CREDIT SUISSE | FIRST BOSTON

Credit Portfolio Modeling Handbook

QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUAN STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUAN QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUAN QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY • QUANTITATIVE CREDIT STRATEGY

MAIN AUTHOR

Richard Martin, Director
+44 20 7888 4736
richard.martin@csfb.com

FIXED INCOME PORTFOLIO STRATEGY

NORTH AMERICA	Jonathan Song, Director	Fer Koch, Vice President	Jason Rayman, Vice President, CFA
	(Global Group Head)	+1 212 325 2314	+1 212 325 4908
	+1 212 325 1768	fer.koch@csfb.com	jason.rayman@csfb.com
	jonathan.song@csfb.com		
	David Crick, Associate		
	+1 212 325 3070		
	david.crick@csfb.com		

EUROPE	Richard Martin, Director	Recai Güneşdoğdu, Vice President	Jon Zucker, Vice President
	+44 20 7888 4736	*(European Head)*	+44 20 7888 8109
	richard.martin@csfb.com	+44 20 7883 7978	jon.zucker@csfb.com
		recai.gunesdogdu@csfb.com	
	Derek Hynes, Associate	Donato Guarino, Associate	Andrea Cicione, Associate
	+44 20 7888 8107	+44 20 7883 6220	+44 20 7888 0757
	derek.hynes@csfb.com	donato.guarino@csfb.com	andrea.cicione@csfb.com

Table of Contents

Unless otherwise indicated the sources for all charts are Credit Suisse First Boston

THIS PAGE IS INTENTIONALLY LEFT BLANK

1

The past, present and future of credit risk

Introduction

Credit risk has been in the headlines for the last few years. The volatility in the credit market a few years ago and the turmoil surrounding the major defaults of Enron, WorldCom and Marconi have forced credit investors to pay close attention to risk and reward. But despite these, the appetite for credit risk has not diminished, and historically low interest rates have made investors turn to credit to provide extra return. Every year the credit derivatives and structured credit markets have grown at a great pace. According to the British Bankers' Association survey published at the end of September 2004, total outstanding notionals of credit derivatives have jumped during the first half of 2004 from $3.55 trillion at the end of 2003 to $5.02 trillion, and are predicted to surpass the $8.2 trillion mark by 2006.

The volatility in the investment grade market in 2002 also forced dealers and end-users to develop analytics to support their investment decisions, as it was found that the traditional rating and fundamental research proved insufficient to evaluate credit risk. This has resulted in an increased interest in credit risk modeling in the academic world, and only recently these techniques are being implemented in real-world applications.

We must begin with the most fundamental question of all:

- What is credit risk?

Further questions follow from this, which go to the heart of trading, structuring and risk management:

- How is credit risk priced?

- How should one invest in credit risk?

- How should one manage credit risk?

None of these questions is straightforward and all are interrelated. But we have to start somewhere, and the natural place to start is by defining a credit-risky asset.

A credit-risky asset is one that might not honor its financial obligations.

There are several approaches to modeling the default risk in an asset, commonly known as the reduced-form models, hybrid models, and structural (or Merton) models. These are described in more detail in the Appendix. In essence, reduced-form models can be thought of as generalizations of interest rate models and are direct models of the term structure of credit spread. By contrast, structural models seek to understand the dynamics of the financial structure of a firm, with default occurring when it is unable to meet its financial obligations. The term structure of credit spread then comes from the asset model. Structural models are essential to understanding the relation between equity and credit. Finally, hybrid models are essentially equity models with credit risk bolted onto the side. They are generally easier to calibrate to the equity and debt market than structural models and are a popular choice for pricing credit-risky equity products such as convertible bonds.

This primer is concerned with the modeling of portfolio credit risk. When an investor has a portfolio of credits to deal with, many of the following questions arise:

- What is the distribution of loss in my portfolio?

- What assumptions are being made when I derive this loss distribution?

- What does correlation mean?

- What does the market tell me about credit risk?

- How do I allocate risk amongst the constituents of my portfolio?

- Which constituents generate the most return on risk and which the least?

- How do changes in the portfolio composition manifest themselves at the portfolio level?

- How do I price credit default baskets and CDOs?

- How do I quantify and manage risk when I buy or sell tranches from baskets or CDOs?

We shall tackle all of these except those on baskets and CDOs, which will be covered in later publications. (However, the technical ideas used in the pricing of these products rest on what is contained herein.) This primer is aimed at a technical audience, but not all of the pieces will use mathematics: we think it is just as important to develop an intuitive understanding of all subjects discussed. Given that objective, those clients most interested in the series would include:

- Managers of bond portfolios;

- Credit vs. equity investors;

- Loan portfolio managers;

- CDO managers;

- CDO investors;

- Securities traders;

- Insurance companies; and

- Pension funds.

Charting the course

It is well known that portfolio modeling, whether in the context of derivatives or of credit risk, is difficult because it requires the aggregation of correlated risky positions with nonlinear payoffs. There are many issues to be covered, and a brief overview is as follows:

- First, we introduce factor models and large-portfolio approximations to the "loss distribution" or "distribution of P&L".

- Next, we examine how to deal with portfolios that do not fit into this idealized picture.

- Having obtained the distribution, which represents uncertainty, we wish to measure it, and this brings in the notion of risk measures.

- Next we address several difficult questions on the subject of risk contributions and portfolio optimization.

Before we begin we examine an important aspect of risk and portfolio management that is not treated very satisfactorily in literature on the subject.

Views of credit

There are different ways of looking at asset management, portfolio management or risk management. It is possible to define two "extremes", as follows.

Trading view

Equity and debt are tradable instruments and indeed they have both been traded in Europe for well over 200 years. If we take this notion to its extreme, and regard trading as a continuous activity, we can argue that all "credit risk" is just "market risk", i.e. trading or mark-to-market (MTM) losses, and look at it over a very short time horizon (a few days). We shall refer to this as "PV sensitivity", as the basic notion in market risk management is the sensitivity of the *value* of the portfolio (PV or P&L) over a short time horizon to small changes in the underlying *explanatory variables*. Examples of explanatory variables could be, in Fixed Income:

 EUR/USD exchange rate; 5Y JPY interest rate; 3M-into-10Y GBP/USD swaption vol; …

or in Equity Derivatives:

value of the S&P500; implied volatility of IBM shares; … .

First-order sensitivities to small changes are generally known as deltas, while convexities (second-order sensitivities) are known as gammas and cross-gammas. Collectively they are known as "greeks". The "greeks" approach has one great advantage, in that sensitivities can be added in a portfolio. So, referring to one particular explanatory variable, e.g. EUR/USD spot, one can take the deltas and gammas of the trades of one trader, add them to get the trader's greeks, aggregate over traders to get the desk's greeks, aggregate over desks to get the business line's greeks, and aggregate over business lines to get the greeks for the whole business. After this relatively simple process one arrives at the sensitivities of the global P&L to all the explanatory variables, and it is attractive to view the whole P&L as a complex derivative

that can be hedged using vanilla instruments once its deltas and gammas are known. Philosophically the end product of the analysis is a sort of "top-down" view. By this we mean that the microstructure of the portfolio can be ignored and the portfolio is consolidated into a single instrument whose value depends upon a small number of explanatory variables. This is attractive to heads of trading desks because a reasonably complete picture of the risk can be conveyed without "information overload" occurring.

This approach works very well when coupled with the Merton model for the valuation of equity and credit (implemented in our CUSP™ model), because we can analyze credit and equity in a consistent way and hence provide a unified methodology for the risk analysis of a variety of products: bonds, CDS, equity, equity derivatives and so on. This is particularly useful for the analysis of "capital structure arbitrage", which is the common name given to a variety of trades based on the supposed misalignment of volatility in the equity and credit markets, for example, long the stock and long the CDS, or long the equity put and short the CDS.

The *distribution* of P&L requires more information, namely how volatile the explanatory variables are, and how they are correlated. To express the joint distribution of *small* movements of all the explanatory variables it is common to use volatilities and correlations. What typically happens on trading desks is that these are taken into account "by judgment". For example, a swaps trader knows which parts of the yield curve are more volatile and which are less, and also that being long the 5Y and 7Y rate and short the 6Y rate is likely to result in a fair degree of cancellation.

The time when this approach can come badly unstuck is when large movements occur: sensitivity analysis does not tell us about changes that are larger than infinitesimal. It is possible for an instrument to be locally insensitive to some underlying variable (have zero delta and gamma[1]), but for it to change in value when the underlying moves a long way, as for example in Figure 1.1.

Figure 1.1. (a) local analysis shows no sensitivity to underlying variable; (b) full picture may have a nasty surprise

(a)

P&L

? ------ ------ ?

Change in underlying variable

(b)

P&L

Change in underlying variable

---
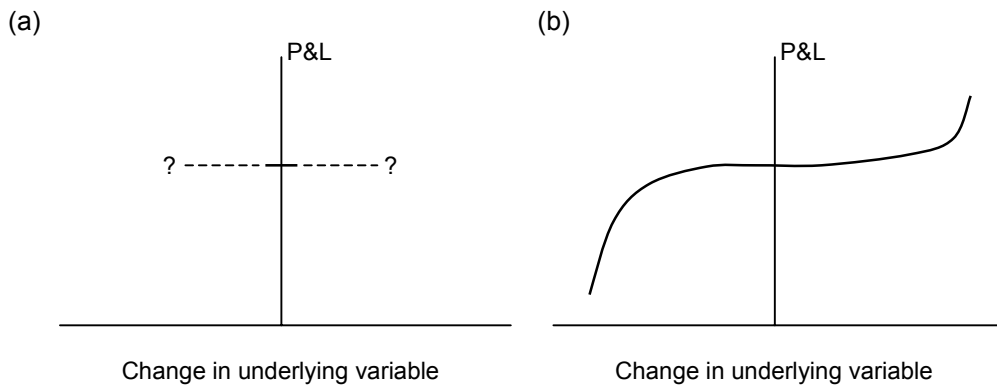
[1] and even for all derivatives to be zero at some point, i.e. the function is "locally flat". For example, $\exp(-1/x^2)$ in the vicinity of $x=0$.

By expressing the portfolio value through asset models such as Figure 1.1(b), and a joint distribution of all explanatory or underlying variables, one ends up with a complete picture of the distribution of P&L at some time horizon. This is then commonly measured using VaR (hence the term "10-day VaR" which is ubiquitous in market risk departments).

Buy-and-hold view

By contrast, the common meaning of "credit risk management", as it is known in banks, is different. There, the emphasis is on the *distribution* of default losses rather than MTM, and the time horizon is longer (often 1 year, or for counterparty risk the full life of the trade, which may be many years). Certain types of loans cannot be traded, and traditionally counterparty risk has been seen as a "sit on it and hope for the best" type of risk (though that is changing). For non-tradable assets MTM makes little sense and a buy-and-hold analysis is more appropriate, where one looks at real loss. With the emphasis being on a *loss distribution*, the idea is first to find a *risk measure* (we shall talk about this later) and then find the sensitivity of that risk measure to various parameters such as change in asset allocations. The whole subject has more of an "insurance-company" feel about it, whereas a securities house would not hold risk. For example, a securities house trades catastrophe bonds, but an insurance company buys and holds them. Rather than hedging risk, as happens in MTM, one seeks to diversify it instead, and achieve sufficient return to cover the risk. The PV sensitivity says nothing about diversification.

How does the modeling work? The usual way of setting about it is to model each asset individually. In commercial implementations such as KMV and CreditMetrics, the Merton model has found favor. So, in fact, the Merton model is useful in both the trading and the buy-and-hold setting: in the former, one trades the embedded put option[2] in credit-risky debt, and in the latter, one holds onto it until maturity. Then, once the individual assets have been modeled, it is necessary to find a correlation model that describes the co-movements. At this point it is currently fashionable to drag in copulas[3], though if one persists with the Merton model it is not strictly necessary to do so, as correlation can be introduced simply by prescribing that the joint distribution of asset returns be multivariate Normal. Generally the resulting models are analytically intractable and so Monte Carlo is used to obtain the loss distribution (as for example in CreditMetrics and KMV). The main objections to this are that it is slow and that it is hard to do further analysis such as sensitivity analysis and portfolio optimization. On the other hand, a large amount of progress has been made in portfolio analytics: when reviewing this, which we shall do shortly, we shall obtain a good insight into how portfolios "work".

