}\right).$$

The conditional expectation for each name is

$$\mathbb{E}[\mathbb{1}_{\{Y_n \le H_n\}} \mid Z] = \mathbb{P}\left(\varepsilon_n \le \frac{H_n - \beta_n Z}{\sqrt{1-\beta_n^2}} \middle| Z\right) = \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \beta_n Z}{\sqrt{1-\beta_n^2}}\right).$$

Proof. Conditional on the systematic factor Z, the default indicators

$$\mathbb{1}_{\{Y_n \le H_n\}}, \quad n = 1, \dots, N,$$

are independent, because each one depends on Z and on its own idiosyncratic shock ε_n, and the shocks ε_n are independent across names. Therefore

$$\mathbb{E}[L_T \mid Z] = \sum_{n=1}^{N} \mathrm{EAD}_n(1 - \mathrm{RR}_n)\mathbb{E}[\mathbb{1}_{\{Y_n \le H_n\}} \mid Z].$$

Since

$$Y_n = \beta_n Z + \sqrt{1-\beta_n^2}\,\varepsilon_n \qquad \text{and} \qquad H_n = \Phi^{-1}(\mathrm{PD}_n),$$

we obtain

$$\mathbb{E}[\mathbb{1}_{\{Y_n \le H_n\}} \mid Z] = \mathbb{P}(Y_n \le H_n \mid Z) = \mathbb{P}\left(\varepsilon_n \le \frac{H_n - \beta_n Z}{\sqrt{1-\beta_n^2}} \middle| Z\right) = \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \beta_n Z}{\sqrt{1-\beta_n^2}}\right),$$

which gives the stated expression for $L_S = \mathbb{E}[L_T \mid Z]$.

Now work along the given sequence of portfolios and write

$$p_n(Z) = \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \beta_n Z}{\sqrt{1-\beta_n^2}}\right).$$

Conditional on Z, the variables $\mathbb{1}_{\{Y_n \le H_n\}}$ are independent Bernoulli random variables with parameter $p_n(Z)$, so

$$\mathrm{Var}\left(\frac{L_T}{C_N} \middle| Z\right) = \sum_{n=1}^{N} \left(\frac{c_n}{C_N}\right)^2 \mathrm{Var}(\mathbb{1}_{\{Y_n \le H_n\}} \mid Z) = \sum_{n=1}^{N} \left(\frac{c_n}{C_N}\right)^2 p_n(Z)(1 - p_n(Z)) \le \frac{1}{4} \sum_{n=1}^{N} \left(\frac{c_n}{C_N}\right)^2.$$

The deterministic upper bound tends to zero. Taking expectations therefore gives

$$\mathbb{E}\left[\left(\frac{L_T - \mathbb{E}[L_T \mid Z]}{C_N}\right)^2\right] \to 0,$$

which is the claimed L^2 convergence. This is the precise sense in which idiosyncratic fluctuations diversify away. □

Because the factor loadings are nonnegative, the systematic loss is a decreasing function of the global factor Z: smaller values of Z produce larger losses. Therefore its quantiles can be computed explicitly:

$$L_{S,\alpha} = \sum_{n=1}^{N} \mathrm{EAD}_n(1 - \mathrm{RR}_n)\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \beta_n z_{1-\alpha}}{\sqrt{1-\beta_n^2}}\right).$$

Thus **portfolio VaR** is the sum of the individual instrument contributions in the asymptotic single factor model.

Gordy (2003) proved that in the asymptotic single factor model, the capital charge of a portfolio is the sum of the capital charges of the individual loans. The capital charge of a loan does not depend on the portfolio in which it is held. This is the **portfolio invariance** property, which is convenient for regulation, although not always realistic. It makes the charge depend on the loan's own characteristics rather than on the actual composition or concentration of the portfolio in which it is held, so portfolio-specific diversification effects are absent from the charge.

Basel II and Extensions

The single factor model underlies the credit risk capital calculations of the Bank for International Settlements (BIS), the organization that promotes cooperation among central banks and sets global banking standards. Its **Basel II** framework, implemented in 2007, determines minimum credit risk capital from one-year VaR at the 99.9% confidence level; the core inputs are probability of default PD, exposure at default EAD, loss given default LGD, and maturity M.