---

[2] i.e. the embedded short put option on the firm's assets
[3] "Demystifying copulas", Chapter 4.

Comparison

In comparing the two approaches some questions become apparent, which we state and answer now.

- What does PV sensitivity say about proprietary trading and risk/reward?

In short, it doesn't, and this is the most important distinction between them. PV sensitivity is designed for people who are hedging their risk out. This is mainly suited to flow trading in which the idea is to take no risk and make a small margin on each trade. The buy-and-hold viewpoint explicitly acknowledges that risk is being taken and seeks adequate compensation for it. PV sensitivity says nothing about correlation or diversification.

- What is the right time horizon?

To answer this question we have to understand what we are trying to do. If we want to hedge our risk out then a very short time horizon is right. If we often hold trades for a while, perhaps a few months, then that should be the horizon. It is unlikely to be right to look over a period of years, however, even with so-called "economic capital" management where one looks at risks that are likely to sit around for that sort of period: as more and more risks are becoming tradable, the time horizon will naturally decrease from the maturity of the trade to the trading (or hedging) horizon.

For this reason we suggest that a period of about 3 months is sensible. Portfolio managers often tend to think on this kind of horizon, and so do people who put on proprietary trades such as basis or cap-arb trades: although some trades will be expected to move into profit more quickly, one often has to wait a little longer to derive maximum benefit from the position.

There is another reason why we choose 3 months, but we shall come to that later.

- What is understood by "credit risk"?

For buy-and-hold, credit risk is the inability of a market participant[4] to repay, and can be thought of as event risk: the credit riskiness of a bond is its *default probability*, often in the form of a credit rating. In MTM, credit risk is a form of *optionality* that can be traded, and the credit riskiness of a bond is its *volatility*, which can be extracted from the volatility of the issuer's assets and the leverage of the debt. CSFB's CUSP™ model extracts this information from the debt and equity markets.

- Why, in the buy-and-hold setting, cannot one seek to explain the risk in terms of explanatory variables, as in market risk?

In market risk, the number of explanatory variables is generally quite small. For example, if we run a USD swaps desk, the explanatory variables are the USD swap rates at all maturities. It doesn't matter if our desk has done ten swaps, or ten thousand: they can all be aggregated. But in credit risk there is a fundamental heterogeneity and the number of explanatory variables is too large: one is sensitive to the default of each

---

[4] e.g. a bond issuer (credit risk) or a market counterparty (counterparty risk)

issuer, and so one has as many explanatory variables as issuers. And working out the sensitivity of the portfolio value to each default event is not very helpful: if ABC defaults we lose $34M, if DEF defaults we lose $58M — this ignores correlation, and is therefore rather useless.

- It would be nice if the buy-and-hold setting had a top-down view. A detailed bottom-up model culminating in a single VaR number is not much help in portfolio management, so some sort of "the portfolio value depends on a small number of risk factors" analysis would be nice. Can anything be done?

Yes, provided that the portfolio is diversified: large portfolio approximations then come into play. In fact, if a portfolio is large enough, nearly all of the risk comes from the variation of the conditional expected loss, by which we mean the expected loss conditional on the set of risk-factors to which the portfolio is sensitive. This bears a good deal of resemblance to the top-down approach described earlier: the finer details are ignored and only the sensitivity of the portfolio to a small number of risk factors is important. The difference between an explanatory variable and a risk factor is that there are many explanatory variables, whose correlation is expressed through a few risk factors that capture the movements of large parts of the investment universe. So for example, the asset level of ABC Inc (a Consumer Products company) is an explanatory variable in the Merton model, but the average asset return of Consumer Products is a risk factor.

Notice that if a portfolio is not diversified (by reason of single-name exposures), the large portfolio approximations only tell us about part of the risk. We therefore have to report large exposures separately.

CUSP/PortfolioRisk+

For the record, we round off this chapter with a very brief commentary on CSFB's PortfolioRisk+, and explain how it fits into the picture we have described above.

The first point is that CUSP/PR+ embraces MTM valuation in a way that KMV and CreditMetrics do not. This is why, for example, KMV always talks about expected default frequencies (EDFs), whereas CUSP generates a continuous distribution of credit spread. For investors in investment-grade paper, EDFs are of almost no interest at all. In other words, KMV and CreditMetrics are suitable for calculating economic capital on loan portfolios, but not for "running money".

On the matter of horizon, we opt for 3 months not just for the reasons stated earlier (principally, that this is the horizon of interest of our investors) but also because we use equity option volatility to calibrate CUSP, and that is liquid in the 3-12 month range (there ain't no such animal as a 10-day option).

The table below compares a "typical" market risk system, a "typical" credit risk system and PortfolioRisk+2.

	FI[5] Market risk	CUSP/PortfolioRisk+	Credit risk
Horizon	1-10 days	3 (-12) months	1-10 years or more
Output	PV sensitivities to explanatory variables	Sensitivity of risk measure (VaR etc) to asset allocations and risk factors	Risk measure (VaR etc)
Treatment of credit risk	MTM, usually by consideration of spread volatility within some rating bucket	MTM, Merton model used to give spread distribution	Buy-and-hold: default probabilities (EDFs) from credit ratings or from Merton
Explanatory variables	Rates, FX, sector or rating spread curves	Credit: individual firm's-asset returns IR/FX: swap curve and FX spot movements	Individual firm's-asset returns (KMV) or credit ratings (CreditMetrics)
Risk factors	Obtained from PCA[6] on set of explanatory variables	Credit Risk: Sector factors IR/FX: from PCA on set of explanatory variables	Country/Sector factors

Source: Credit Suisse First Boston

---

[5] Fixed Income.
[6] Principal Components Analysis.

2

The default/no-default world, and factor models

In this chapter we lay the groundwork for future discussions by examining the basics of credit portfolio modeling, factor models, systematic and unsystematic risk.

How are portfolios modeled?

This most basic question must be addressed first. The problem of modeling a portfolio can be expressed thus: we have a large number of assets, each of which can go up or down in value in the course of time. Importantly, their price movements are correlated: they will have a tendency to move in the same direction, though in derivatives markets they can also have a tendency to move in opposite directions instead. The problem is: given some large set of assets, model the distribution of the total portfolio value. However, 'risk' means that we do not know what the gain or loss in portfolio value will be.

One's first inclination is to try to model each asset individually, coming up with a distribution for each. However, we quickly encounter the problem of how to model the correlation: how to glue the distributions together to obtain a statistical description of the probabilities of all possible combinations of movements of the assets, rather than just an expression for each asset in isolation from the others. This 'bottom-up' approach is extremely difficult, however, particularly when the portfolio is large.

An alternative idea is to take a 'top-down' view. The example that we shall concentrate on here is a simple basket of bonds (or loans) subject to risk of default. A manager periodically looking at the performance of the portfolio will notice that the profit-and-loss has been volatile. He will then want to know why, and will probably compare his portfolio with those of other firms. Crucially, there are two important reasons why this volatility occurs, as shown by the following two possible explanations:

A. "From 1996 to 1998 we had bad losses. So did other accounts. This was because of bad economic conditions."

B. "In second quarter 1999, we should have done well, but took a hit on ABC Inc, which pulled us down."

By understanding the difference between these explanations, we converge on the factor approach to credit portfolio modeling, which is the basis of the modern approach. In the first case we can hypothesise the existence of a factor that has caused the default risk of this, and indeed all, portfolios to increase; in the second, the loss was due to a specific event. The second is known as specific risk: if the portfolio manager had had a

smaller exposure to ABC, 'putting fewer eggs in one basket', the loss would have been less. By contrast, the first is known as systematic risk, and it is not alleviated by diversification.

In more detail, the factor approach is based on the following line of argument:

- There are latent (hidden) factors that determine the average loss across the market or sectors in it. Systematic risk occurs because we do not know what the state of the world will be.

- For a large enough portfolio, the observed loss will be determined only by the states of these factors.

- Conditional on these factors being known, losses from individual assets are independent. Independence is a useful property because it is relatively easy to deal with—and whereas there are many ways in which assets can be correlated, there is only one way in which they can be independent.

- Even if we know the state of the world, we do not know exactly what the loss will be in our own portfolio. This 'unsystematic' risk will be particularly significant if our portfolio is small, or contains a few large exposures.

Unsystematic risk can be understood in the following way: suppose that you have a portfolio of 100 loans of similar credit worthiness and the underlying default probability for loans of this credit quality is 1%. Then across the market we see 1% of the loans defaulting. However, the proportion of defaults in your sample might not be 1%: most likely, you will see none, but you might have several defaults. This uncertainty is unsystematic risk. Further, suppose that you do suffer one default. If the loan exposures differ, you have even more risk because your financial loss depends on which one defaults. So unsystematic risk increases when there are concentrations.

To put all of this into perspective, take, for example, credit card loans, for which managers are sometimes heard to say that there is virtually no credit risk. What this means is that over the years it has been observed that credit card default frequency is fairly static. Also a credit card company's portfolio contains a huge number of small exposures. So both the systematic and the unsystematic risk are quite small. Fees and interest payments compensate the expected loss.

The separation of portfolio risk into systematic and unsystematic parts is important because systematic risk is very model-dependent, whereas the analysis of unsystematic risk is purely a mathematical problem. To specify a factor model, we must provide the following information:

- The 'states of the world', i.e., what states it takes and with what probabilities (or it may have a continuous distribution).

- The conditional default probability, or more generally the distribution of the value of each asset, in each state of the world.

Pictorial example

Figure 2.1 gives a good picture of what is going on, in the context of a credit portfolio. We shall make some simplifying assumptions, which will all be relaxed as we develop the subject further. We shall assume that there are five states of the world (more complex models would require many more states, or possibly a continuum of states, but the following is sufficient to illustrate the principle).

Figure 2.1 Five States of the World
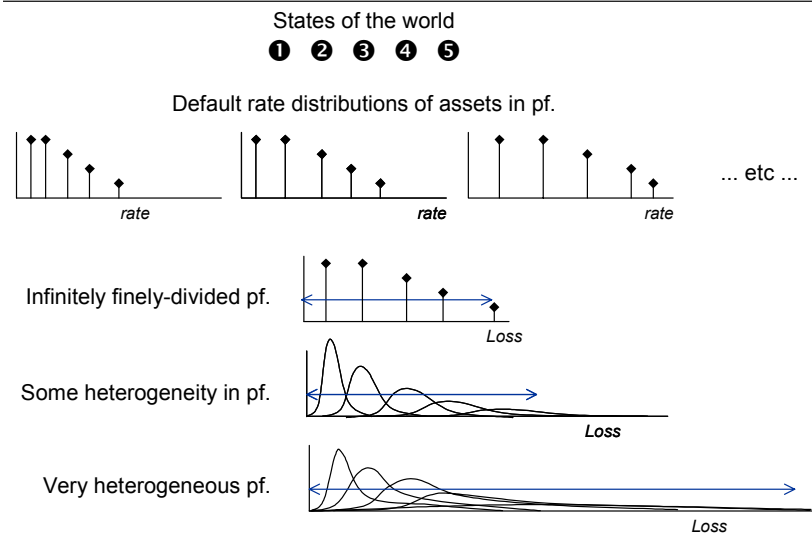
Source: Credit Suisse First Boston

At the top of the figure we have illustrated symbolically five states of the world (❶❷❸❹❺); we must also decide their probabilities. Underneath we have sketched the conditional default probabilities, in each of the five states of the world, for three of the assets in the portfolio. The bottom part of the figure shows loss distributions for different types of portfolio:

1. First, we have the so-called 'infinitely finely divided portfolio'. It is very easy to calculate its loss distribution: in each state of the world, we find the conditional expected loss, which is easily found because it is the sum of the conditional expected losses of the assets in the portfolio. The arrow measures the "risk" of the portfolio, by which we mean the loss at some particular level of confidence, or "VaR". (In fact, Basel II regulatory capital is calculated this way, but we shall not get side-tracked with that.)

2. Second, we have the loss distribution for a portfolio that has some unsystematic risk, which is more realistic than the first example, but harder to analyze. Because there is some uncertainty in portfolio loss even when we know the state of the world, each of the spikes in the infinitely fine-grained portfolio becomes fattened, though its mean position is the same.

3. Third, we have the loss distribution where there is a large amount of unsystematic risk, usually caused by a few big exposures. If one of the bigger ones defaults, the loss will be greater.

What is default probability?

What do we mean by default rate and default rate distribution? Default rate goes under a variety of names, including 'default probability' (PD) and 'expected default frequency' (EDF). Most people have an intuitive idea about what a probability is, but a precise definition is quite elusive. For example, we say that the probability of an unbiased die rolling a six is one-sixth, but what does that actually mean? There are two possible definitions. First, over a very large number of repetitions, one-sixth of the time the die will roll a six. This brings in a notion of replication: we either roll a huge number of 'identical' dice simultaneously and look at the outcomes, or we roll one die many times over. Secondly, a completely different and more abstract approach is to define a probability as a measure of degree of belief. It must satisfy some basic criteria: the probability must lie between 0 and 1; the probability of an impossible event is 0; the probability of an event that must occur is 1; and the probability of mutually exclusive events (i.e., events that cannot occur together) is just the sum of their individual probabilities. In this case we would say that the statement that 'the probability of an unbiased die rolling a six is one-sixth' is not a definition of 'probability', but rather a definition of 'unbiased'! Both definitions have their uses: the first is fundamental to statistical theory, while the second is fundamental in pricing.

For our purposes, we shall need both definitions. For the moment, we are most interested in the statistical notion of probability, but we are faced with the question of where the large number of repetitions is supposed to come from. In fact, default rate is meaningful only when applied to asset *classes*, by which we mean a large group of assets of similar type (industry, sensitivity to economic and sectorial factors, etc.) and differing only in idiosyncratic (company-specific) risk. The proportion of defaults across the asset class is then called the default rate. This is rather like rolling a truckload of dice in order to find the probability of a six: each die has the same basic characteristics (size, shape, distribution of mass) but differs in 'specific' ways such as its position, linear and angular velocity when it is thrown, and so on.

Finally, default rate or probability must also be defined for a particular time horizon. The default probability over one year might be 0.50%, but for a two-year horizon it is not necessarily double this. (To argue at a trivial level, it would not be exactly double this anyway, but instead $1-(1-0.0050)^2$. However, this is unimportant in context.) For example the two-year default probability could be 1.20%, which would mean that even if the obligor survives one year, there is a higher probability of default in the second year (roughly 0.70%), reflecting the view that a downgrade in the firm's credit quality is likely.

What is default rate distribution?

Why are default events correlated? There are several ways in which default events of two obligors, ABC and XYZ, can be, or appear to be, dependent. One is that the fortunes of XYZ are directly linked to those of ABC. This linkage would occur if ABC were one of XYZ's biggest customers. This relationship is very difficult to model and it ignores a much more important effect already discussed: systematic risk. In this view, default events are not really dependent at all: rather, default of ABC is indicative of the fact that we are in a bad state of the world, and so XYZ (and similar obligors) are more likely to default.

Viewed this way, correlation between default events of obligors can be thought of as a measure of how dependent they are on systematic factors. The figure below gives a good pictorial view:

Figure 2.2 Default rate volatility and default correlation are similar concepts
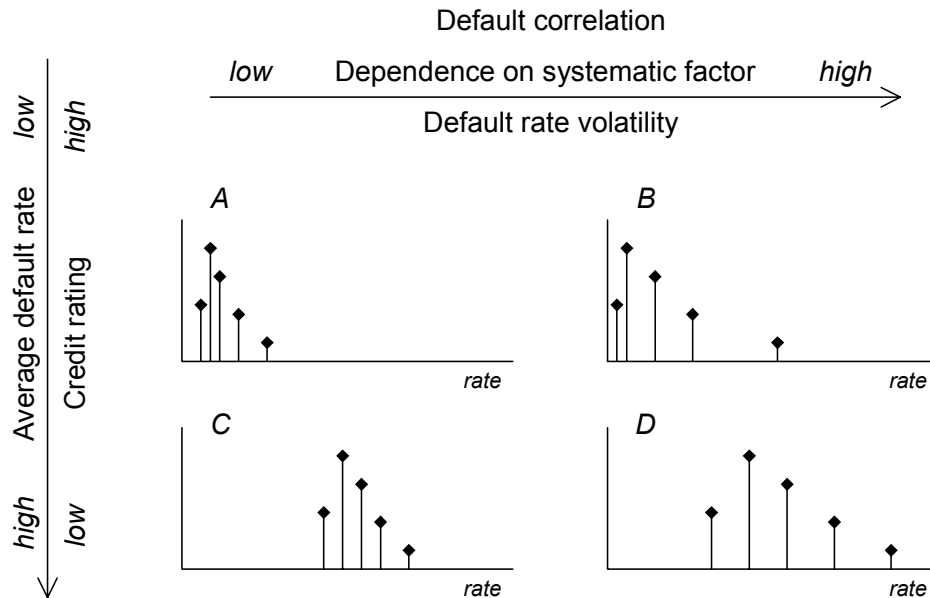
Source: Credit Suisse First Boston

Each of the graphs shows the default rate in each state of the world for a particular type of obligor. Graphs A and B show the situation where the average default rate is low, and C and D where it is high. More subtle is the distinction between the left two and right two graphs. For A and C we see that the default rates of the obligors are mildly sensitive to the state-of-the-world, but for B and D the sensitivity is considerably higher (remember that in going from left to right the average default rate remains the same). Consequently a portfolio of type D will be riskier than C because there is more uncertainty in the default rate: in bad years it will suffer much worse. Of course, D is riskier than B also, but that is largely due to the difference in credit quality.

If the default rate of an obligor does not depend at all on the state-of-the-world, then its default events will be uncorrelated with those of all other obligors.

Numerical example

Persisting with the idea of five states of the world, let us build the following hypothetical example.

Table 3.1

Class	<<<<<<< States of the world >>>>>>					Mean	Stdev
	34%	40%	20%	5%	1%		
P0	0.25%	0.25%	0.25%	0.25%	0.25%	0.25%	0.00%
P1	0.20%	0.23%	0.30%	0.44%	0.80%	0.25%	0.08%
P2	0.10%	0.20%	0.40%	0.80%	1.60%	0.25%	0.22%
Q0	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	0.00%
Q1	0.80%	0.92%	1.20%	1.76%	3.20%	1.00%	0.32%
Q2	0.40%	0.80%	1.60%	3.20%	6.40%	1.00%	0.86%

Source: Credit Suisse First Boston

The probabilities of the five states occurring are given in the top row (34% ... 1%). The figures underneath show the default rates for each of six classes of obligors in each state. Obligors in class P0, P1 or P2 all have the same average default rate (0.25%), and those in class Q0, Q1 or Q2 also have the same average default rate (1%). For example, P0-P2 would roughly correspond to S&P credit rating BBB+ and Q0-Q3 to credit rating BB+. To check that the mean default rates are correct is quite simple, e.g. for P1:

$$34\% \times 0.20\% + 40\% \times 0.23\% + ... + 1\% \times 0.80\% = 0.25\%.$$

The difference between P0, P1 and P2 is in their default rate volatility, and hence correlation. The last column of the table shows the standard deviation of the default rate. This is a measure of variability, defined as the square root of the mean square deviation from the mean. For P0 there is no deviation from the mean, so the standard deviation is zero. For P1 the mean square deviation is

$$34\% \times (0.10\% - 0.25\%)^2 + 40\% \times (0.23\% - 0.25\%)^2 + ... = 6.34 \times 10^{-7}$$

and taking the square root gives 0.08% as in the table.

Correlation numbers: A useful construct?

Once these calculations with default rates have been mastered it is easy to find default correlations. The probability that an obligor of class P1 and one of class Q1 default is

$$34\% \times 0.20\% \times 0.80\% + 40\% \times 0.23\% \times 0.92\% + ... = 2.75 \times 10^{-5}$$

The product of their individual default probabilities is smaller than this:

$$0.25\% \times 1.00\% = 2.5 \times 10^{-5}$$

so there is positive correlation between the default events of the obligors. Numerically the correlation coefficient is given by

$$\frac{2.75 \times 10^{-5} - 2.5 \times 10^{-5}}{\sqrt{0.25\%(1 - 0.25\%)}\sqrt{1.00\%(1 - 1.00\%)}} \approx 0.00051$$

This number is generally referred to as the *default (event) correlation*. By a similar calculation the default event correlation between two obligors of type P1 is 0.00025. At first sight this looks very small: a correlation coefficient is standardized to lie between -1 and 1, with perhaps 0.2 or 0.3 being regarded as a low or moderate level of correlation, and 0.7 or 0.8 a high level. Surely then a correlation coefficient less than 0.001 is so small it can be regarded as negligible? In fact, as we are about to see, this is not the case. The default event correlation is not a particularly good guide to the effects of correlation on portfolio risk. When the risk comes in the form of rare events, the event correlations will typically be very small also.

It is not difficult to show that when an obligor has the same default rate in each state of the world, its default is not correlated with that of any other obligor.

We shall not have much cause to refer to correlation coefficients again, but for the sake of completeness here is the mathematical definition:

$$\rho^D = \frac{p_{12} - p_1 p_2}{\sqrt{p_1(1-p_1)}\sqrt{p_2(1-p_2)}} \,.$$

Here p_1 and p_2 are the individual default probabilities and p_{12} is the joint default probability (probability that both obligors in question default). We said that the correlation coefficient must lie between -1 and 1 but in fact we can do better. Suppose that $p_1 \leq p_2$ (if not, interchange p_1 and p_2). Now $p_{12} \leq p_1 \leq p_2$ (the probability that both occur must be less than either of the individual default probabilities) and so

$$\rho^D \leq \sqrt{\frac{p_1}{1-p_1} \bigg/ \frac{p_2}{1-p_2}} \leq 1$$

This upper bound is only equal to 1 when the two obligors have equal default probability; otherwise it must be less. For example, if $p_1=0.25\%$ and $p_2=1.00\%$ then

$$\rho^D \leq \sqrt{\frac{0.0025}{0.9975} \bigg/ \frac{0.01}{0.99}} \approx 0.5$$

So it is simply not possible for the default correlation to be 60%. **To try and impose a constant default event correlation across a portfolio is bound to lead to nonsense.**

Portfolio analysis

If the portfolio consists entirely of obligors of class P0, there is no correlation and no systematic risk: whatever state of the world we are in, the expected proportion of defaults in the portfolio is always the same (0.25%). However, if all the obligors are of type P1, the proportion of defaults does depend on the state of the world, varying from 0.20% in the best case to 0.80% in the worst case. If the portfolio consists entirely of obligors of class P1, then we can be 95% confident that the default rate will be 0.44% or less; to be 99% sure we would have to alter this figure from 0.44% to 0.80%. To take a specific example, suppose there are 10000 loans (a large portfolio) of notional 1M$. Then the credit VaR (value at risk) at 95% confidence is

$$10000 \times 1M\$ \times 0.44\% = 44M\$$$

and at 99% (or, indeed, at confidence levels higher than that because there are no worse states in this simple example and we are neglecting unsystematic risk[7]) the VaR is

$$10000 \text{ x } 1M\$ \text{ x } 0.80\% = 80M\$.$$

For a portfolio of P0 obligors both these figures would have been only 25M$. This shows the impact of correlation: the VaR at 99% confidence has been multiplied by more than 3 (80M$ ÷ 25M$) when a superficially tiny degree of correlation has been introduced (tiny as measured by the default event correlation coefficient, which we said was less than 0.001).