The **worst-case default rate** (**WCDR**) at confidence level 99.9% is

$$\mathrm{WCDR} = \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}) - \beta z_{0.001}}{\sqrt{1-\beta^2}}\right),$$

so the instrument **capital charge** is

$$\mathrm{EAD} \cdot \mathrm{LGD} \cdot (\mathrm{WCDR} - \mathrm{PD}) \cdot \mathrm{MA}\,.$$

Basel often expresses capital requirements in terms of **risk-weighted assets (RWA)** rather than directly in dollars of capital. RWA are not the actual book value of the assets; instead, they are a regulatory rescaling of exposure that puts safer and riskier positions on a common basis. The idea is that a bank must hold capital equal to a fixed fraction of its RWA, so a larger capital charge corresponds to a larger amount of risk-weighted assets. Under Basel II, the minimum total capital requirement is 8% of RWA, so the corresponding RWA for the exposure are

$$\mathrm{RWA} = \frac{\text{capital charge}}{0.08} = 12.5 \times \text{capital charge}.$$

Thus, with the capital charge formula above,

$$\mathrm{RWA} = 12.5 \cdot \mathrm{EAD} \cdot \mathrm{LGD} \cdot (\mathrm{WCDR} - \mathrm{PD}) \cdot \mathrm{MA}\,.$$

Equivalently, if we define

$$K = \mathrm{LGD} \cdot (\mathrm{WCDR} - \mathrm{PD}) \cdot \mathrm{MA},$$

then K is the capital requirement per dollar of exposure and

$$\mathrm{RWA} = 12.5\, K\ \mathrm{EAD}\,.$$

For the admissible regulatory range of PD and M, Basel specifies the factor correlation through an empirical function of PD:

$$\beta^2 = 0.12 \frac{1 - e^{-50\,\mathrm{PD}}}{1 - e^{-50}} + 0.24 \left(1 - \frac{1 - e^{-50\,\mathrm{PD}}}{1 - e^{-50}}\right),$$

and the **maturity adjustment** is

$$\mathrm{MA} = \frac{1 + (M - 2.5)b}{1 - 1.5b} \qquad \text{and} \qquad b = \big(0.11852 - 0.05478 \log(\mathrm{PD})\big)^2.$$

Here M is the actual maturity of the exposure, and the regulatory range keeps $1 - 1.5b$ positive. As the company's default probability increases, β decreases. The interpretation is that less creditworthy firms tend to be driven more by idiosyncratic risk and less by the market factor.

Under the **foundation internal ratings-based (F-IRB) approach**, the bank supplies PD while Basel specifies LGD, EAD, and M; under the **advanced internal ratings-based (A-IRB) approach**, the bank supplies all four. Other exposure classes modify the corporate, sovereign, and bank formulas (retail exposures, for example, have no maturity adjustment), while the trading book is more complicated because its exposures are themselves random.

Credit risk and liquidity risk are closely connected. Firms with lower credit risk are viewed as more homogeneous and can usually be bought and sold more easily and in larger quantities without substantial price impact. This liquidity component is reflected in bond spreads: the difference between BBB and AA bond spreads is too large to be explained entirely by default risk.

Example 12.10 Consider two banks with loan portfolios of equal size and quality. Bank A lends only to Texas oil companies, while Bank B lends nationwide to businesses in many sectors. A capital rule based on one factor and portfolio invariance does not reflect the diversification benefit in Bank B's portfolio.

A multifactor model generalizes the single factor credit model by writing

$$Y_n = \sum_{k=1}^{K} \beta_{nk} Z_k + \sqrt{1 - \sum_{k=1}^{K} \beta_{nk}^2}\, \varepsilon_n,$$

where $Z_1, \ldots, Z_K, \varepsilon_1, \ldots, \varepsilon_N$ are iid standard normal random variables. For every name, the loadings must satisfy

$$\sum_{k=1}^{K} \beta_{nk}^2 < 1,$$

so the idiosyncratic scale is real and positive. The threshold rule is unchanged:

$$\mathbb{1}_{\{Y_n \le H_n\}} \qquad \text{and} \qquad H_n = \Phi^{-1}(\mathrm{PD}_n).$$

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_K)$. In the large portfolio limit,

$$\mathbb{E}\left[L_T \mid \boldsymbol{Z}\right] = \sum_{n=1}^{N} \mathrm{EAD}_n (1 - \mathrm{RR}_n) \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_n) - \sum_{k=1}^{K} \beta_{nk} Z_k}{\sqrt{1 - \sum_{k=1}^{K} \beta_{nk}^2}} \right).$$

Unlike the case with one factor, VaR is generally not available explicitly. If all names have the same factor-loading vector, however, only that common linear combination of the factors matters and the model reduces effectively to one factor.