Putting some structure in

One of the nice things about the framework we have set up is that it is quite transparent: as we have seen, we can at least in simple cases perform credit VaR calculations 'on the back of an envelope' without the need for any complex analytics or Monte Carlo simulation. Set against this, however, is the problem that the framework is too 'floppy': we have not said how we obtain the conditional default probabilities in the table, or the 'states of the world'. To an extent they are subjective—in some markets very subjective—but even so we would like some way of obtaining a structure via a small set of parameters rather than having to specify the whole matrix ourselves.

There are several ways to proceed, one of which is through the use of *copulas*. We shall describe these in more generality later but for the moment we shall stick to a simple idea, the Gaussian copula model. This can be thought of as a spin-off from the Black-Scholes-Merton framework, that is, the use of option pricing concepts in corporate finance. We regard it only as a spin-off because some corners have to be cut, and we do not want to go into the details of the Merton framework just yet. However there is a reasonable amount of financial intuition behind it and it is one of the most popular approaches in correlation modeling.

Gaussian copula (quasi-Merton) model

The financial intuition behind this model is that default of an obligor occurs if, at the horizon in question, the level of its assets is below that of its liabilities (strike). The asset process is modeled as a geometric Brownian motion, in common with the well-known Black-Scholes analysis. Then the logarithm of the asset value at the horizon is Normally distributed, and after standardizing can be assumed to have mean zero and variance 1. We call the standardized log-asset Z, and the standardized log-strike ζ. The default probability is then given as

$$p = P(Z < \zeta) = \Phi(\zeta)$$

with Φ denoting as usual the cumulative Normal distribution function. In option-theoretic models, p would be derived from ζ which would in turn depend on the capital structure and the level and volatility of the firm's assets. Here though we shall assume that p is given e.g. through internal ratings.

---

[7] Unsystematic risk, always present in a finite portfolio, causes the VaR at higher percentiles to rise above this figure.

To make two assets have correlated defaults, we just have to correlate their Z's, which is easy because multivariate Normal distributions are well-understood. The correlations between the Z's can be expressed through a correlation matrix, but this is not a very efficient construction, particularly for a large portfolio. A better approach is to write the asset returns in factor form. In doing this, each asset return is written as a sum of correlated parts (factors) and an uncorrelated part. These parts, which are random, are Normally distributed with mean zero and variance 1. The factors, which can be chosen to be independent of each other, are common to all obligors, and the correlation between obligors is expressed through the degree of dependence on each factor, much as before. The uncorrelated parts correspond to firm-specific effects; they are also Normally distributed with mean zero and variance 1, and they are uncorrelated from each other and from all the correlation factors. The simplest specification is the one-factor model, for which an example (with numbers filled in) is

$$Z_1 = 0.60V + 0.80U_1$$
$$Z_2 = 0.28V + 0.96U_2$$
$$\ldots$$

Here V is the common part of the asset returns and U_1, U_2 are the uncorrelated parts. For each obligor the coefficients have to be chosen so that the sum of their squares is equal to 1 (Z_i is to have unit variance, and V is independent of U_i so their variances add):

$$0.60^2 + 0.80^2 = 0.28^2 + 0.96^2 = 1.$$

The correlation between the *asset returns* (**not** the correlation between the default events) can easily be computed:

$$0.60 \times 0.28 = 0.168$$

(The computation is easy because their mean is 0 and their variance is 1.) Symbolically the model can be written

$$Z_1 = c_1 V + \sqrt{1 - c_1^2}\, U_1$$
$$Z_2 = c_2 V + \sqrt{1 - c_2^2}\, U_2$$
$$\ldots$$

and in multiple factor form

$$Z_1 = c_{11} V_1 + \cdots + c_{1m} V_m + \sqrt{1 - (c_{11}^2 + \cdots + c_{1m}^2)}\, U_1$$
$$Z_2 = c_{21} V_1 + \cdots + c_{2m} V_m + \sqrt{1 - (c_{21}^2 + \cdots + c_{2m}^2)}\, U_2$$
$$\ldots$$

Additional factors can be thought of as contributions to the asset return coming from sectorial (i.e. industry) or country effects.

We return to the one-factor model and now need to find the conditional default probability of an obligor in each state of the world. By state of the world, we mean the event that the factor variable V takes some particular value: as V is Normally distributed it can take a continuum of values. The derivation is as follows: we have for one asset

$$Z = cV + \sqrt{1 - c^2}\, U$$

so conditionally on *V* (i.e. fixing V) the only variability is in the obligor-specific part *U*. So the conditional default probability is

$$P(Z < \zeta \,|\, V) = P\left(U < \frac{\zeta - cV}{\sqrt{1-c^2}}\right) = \Phi\left(\frac{\zeta - cV}{\sqrt{1-c^2}}\right)$$

We can make a few observations about this:

When *c*=0 the default rate is always $\Phi(\zeta)$, regardless of *V*. This means that when there is no correlation, there is no variability in default rate, which is what we saw earlier.

When *c* approaches 1, the conditional default probability only takes two values: zero (when $V>\zeta$), and unity (when $V<\zeta$). If several assets have the same default probability (same ζ) then they will all default together or not-default together, so are 'perfectly correlated'. More generally, once *V* is known there is no uncertainty about whether any particular asset defaults or not, and hence there is no unsystematic risk. This is to be expected.

One can show by direct integration over *V* that, regardless of the value of *c*, the average default rate is $\Phi(\zeta)$, i.e. *p*, as it should be.

To plot the actual variation of the default rate, we allow *V* to vary. Here are examples that loosely correspond to A, B, C, and D earlier. The top two graphs are for an average default rate of 1%, the lower two for 4%; the right two have higher default rate volatilities than their counterparts on the left hand side because the correlation parameter *c* is higher (0.10 for the left two, 0.40 for the right).

>>> Increasing default correlation / systematic dependence / default rate vol >>>
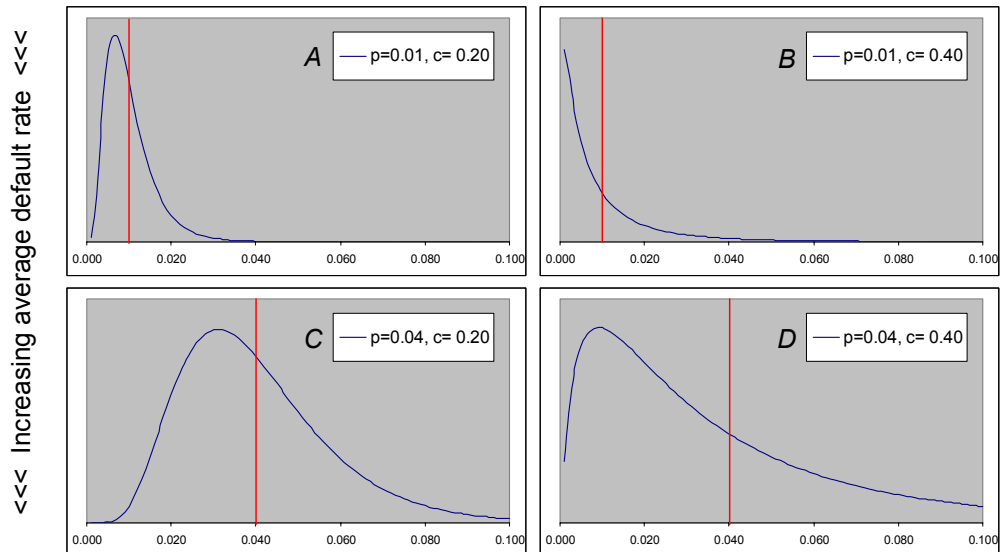
Portfolio analysis with 1-factor Gaussian copula

We gave examples of how to calculate VaR using the discrete 5-state model earlier. We now repeat this exercise with a more complex structure of portfolio and using the Gaussian copula model instead. Again we shall take a large enough portfolio to render the effects of unsystematic risk negligible.

Suppose that a portfolio consists of n bonds with exposures a_j, average default probabilities p_j and correlations c_j. Conditionally on the value of the factor variable V, the expected loss on the portfolio is obtained by multiplying exposure by the conditional default rate and summing over the portfolio:

$$Y(V) = \sum_{j=1}^{n} a_j \Phi\left(\frac{\Phi^{-1}(p_j) - c_j V}{\sqrt{1 - c_j^2}} \right)$$

To work out the VaR at 99% confidence, we simply need to find the '99% worst value' of V and stick it in. For a Normal distribution this is at $V=-2.326$ (note that it is when V is negative that the losses are high). So the VaR is $Y(-2.326)$.

Here is a specific example. The portfolio of 2000 bonds has been partitioned for ease of exposition and is shown below. Also for simplicity we have assumed each bond to have a notional, and hence loss given default, of 1 (we assume zero recovery). Using the formula above we can plot the VaR for different tail probabilities. The graph shows for example that the VaR at 99.9% confidence (tail probability 0.001, $V=-3.09$) is about 235.

Loss distribution of test portfolio
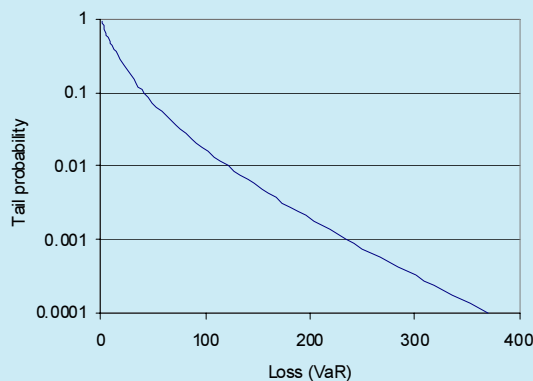
Source: Credit Suisse First Boston

Composition of test portfolio

	Qty	Notional	Def prob	Correl
1	150	1.00	0.0009	0.34
2	200	1.00	0.0014	0.52
3	150	1.00	0.0040	0.55
4	300	1.00	0.0081	0.36
5	450	1.00	0.0082	0.34
6	200	1.00	0.0098	0.46
7	150	1.00	0.0117	0.45
8	200	1.00	0.0127	0.50
9	100	1.00	0.0179	0.44
10	100	1.00	0.0184	0.49

Source: Credit Suisse First Boston

Conclusions

We have shown how to set up models of systematic risk and also shown that portfolio loss calculations can be carried out quite easily in simple cases. Our next two chapters deal with the following issues:

- Moving from a default/no-default to a mark-to-market model that can also be extended across other asset classes;

- Developing copulas in more depth and showing that the copula method is not fundamentally different from the conditional independence approach.

- Understanding unsystematic risk.

3

Risk and optionalities

In the previous chapter we discussed factor models in a default/no-default context. For traded securities this is too restrictive a treatment and we now move towards a mark-to-market treatment of portfolios. We develop a theory of explanatory variables and "transfer functions" and show that these are a convenient way of modeling correlated nonlinear assets.

Nonlinear assets

In olden days, portfolio theory was developed using Gaussian distributions and the joint distribution of all the assets in the portfolio was assumed to be multivariate Gaussian. The advent of derivative products such as options, which have nonlinear payoffs, made this approach untenable. To understand what nonlinear payoffs can do, have a look at Figure 3.1.

Figure 3.1. (a) local analysis shows no sensitivity to underlying variable; (b) full picture may have a nasty surprise.
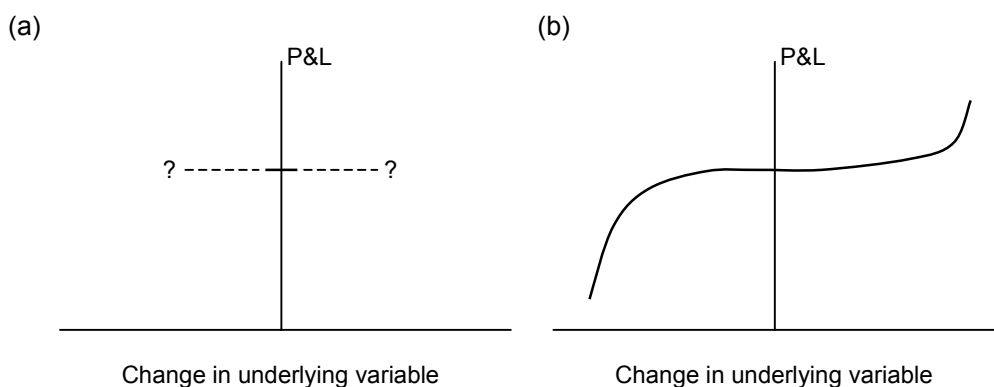
The picture that we have in mind is that there exist explanatory variables (or "underlyings"); the values of the portfolio's assets are deterministic functions of these.

Indeed the Merton model says this, and the deterministic functions alluded to above are just the Black-Scholes formulae. The equity is a call option on the firm value[8] S, with strike equal to the face value of the debt (K):

$$E_t = S_t \Phi(d_+) - Ke^{-r(T-t)} \Phi(d_-),$$

with

$$d_\pm = \frac{\ln(S_t / Ke^{-r(T-t)}) \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}},$$

while the debt is the risk free debt (K) plus a short put option:

$$D_t = S_t \Phi(-d_+) + Ke^{-r(T-t)} \Phi(d_-).$$

(Note that by put-call parity we have that the whole is the sum of the parts: $S_t = E_t + D_t$.)

So both the equity and the debt can be written as functions of the underlying asset level. For reasons of familiarity (and convenience when defining the correlations) it is useful to have Normally distributed explanatory variables. However, S_t is lognormal, so we write it as the exponential of the normalized log-asset-value, whose distribution is N(0,1):

$$S_t = S_0 \exp((\mu - \tfrac{1}{2}\sigma^2)t + \sigma\sqrt{t} Z_t).$$

So now the equity and debt can be expressed as functions of[9] the normalized log-asset-return Z_t:

$$E_t = g_t^E(Z_t), \quad D_t = g_t^D(Z_t),$$

with

$$g_t^E(z) = S_0 e^{\sigma\sqrt{t}z + (\mu - \frac{1}{2}\sigma^2)t} \Phi(d_+(z)) - Ke^{-rT}\Phi(d_-(z))$$

$$g_t^D(z) = S_0 e^{\sigma\sqrt{t}z + (\mu - \frac{1}{2}\sigma^2)t} \Phi(-d_+(z)) + Ke^{-rT}\Phi(d_-(z))$$

$$d_\pm(z) = \frac{z\sigma\sqrt{t} + \ln(S_0 / Ke^{-rT}) + (\mu - r - \frac{1}{2}\sigma^2)t \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

Roughly speaking there are only two important parameters needed to specify this, the simplest form of the Merton model:

- $m = \ln(S_0 / Ke^{-rT})$, the "moneyness" of the call option (on S with strike K), is the log of the ratio of the asset level to the strike. If we adopt the definition that leverage is the ratio of the discounted face value of the debt to the PV of the assets of the firm (PV of equity + PV of debt), i.e. Ke^{-rT}/S_0, then we have leverage = e^{-m}. So the higher m is, the more in-the-money is the call option, and the lower the leverage of the firm. The lower m is, the more in-the-money is the put option, and the higher the leverage of the firm. When $m<0$, the company is technically insolvent, or has "negative equity".

- σ, the volatility of the firm's assets. All other things being equal, raising the volatility causes the equity to become more valuable and the debt less valuable.

---

[8] Note: S does NOT denote stock price.
[9] The subscript $_t$ in g_t indicates that there is time-dependence too, but this is not as important.

That these should be independent is a sensible notion: a firm can control its leverage by issuing or repaying debt, but its volatility is mainly affected by its business environment (called an "exogenous variable"). The parameter μ is the expected growth of the firm value. It has no impact on the pricing of the debt or equity of the firm, but when the "real" distribution of the debt or equity are required at some point in the future it does come into play. For the purposes of this chapter, however, we shall carry out all the analysis in the "risk-neutral" measure and just set $\mu=r$. That means that the expected gain in buying and holding any instrument is zero, and we just concentrate on the risk (uncertainty in value).

Following electrical engineering parlance, we call the function g the *transfer function* for the asset in question (the transfer function contains the information about how a component behaves when a certain input is applied).

If the transfer function is linear, $g(z)=\Delta.z+g_0$, and if the explanatory variable has distribution N(0,1), then the asset in question has a Normal distribution with mean g_0 and variance Δ^2. More generally, if g is nonlinear, then we can think of its gradient at the origin ($z=0$) as a sort of "instantaneous volatility". In essence this is because if Z makes a small movement then $g(Z)$ moves by Δ times as much, and so it is Δ times as volatile as Z is. As we said at the outset, in our discussion of PV sensitivity, Δ is the simplest guide to risk.

But when we are dealing with options we know perfectly well that the transfer functions aren't going to be linear, and indeed the Black-Scholes ones are not. Figure 3.2 shows some examples. In each graph the horizontal axis is normalized so that the mean of the explanatory variable is 0 and the standard deviation is 1. This means that the horizontal axis can be associated with a probability in a simple way (as we shall show by example). Also the debt price is shown as a fraction of its discounted face value, so 1 corresponds to risk-free, and to make things simple we are assuming that the debt is a zero-coupon bond. In (a) the horizon is $t=0.5$yr, the maturity is $T=1$yr, the asset vol is $\sigma=50\%$ and the leverage parameter is $m=0.5$. The debt then prices at $97.39 (for $100 face value), so over the next 6 months it can increase by $2.61 to $100 at best, or go all the way down to zero. The probability of the debt losing one-tenth of its value (0.90 on the vertical axis) is $\Phi(-1.5) = 6.7\%$ (as -1.5 is the corresponding coordinate on the horizontal axis). Figures (b-e) show the effect of varying the volatility, leverage, and term of the debt.

Generally speaking the graphs for debt all have the same kind of shape, the most important point being that they all exhibit negative convexity. The main difference is the degree of asymmetry. The worst possible case (a limiting case, in fact) is always that the debt becomes worthless, though in practice this is often very unlikely. When the volatility or leverage is low, or the debt is short term, the value of the credit risk (i.e. the value of the embedded put) is very small. This means that there is little upside but, relatively, a substantial downside. By contrast, when the volatility or leverage is high, or the debt is long-term, the value of the credit risk is much higher. This has two effects: the "delta" is higher (gradient of the transfer function at the origin), so the debt is more volatile; but there is more symmetry between upside and downside—so the risk is more "volatility risk" than "tail risk".

Figure 3.2. Transfer functions for equity and debt (respectively, long call and short put in the Merton model)

(a) Horizon *t*=0.5yr, maturity *T*=1yr, vol σ=50%, leverage *m*=0.5. Debt spread is 261bp. The debt has a much bigger downside than upside, and the opposite is true for the equity.
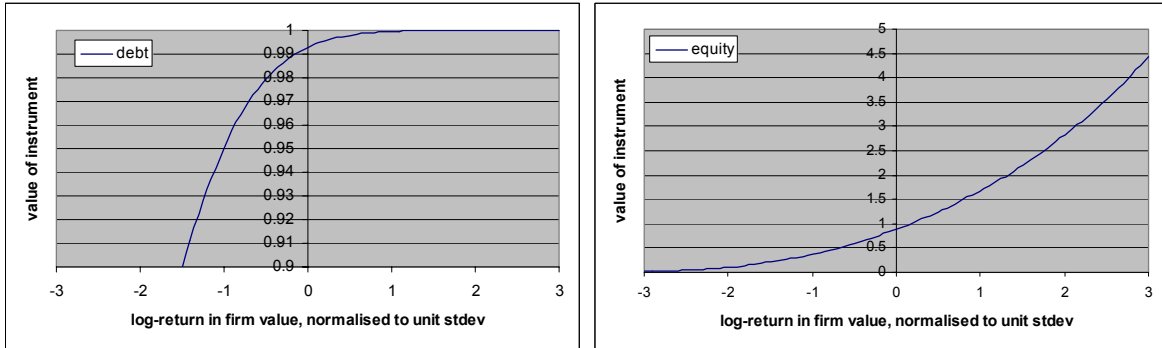
(b) *t*=0.5yr, *T*=1yr, σ=50%, *m*=0.25. Reducing the leverage makes it much less likely that the debt will be hit: so it is worth more (spread 8bp), but has almost no upside.
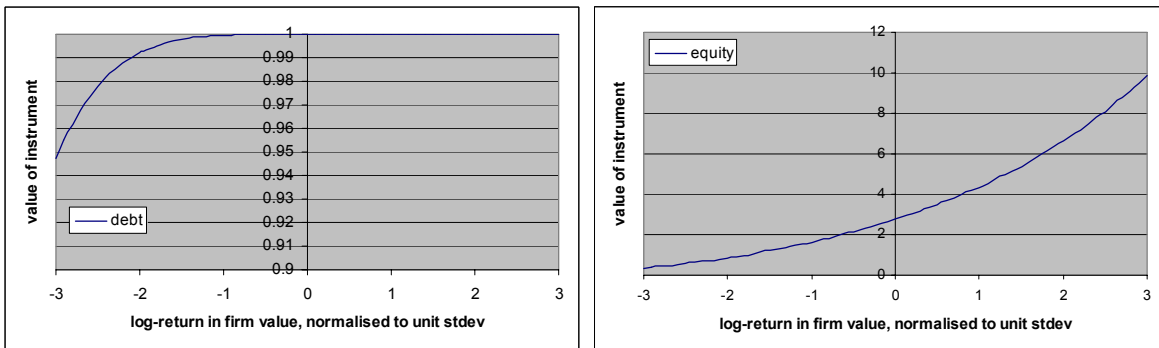
(c) *t*=0.5yr, *T*=1yr, σ=25%, *m*=0.5. Reducing the asset vol also makes it much less likely that the debt will be hit, so, as in going from (a) to (b), the debt is worth more (spread 3bp), but has almost no upside.
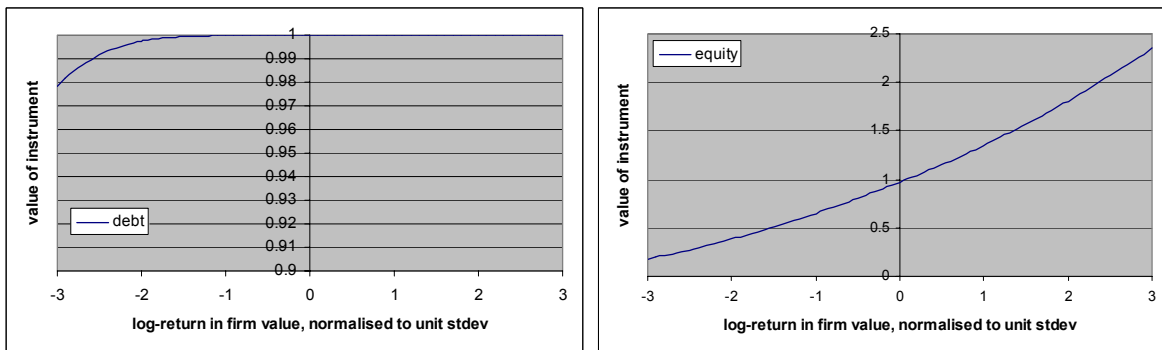
CREDIT SUISSE | FIRST BOSTON

Figure 3.2 cont. Long-term debt is more likely to default, but relatively there is less tail risk associated with it: in (d) the debt price is $99.77 (per $100 face) so its maximum gain is $0.23, but at 97.5% confidence[10] it loses $0.65 which is almost 3 times bigger. The position is much less asymmetric for (e), where the debt maturity is longer: debt price is $97.61, maximum upside is $2.39, downside at 97.5% confidence is $2.84. For the equity there is very little difference between (d/e).