The Gaussian single factor and multifactor models use Gaussian copulas: the creditworthiness indicators have a multivariate Gaussian distribution. More general threshold models allow arbitrary marginal distributions for the creditworthiness indicators and use a copula to determine their joint distribution. The thresholds are calibrated to match default and migration probabilities, while the copula describes codependence. Figure 14 contrasts Gaussian tail independence with the joint lower-tail clustering available under a Student t copula.

The QRM textbook provides a simulation with 10,000 names that illustrates the impact of the copula choice. Each portfolio has a common default probability for all names and a common

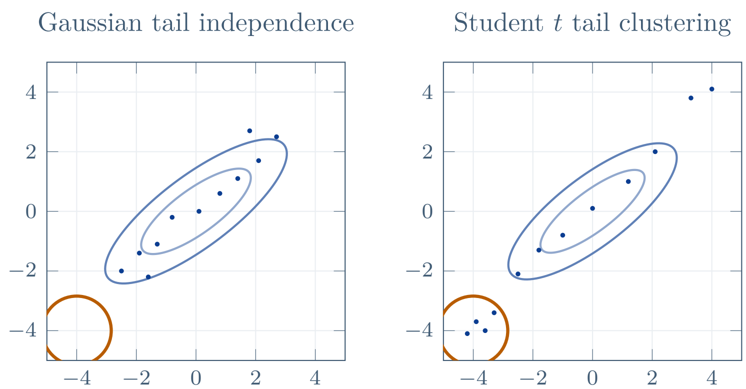


FIGURE 14

correlation between creditworthiness indicators:

Portfolio	PD	Creditworthiness indicator correlation
A	0.06%	2.58%
B	0.50%	3.80%
C	7.50%	9.21%

The following table gives the 99% VaR of M, the number of defaults. The case $\nu = \infty$ is the Gaussian copula:

Portfolio	$\nu = \infty$	$\nu = 50$	$\nu = 10$
A	21	49	118
B	157	261	589
C	2206	2400	3067

For Portfolio A, the 99th percentile of M under the Student t copula with $\nu = 10$ is more than five times the corresponding Gaussian value. Dependence choice therefore has a major effect on credit tail risk.

Proposition 12.11 Consider the one factor threshold model

$$X_i = \sqrt{\rho}\, F + \sqrt{1-\rho}\, \varepsilon_i, \quad i = 1, \dots, m,$$

where $F, \varepsilon_1, \dots, \varepsilon_m$ are iid standard normal. If $\bar{p}_i$ is the default probability of firm i, then for large m,

$$q_\alpha(L) \approx \sum_{i=1}^{m} \delta_i e_i \Phi\left(\frac{\Phi^{-1}(\bar{p}_i) + \sqrt{\rho}\,\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right),$$

where $q_\alpha(L)$ is the 100α percentile of

$$L = \sum_{i=1}^{m} \delta_i e_i Y_i,$$

provided the portfolio is sufficiently granular and $0 \le \rho < 1$. For $\alpha = 0.999$, each term in the sum contributes to the asymptotic loss quantile associated with firm i. Regulatory capital may then subtract expected loss and apply maturity or other adjustment factors.

Example 12.12 An asymptotic loss-quantile calculation for a portfolio with credit loss

$$L = \sum_{i=1}^{m} \delta_i e_i Y_i$$

uses

$$q_{0.999}(L) \approx \sum_{i=1}^{m} \delta_i e_i \Phi\left(\frac{\Phi^{-1}(\bar{p}_i) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}} \right),$$

where ρ is the correlation between X_i and X_j for all $i \neq j$. A portfolio contains investments in 100 firms. For 50% of the firms, the exposure is 5,000,000; for 30% it is 7,500,000; and for 20% it is 15,000,000. Suppose the loss given default is 100%, so $\delta_i = 1$, and the probability of default is 0.05 for all firms. Compute $q_{0.999}(L)$ for $\rho = 0.2$ and for $\rho = 0.7$.