(d) t=0.5yr, T=5yr, σ=25%, m=0.25.
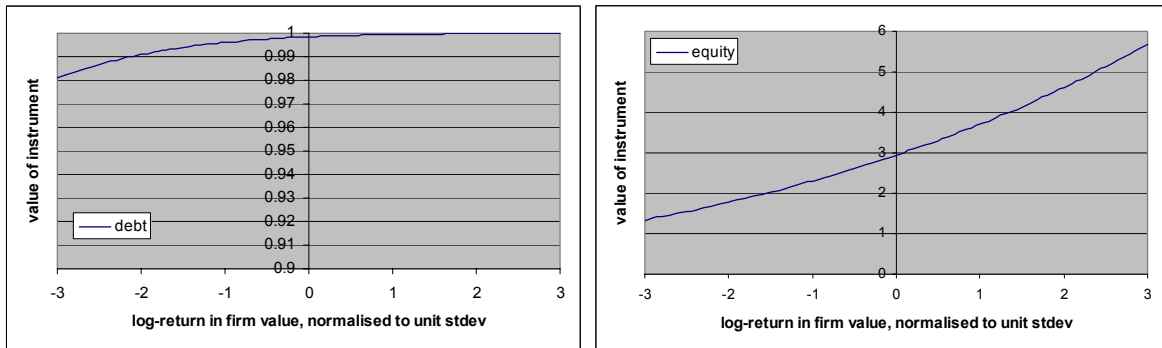
(e) t=0.5yr, T=10yr, σ=25%, m=0.25.

---
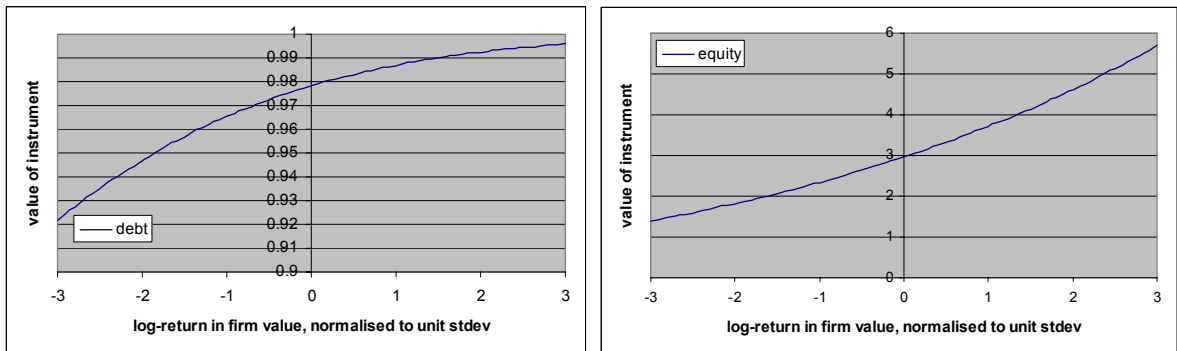
[10] -1.96 standard deviations, hence -1.96 on horizontal axis.

As another example of a transfer function, consider the buy-and-hold (default/no-default) model of credit risk, that we introduced in the last chapter, expressed as follows. The value of the asset (loan, bond, etc.) is given by

$$X = g(Z) = \begin{cases} 1, & Z > \zeta \\ r, & Z < \zeta \end{cases}, \quad \zeta = \Phi^{-1}(p),$$

where ζ is the "threshold of default", p is the default probability (EDF) and r is the recovery rate. Obviously this g is not linear either:
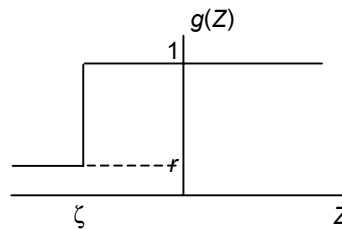
Transfer function for a default/no-default model.

Asset distributions

Let us now turn to questions of portfolio modeling.

It is apparent that if the transfer function of some particular asset is nonlinear, and the explanatory variable is Normally distributed, then the distribution of the asset's return can no longer be Normal. It is also apparent that, once the transfer function and the explanatory variable's distribution have been defined, the distribution of the asset's value must be uniquely determined. This seems obvious, but it is not the only way of going about portfolio modeling: as we intimated earlier, another seemingly "obvious" approach is to find the marginal distribution of each asset and then work out how to correlate them (and in fact the copula approach points to this method). We are going to show that directly modeling the distributions of assets is not a good idea, and that the transfer function / explanatory variable approach is preferable.

First, if we wish to find the marginal distributions of the assets from the transfer function and the distribution of the explanatory variable, we can use the following result. Suppose that the probability density function (pdf) of Z is $f(z)$, and suppose that $X = g(Z)$. Then provided that g is differentiable and strictly monotonic (increasing or decreasing)[11], the pdf of X is given by[12]

$$h(x) = \frac{f(g^{-1}(x))}{|g'(g^{-1}(x))|},$$

---

[11] Has no turning-points.
[12] g' denotes the first derivative of g, and g^{-1} the inverse of g, i.e. by $y = g^{-1}(x)$ we mean that y satisfies $g(y) = x$.

which is a well-known result[13]. But the problem starts to become awkward if g has one or more turning-points. In that case there may be several values of z that satisfy $g(z) = x$; however, by cutting the function g into pieces, the same method can be followed. One arrives at the following generalization:

$$h(x) = \sum_{z:g(z)=x} \frac{f(z)}{|g'(z)|}.$$

In essence, what is happening is that Z is 'folded up' to make X, and so the density of X at some point (x say) comes from contributions from various points on the distribution of Z. There are further complications if the derivative of g vanishes on some set of non-zero measure, as the pdf of $g(Z)$ then contains discrete bits (delta functions).

So we can find an expression for the pdf of X, but this formula is difficult to use in practice. For one thing, it involves searching the graph of $x = g(z)$ making sure that all the solutions have been found for the value of x in question, which is a tedious and open-ended problem. Therefore, it might therefore appear that working directly with the distributions is a better idea than working with transfer functions (which would then remove the need to calculate the distribution from the transfer function). However, this is an illusion, for two reasons.

Reason 1

First, given the distribution of the explanatory variable Z, and the transfer function g, we can *in principle* find the distribution of the asset $X=g(Z)$ (even though, as we have said, it is awkward in practice). But the converse is false: given the distribution of the asset X, we cannot *even in principle* find the transfer function.

This is fairly obvious, but here are two examples. We are assuming that Z is Normally distributed, but the principles are valid for more general distributions too.

Example 1. Take $X_1 = g_1(Z) = Z$, and $X_2 = g_2(Z) = -Z$. Then X_1 and X_2 have the same distribution (because the distribution of Z is symmetrical about the origin), and so we cannot tell the difference between a long and a short position in an asset. In principle this ambiguity would be resolved if we stored extra information, i.e. long or short, but the next example shows that the problem is much worse.

Example 2. Take

$$X_1 = g_1(Z) = Z^2$$

$$X_2 = g_2(Z) = \mathrm{Ga}^{-1}(\tfrac{1}{2};2;\Phi(Z))$$

$$X_3 = g_3(Z) = \mathrm{Ga}^{-1}(\tfrac{1}{2};2;\Phi(-Z)).$$

---

[13] Either by Jacobians or by the following argument. Denote by F the cumulative distribution function of X. If g is an increasing function then

$$P(X < x) = P(g(Z) < x) = P(Z < g^{-1}(x)) = F(g^{-1}(x))$$

$$\therefore h(x) = \frac{d}{dx}P(X < x) = \frac{f(g^{-1}(x))}{g'(g^{-1}(x))} = \frac{f(g^{-1}(x))}{|g'(g^{-1}(x))|}$$

as required (the last line follows by the chain rule of differentiation). The case for decreasing g is similar.

Figure 3.3. Different instruments can have the same P&L distribution. Transfer functions for Example 2: (top row, left to right) $g_1(z)$, $g_2(z)$, $g_3(z)$ vs. z. When the 'input' (Z) is Normally distributed (bottom left), the 'output' (respectively X_1, X_2, X_3) has the same distribution in each case (bottom right). Asset X_2 is 'call-like', asset X_3 is 'put-like', and asset X_1 is 'like a call plus a put'. The conclusion is that the distribution of X contains less information than the transfer function.
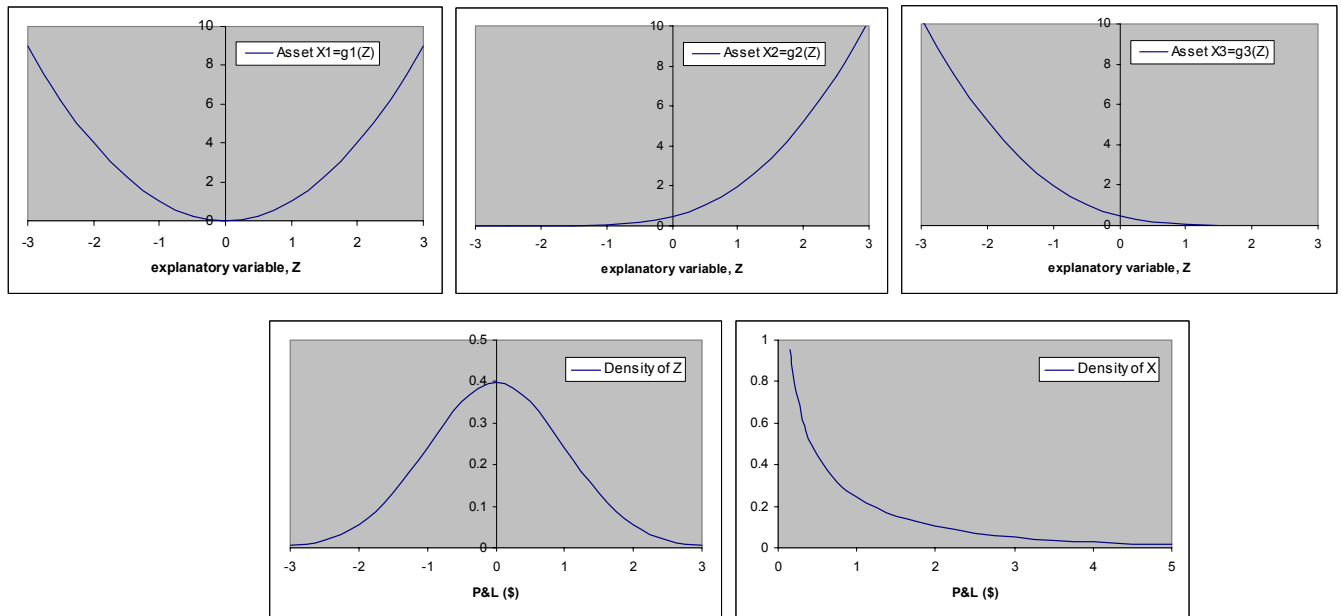
Here $Ga(\alpha,\beta,\cdot)$ denotes the cumulative distribution function of the Gamma distribution with parameters α, β. As it is not obvious where this example comes from, we explain it now. The pdf (density function) of X_1 is

$$h_1(x) = \frac{e^{-\frac{1}{2}x}}{\sqrt{2\pi x}} \quad (x > 0)$$

i.e. Gamma with parameters $\alpha=\frac{1}{2}$, $\beta=2$ (also known as chi-squared with one-degree of freedom). The variables X_2 and X_3 are then constructed to make them have that distribution too. The logic is that $\Phi(Z)$ is uniformly distributed, and to sample from a distribution of your choice you just take a uniformly distributed variable and pass it through the relevant inverse cumulative distribution function. That X_2 and X_3 have the same distribution is just the same symmetry argument used in Example 1.

Why are there totally different ways of ending up with the same distribution? Answer: the transformation for X_1 is two-to-one (because $g_1(z) = g_1(-z)$), whereas those for X_2 and X_3 are one-to-one. You can think of X_2 as being a bit like the value of a long call option (before expiry) on Z, and that of X_3 as being a bit like a long put option; but X_1 is more subtle, being similar to the value of a long put and long call combined.

So the transfer function contains information that is lost in going to the density. But is that information important? Yes, very much so. In Figure 3.3, focus on the bottom right-hand figure, which shows the distribution of the asset X = any of (X_1, X_2, X_3) (they all have same distribution). For each of the three assets (X_1, X_2, X_3) there are three different interpretations:

- For X_1: the gain is highest when Z moves a long way in *either* direction;

- For X_2: the gain is highest when Z *increases*;

- For X_3: the gain is highest when Z *decreases*.

Suppose to make things concrete that Z is the return on IBM stock. A risk manager has no idea, looking only at the distribution of X, whether his upside comes from IBM going up, down or both!

Reason 2

It is often necessary to aggregate, or at least to model in a coherent way, different assets that have the same explanatory variable. Here are some examples:

- long 5Y CDS and long 7Y bond of the same issuer;

- long equity and long CDS of the same issuer;

- short equity put option and long CDS of the same issuer.

(In each case we are assuming that there is only one explanatory variable, the firm's asset level. A more complex model would require the volatility to be taken into account as well: the equity, CDS etc., are options on the firm's assets, and hence their price is dependent on volatility.) Such situations are easily dealt with using transfer functions, because they simply add.

So, if two or more instruments are sensitive to the same explanatory variable, the transfer function corresponding to their sum is just the sum of their transfer functions. So if X_1 is one asset and X_2 is another, represented by g_1, g_2 respectively, and we wish to combine them to make one asset X_3, with weights a_1, a_2, then the transfer function that represents X_3 is just

$$g_3(z) = a_1 g_1(z) + a_2 g_2(z).$$

The distribution, on the other hand, doesn't obey a simple rule. So transfer functions are easier to deal with. Incidentally, once the decision has been made to represent assets via their transfer functions, it is necessary to decide how to store that representation in numerical form. There is no 'one correct' way of doing this, but it is important to choose a method that allows transfer functions to be added and scaled[14]. A polynomial approximation is the easiest.

---

[14] More formally, the space of functions is closed under addition and under scaling, or, for short, is a real vector space.

This point about aggregation takes us back to Reason 1. Although (in the example in Figure 3.2) X_1, X_2, X_3 have the same distribution, it is *not* true that the following have the same distribution[15]:

- $2X_1$ (long asset X_1 in twice the size),

- X_1+X_2 (long one unit of each of X_1 and X_2),

- X_2+X_3 (long one unit of each of X_2 and X_3).

Nor, and in fact much more obviously, is it true that these two have the same distribution:

- X_1-X_1 (long and short one unit of asset X_1, which of course exactly cancel each other out),

- X_1-X_2 (long one unit of X_1 short one unit of X_2: these clearly don't hedge each other).

So, it all boils down to this: representing assets by their density or distribution functions isn't a very good idea, because information is lost in doing so.

In fact, this strikes at the heart of copula-based methods, because the joint distribution of (X_1, X_2, X_3) is not easily represented by an off-the-shelf copula. To see why, notice that the correlation between X_1 and X_2 is somewhat unusual, in the following respect (please refer to Figure 3.3):

A. If I know X_2, then I know Z and hence I know X_1. For example, if X_2=7.5, then Z=2.5 and so X_1=6.25.

B. If I know X_1, then I know Z *only up to sign*[16], and hence I know that X_2 must take one of two values. For example, if X_1=6.25 then $Z=\pm2.5$ and X_2 is either almost 0 or else 7.5.

An attempt to represent this using one of the well-known copulas (e.g. Gaussian, Student-t, or Archimedean) is bound to fail. This is because these are essentially 'the wrong shape' (see Figure 3.4). Of course, a copula can always be found that describes their joint distribution, simply by transforming the marginals so that they both become uniform[17]. But to construct such a copula one needs to know the joint distribution in the first place, making the exercise pointless.

---

[15] A rigorous proof is a bit messy, but one can always justify it numerically.
[16] In other words we know the magnitude of Z but not its sign, e.g. it might be 2.3 or –2.3.
[17] One would redefine the X_1= Ga(½;2; Z^2) and X_2= $\Phi(Z)$. Both are uniformly distributed, and their joint distribution is a strange beast as shown in Figure 3 (top picture).

Figure 3.4. (Top) Codependence of X_1 and X_2 in Example 2 (Figure 3.3). Dots represent samples from an 'imaginary Monte Carlo simulation'. (Bottom) Typical form of codependence representable by one of the 'standard' copulas, with positive and negative correlation. Neither of the bottom two graphs looks like the top one, so a standard copula is a poor representation of the codependence. The transfer function approach, on the other hand, captures it in a natural way.
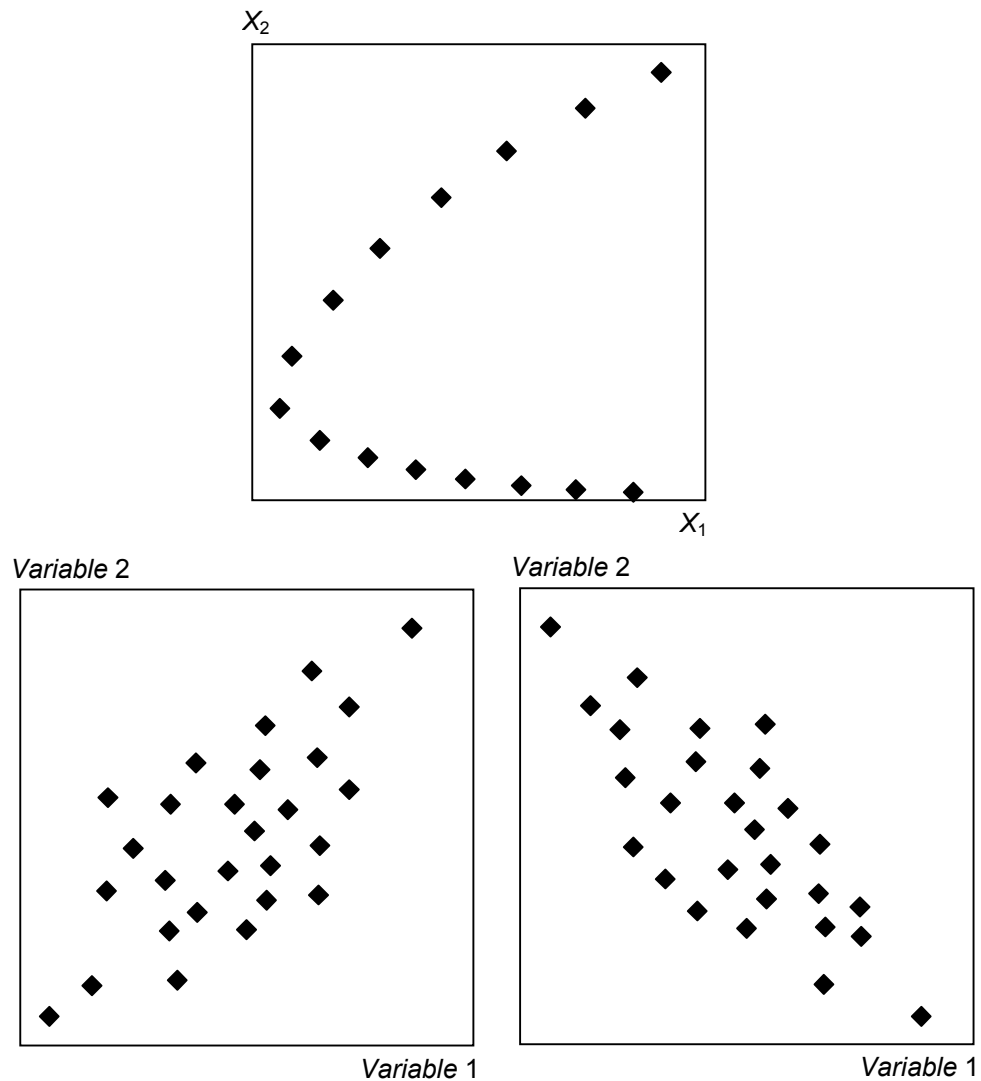
We are not, however, dismissing the copulas as entirely irrelevant. Their proper role is in correlating the explanatory variables together, which is what we discuss next.

Correlation

Another important issue is how to correlate the assets in a sensible way. From what we have seen, the obvious way to do this is to correlate the explanatory variables together. It is at this point that copulas become useful, though even now they are not strictly necessary. In the examples we have considered above, the explanatory variables have always been Gaussian. Then the most natural choice of joint distribution of the explanatory variables is the multivariate Gaussian, and it is completely specified by its covariance matrix. If we wish, we can say the explanatory variables are linked by the Gaussian copula, but this does not add anything to the discussion.

In PR+ this is almost exactly what we do, with an extra step: we represent the correlation in a factor form[18]. The only easily-implementable copulas are the factor ones[19], by which we mean that we identify risk factors conditionally on which the explanatory variables, and hence the asset values, are independent. In the Gaussian framework this is easily done:

$$Z_1 = c_{11}V_1 + \cdots + c_{1m}V_m + \varepsilon_1$$
$$Z_2 = c_{21}V_1 + \cdots + c_{2m}V_m + \varepsilon_2 \quad \text{or} \quad Z_i = c_i \cdot V + \varepsilon_i \text{ for short}$$
$$\cdots$$

where the Z's are the explanatory variables, the c's are the factor weights, the V's are the risk factors (which need not be independent), and the ε's represent the issuer-specific risk. PR+ uses sectorial risk factors, i.e. there is one factor for each industrial sector[20].

A particularly simple demonstration of this in action is the default/no-default model that we talked about earlier. Recall that

$$X = g(Z) = \begin{cases} 1, & Z > \zeta \\ r, & Z < \zeta \end{cases}, \quad \zeta = \Phi^{-1}(p),$$

where ζ is the "threshold of default", p is the default probability (EDF) and r is the recovery rate. Then conditionally on the risk factor V, the conditional default probability is[21]

$$P(Z < \zeta \mid V) = P(\varepsilon < \zeta - c \cdot V) = \Phi\left(\frac{\zeta - c \cdot V}{\sqrt{1 - |c|^2}}\right),$$

the well-known formula; and the expected value of the asset[22] conditionally on the risk factor is

$$E[X \mid V] = r \times \Phi\left(\frac{\zeta - c \cdot V}{\sqrt{1 - |c|^2}}\right) + 1 \times \Phi\left(-\frac{\zeta - c \cdot V}{\sqrt{1 - |c|^2}}\right).$$

---

[18] As where we started out: "The default/no default world, and factor models", Chapter 2.
[19] "Demystifying copulas", Chapter 4.
[20] See Chapters 11 and 12.
[21] The $\sqrt{1 - |c|^2}$ arises because it is the standard deviation of ε; this is because the variance of Z is to be 1.
[22] Assuming that p denotes the market-implied default probability.

When the explanatory variables themselves are non-Gaussian, we do not have an obvious solution to the problem of correlating them. If we wish to use the Gaussian copula, the simplest way of expressing this is to say that the joint distribution of appropriately transformed explanatory variables is multivariate Gaussian. In doing so we go from the simple model

$$X_i = g_i(Z_i)$$
$$Z_i = c_{i1}V_1 + \cdots + c_{im}V_m + \varepsilon_i \tag{$\aleph$}$$

to

$$X_i = g_i(\tilde{Z}_i)$$
$$\tilde{Z}_i = F_i^{-1}(\Phi(Z_i))$$
$$Z_i = c_{i1}V_1 + \cdots + c_{im}V_m + \varepsilon_i$$

where the purpose of the transformation $F_i^{-1}(\Phi(Z_i))$ is to transform a standard Normal variable into one with the same distribution as $\tilde{Z}_i$ (F_i is the cumulative distribution function of $\tilde{Z}_i$). All this can be rolled into one, which returns us to the original model ($\aleph$) with a different transfer function:

$$X_i = g_i(F_i^{-1}(\Phi(Z_i))) \equiv \tilde{g}_i(Z_i)$$
$$Z_i = c_{i1}V_1 + \cdots + c_{im}V_m + \varepsilon_i$$

So having non-Gaussian explanatory variables with a Gaussian copula does not actually constitute a generalization of the basic model ($\aleph$). The model only becomes more general when one moves away from Gaussian copulas.