Solution. The total exposure is

$$\sum_{i=1}^{m} \delta_i e_i = 50(5 \cdot 10^6) + 30(7.5 \cdot 10^6) + 20(15 \cdot 10^6) = 775 \cdot 10^6.$$

Since all firms have the same default probability $\bar{p}_i = 0.05$, the quantile formula simplifies to

$$q_{0.999}(L) \approx 775 \cdot 10^6 \Phi\left(\frac{\Phi^{-1}(0.05) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}} \right).$$

Using

$$\Phi^{-1}(0.05) \approx -1.6449 \qquad \text{and} \qquad \Phi^{-1}(0.999) \approx 3.0902,$$

we obtain the following.

If $\rho = 0.2$, then

$$\Phi\left(\frac{-1.6449 + \sqrt{0.2}\,(3.0902)}{\sqrt{0.8}}\right) \approx 0.3844,$$

so

$$q_{0.999}(L) \approx 775 \cdot 10^6 \times 0.3844 \approx 2.979 \times 10^8.$$

Thus

$$q_{0.999}(L) \approx 297.9 \text{ million}.$$

If $\rho = 0.7$, then

$$\Phi\left(\frac{-1.6449 + \sqrt{0.7}\,(3.0902)}{\sqrt{0.3}}\right) \approx 0.9570,$$

so

$$q_{0.999}(L) \approx 775 \cdot 10^6 \times 0.9570 \approx 7.417 \times 10^8.$$

Thus

$$q_{0.999}(L) \approx 741.7 \text{ million}.$$

The much larger value for $\rho = 0.7$ reflects the stronger default clustering when dependence is high. $\square$

Example 12.13 Consider again the asymptotic approximation

$$q_{0.999}(L) \approx \sum_{i=1}^{m} \delta_i e_i \Phi\left(\frac{\Phi^{-1}(\bar{p}_i) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right),$$

which comes from the model

$$X_i = \sqrt{\rho}\,F + \sqrt{1-\rho}\,\varepsilon_i,$$

where $F, \varepsilon_1, \ldots, \varepsilon_m$ are iid standard normal. Suppose $m = 100$, $\delta_i e_i = 10^6$, $\bar{p}_i = 0.08$ for all i, and $\rho = 0.5$. Determine $q_{0.999}(L)$.

Solution. Here every term in the sum is identical, so

$$q_{0.999}(L) \approx 100 \cdot 10^6 \Phi\left(\frac{\Phi^{-1}(0.08) + \sqrt{0.5}\,\Phi^{-1}(0.999)}{\sqrt{0.5}}\right).$$

Using

$$\Phi^{-1}(0.08) \approx -1.4051 \qquad \text{and} \qquad \Phi^{-1}(0.999) \approx 3.0902,$$

we get

$$\Phi\left(\frac{-1.4051+\sqrt{0.5}\,(3.0902)}{\sqrt{0.5}}\right)\approx 0.8650.$$

Therefore

$$q_{0.999}(L)\approx 100\cdot 10^6\times 0.8650 = 8.650\times 10^7.$$

So the portfolio 99.9% loss quantile is approximately

$$q_{0.999}(L)\approx 86.5 \text{ million}.$$

□

Appendix: Lecture and Tutorial Index

1. Lecture 1 — Tuesday, September 15, 2015
2. Lecture 2 — Tuesday, September 22, 2015
3. Lecture 3 — Tuesday, September 29, 2015
4. Lecture 4 — Tuesday, October 06, 2015
5. Lecture 5 — Tuesday, October 13, 2015
6. Lecture 6 — Tuesday, October 20, 2015
7. Lecture 7 — Tuesday, October 27, 2015
8. Lecture 8 — Tuesday, November 03, 2015
9. Lecture 9 — Tuesday, November 10, 2015
10. Lecture 10 — Tuesday, November 17, 2015
11. Lecture 11 — Tuesday, November 24, 2015
12. Lecture 12 — Thursday, December 03, 2015