Conclusions

We have shown that the problem of modeling asset return distributions in a multivariate context is best tackled not by working with the asset distributions themselves, and trying to correlate the assets, but rather by modeling the assets as functions of explanatory variables and correlating the explanatory variables. This is a flexible approach that is easily extended to model other asset classes.

THIS PAGE IS INTENTIONALLY LEFT BLANK

4

Demystifying copulas

So far we have discussed the basics of portfolio modeling and introduced the concepts of systematic and unsystematic risk, bringing in the notion of factor models and conditional independence for modeling these risks. We have stated that this framework greatly facilitates computation of portfolio risk in simple cases, and that it is very flexible, but that it is too flexible for practical use without some 'structural enhancement'. One such structure is the Gaussian factor model, which we discussed while alluding to copulas. In this chapter we discuss copulas in more depth and show that the copula method is not fundamentally different from the conditional independence approach. We also discuss the concept of tail dependence and present some rather counter-intuitive results about its effect on the loss distribution in practice.

Modeling multivariate distributions

To answer questions about the loss distribution of a portfolio we have to model the behavior of the individual assets. Before beginning any analysis, a probability must be assigned to every possible outcome, i.e. every possible combination of asset movements bearing in mind that movements are correlated. In its fullest generality such a specification is large and complicated. As we said in Chapter 2, there are two approaches to solving the problem:

- (Bottom-up) Model each asset individually, identifying the distribution of each. Then work out how to model the correlation, i.e. how to glue the distributions together to obtain a statistical description of the probabilities of all possible combinations of movements of the assets, rather than just an expression for each asset in isolation from the others.

- (Top-down) Identify the factors by which asset movements depend and specify the distribution of each asset conditionally on those factors. Conditional on the factors, the asset values or movements are assumed to be independent, which means that their joint distribution is then easily determined.

Provided that there is only one factor, and that unsystematic risk can be ignored, the top-down approach allows portfolio computations to be easily done "on the back of an envelope" (or, at least, on an Excel spreadsheet). The advantage of this is clear: one can draw conclusions about the loss distribution without requiring any fancy mathematical techniques or Monte Carlo simulation. The reason why the top-down approach works is that portfolio problems are about sums. The top-down approach says

that the portfolio loss of a large portfolio is its expected loss conditionally on the latent factor, and this is obtained *simply by adding the conditional expected losses of the individual assets*. The distribution of loss is then built up by computing the expected loss for each value that the latent factor takes, and then invoking the probability distribution of the latent factor. On the other hand, one has to do quite a lot of work to specify a factor model: one needs to specify the distribution of the factor(s) and the distribution of each asset conditionally on every possible value of these factors. Without any particular structure there is no guidance on how to do this, so the framework is too "floppy" to be useful and requires structure to be imposed. In the previous chapters we introduced the Gaussian model as a way of imposing the needed structure.

We have not discussed the bottom-up approach, beyond saying that it has the attraction of modeling the marginal distribution of each asset (i.e. the distribution of the asset without reference to what the others do), but that one then has the problem of gluing the distributions together. To see why this is not easy, consider the following:

- It is not sufficient to specify correlations between each pair of assets. Worse, specifying "3-wise", "4-wise" and higher correlations is impractical.
- Specifying the joint distribution of pairs of assets is difficult because, in general, there are many ways in which it can be done.

We shall now explore these issues in more depth. For the first point, take as an example a simple default/no-default model. By specifying pairwise correlations we would know the probabilities of pairs of assets defaulting, but we would not know the probabilities of three, four, and so on. Hence the full loss distribution would not be completely specified. But to specify the probabilities of every single outcome (e.g. assets 1, 3, 7, 8, ... default, assets 2, 4, 5, 6, 9, ... do not) would be a huge undertaking. If there are n assets then there are 2^n different outcomes. For $n=100$, this number is about 1,000,000,000,000,000,000,000,000,000,000.

The second point is best explored by way of an example. Starting with two random variables X and Y, we know the distribution of X in isolation from Y, and we know the distribution of Y in isolation from X. How then do we come up with a joint distribution? Let us take the discrete case, where X takes five values (❶-❺) and Y four (❶-❹) with the "marginal" probabilities shown in the table. (Hence the term marginal distribution: the sums are written on the top and left margins of the table.) The problem is how to fill in the grid so that the rows and columns add up to the correct totals (see Fig. 4.1).

Figure 4.1. Correlating two discrete variables. (Top left) the basic problem; (top right) independence; (bottom left) "strong dependence"; (bottom right) other!

			X									X			
			❶	❷	❸	❹	❺				❶	❷	❸	❹	❺
			0.1	0.3	0.3	0.1	0.2				0.1	0.3	0.3	0.1	0.2
Y	❶	0.2							❶	0.2	0.02	0.06	0.06	0.02	0.04
	❷	0.1						Y	❷	0.1	0.01	0.03	0.03	0.01	0.02
	❸	0.3							❸	0.3	0.03	0.09	0.09	0.03	0.06
	❹	0.4							❹	0.4	0.04	0.12	0.12	0.04	0.08

			X									X			
			❶	❷	❸	❹	❺				❶	❷	❸	❹	❺
			0.1	0.3	0.3	0.1	0.2				0.1	0.3	0.3	0.1	0.2
	❶	0.2	0.1	0.1	0.0	0.0	0.0		❶	0.2	0.1	0.0	0.1	0.0	0.0
Y	❷	0.1	0.0	0.1	0.0	0.0	0.0	Y	❷	0.1	0.0	0.0	0.0	0.1	0.0
	❸	0.3	0.0	0.1	0.2	0.0	0.0		❸	0.3	0.0	0.1	0.1	0.0	0.1
	❹	0.4	0.0	0.0	0.1	0.1	0.2		❹	0.4	0.0	0.2	0.1	0.0	0.1

Source: Credit Suisse First Boston

In the top right-hand figure the numbers have been constructed so as to make X and Y independent, and are obtained by multiplying the relevant marginal probabilities. One can recognize independence very easily: each row is just a multiple of the top (marginal) row, and each column is a multiple of the left (marginal) column. So the distribution of X, conditionally on Y, does not depend on what value Y happens to take—and vice versa.

There is only one way to make random variables independent. But there are many ways of making them dependent, as seen for example in the bottom two tables in Fig.1. (In the left-hand table, the probability masses have been arranged on or near the leading diagonal of the table. In the right-hand table the construction is arbitrary.)

This is beginning to look tricky. First, there are inconveniently many specifications for a pair of discretely distributed assets. Even when we have sorted that out, we must tackle larger numbers of assets. Then, we must be able to deal with continuously distributed variables: in a portfolio valuation problem, for example, there is a continuous spectrum of values the portfolio can take, not a countable set of values[23].

---

[23] To elaborate on the problem of continuously distributed variables, we can formulate the problem as follows. Find a bivariate density function g satisfying:

$$\int_{-\infty}^{\infty} g(x_1, x_2) dx_2 = f_1(x_1) \quad \text{for each } x_1$$

$$\int_{-\infty}^{\infty} g(x_1, x_2) dx_1 = f_2(x_2) \quad \text{for each } x_2$$

which are the equivalent of the row and column sums earlier. Put this way, the problem looks hard. In three dimensions it is even more difficult: one must now arrange for

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2, x_3) dx_2 dx_3 = f_1(x_1) \quad \text{for each } x_1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2, x_3) dx_1 dx_3 = f_2(x_2) \quad \text{for each } x_2$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2, x_3) dx_1 dx_2 = f_3(x_3) \quad \text{for each } x_3$$

We now introduce copulas as a solution to the problem.

Copulas

A copula is a method for correlating any two random variables together, irrespective of their marginal distributions. **The idea behind copulas is to separate marginal distribution from correlation.** The "removal" of the marginal distributions is effected by transforming each variable so that it becomes uniformly distributed. For continuous distributions there is a well-known trick for doing this. Let F denote the cumulative distribution function of the random variable X, i.e.

$$F(x) = P(X<x).$$

Then it is easily shown that the variable $U=F(X)$ is uniformly distributed. Let u be a number between 0 and 1; then

$$P(U<u) = P(F(X)<u) = P(X< F^{-1}(u)) = F(F^{-1}(u)) = u$$

Here F^{-1} denotes the transformation that "undoes" F. The statement $P(U<u)=u$ is equivalent to saying that U follows a uniform distribution. This trick is more often used the other way round: given F, we can generate observations that have the same distribution as X by simulating uniform random variables[24] and then transforming them using F^{-1}.

Consequently, if we are able to glue together (i.e. correlate) uniform distributions, then we have solved the problem. In fact we have arrived at the definition of a copula:

> **A copula is the distribution function of a multivariate random variable with uniform marginals.**

This is not the mathematical definition, but it is close enough. Symbolically the copula is defined by C where

$$C(u_1, ..., u_n) = P(U_1<u_1, ..., U_n<u_n).$$

At this point one might reasonably frown and say, "Well, this is not a solution to the problem, is it? All you've done is thought of a new word and then use that for the answer." There is some truth in this. In fact, much of the theory behind copulas boils down to using mathematical trickery to come up with new copulas, which can then be labelled and placed in a display cabinet like rare butterflies. That said, there are some important results, one of which is the concept of tail dependence, which we shall come to later. It is also worth bearing in mind that we have considerably reduced the problem by transforming each random variable so that its marginal distribution is uniform.

---

which looks even harder.

[24] e.g. RAND() in MS Excel. So to simulate Normal variables, use NORMSINV(RAND()).

Examples of copulas

Independence copula

This is the easiest one to deal with. There is only one way of correctly specifying the probability of multiple events if they are independent: that is, multiply the probabilities together. Hence

$$C(u_1, ..., u_n) = P(U_1 < u_1, ..., U_n < u_n) = P(U_1 < u_1) \, P(U_2 < u_2) \, ... \, P(U_n < u_n) = u_1 \, u_2 \, ... \, u_n.$$

Gaussian copula

We introduced the Gaussian copula in Chapter 2 (but derived it through a different and rather more natural route, without the need for copulas). Having transformed our random variables so that their marginal distributions are uniformly distributed, we now transform them again so that they become Normally distributed using the inverse Normal probability function. The joint distribution of these Normal variables is then assumed to be multivariate Normal, with a given correlation matrix Σ. Writing Φ_Σ for the probability $P(Z_1 < z_1, ..., Z_n < z_n)$, where the Z's are multivariate Normal with correlation matrix Σ, we have

$$C(u_1, ..., u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)) \, .$$

We can apply this to a default/no-default model as follows. Assign to the i'th obligor a uniformly distributed random variable U_i. Default occurs when $U_i < p_i$, where p_i is the (marginal) default probability of that asset ($P(U_i < p_i)$ is just p_i because U_i is uniformly distributed). The probability of any combination of defaulting events can then be obtained from the copula. For example, the probability that assets 1,2,5 default in a portfolio of size 7 is

$$C(p_1, p_2, 1, 1, p_5, 1, 1) \tag{*}$$

(Note that this is the probability of those three assets defaulting, not the probability that *only* those assets default: the other four may or may not.)

One thing worth noting about the Gaussian copula is that it is uniquely specified by pairwise correlations. We said previously that to specify a joint distribution one had to specify the higher correlations as well. The point is that the Gaussian copula "fills in" this remaining structure from the pairwise correlations. So the number of parameters needed to specify the full distribution, rather than growing exponentially with the number of assets (2^n in our previous discussion), instead grows quadratically ($n(n-1)/2$ correlation numbers and n default probabilities).

There is a problem, however. Although we have written down a neat expression (*) for a multiple default probability, it is not obvious how we calculate it. This lack of transparency is due to the fact that calculation of Φ_Σ requires a multidimensional integral to be evaluated:

$$\Phi_\Sigma(z_1, ..., z_n) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} \frac{\exp\left(-\frac{1}{2}\sum_{i,j=1}^{n} z_i (\Sigma^{-1})_{ij} \, z_j\right)}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \, dz_1 \cdots dz_n$$

If Σ is reasonably large and has no particular structure then this is very hard and one has to resort to Monte Carlo simulation, which is not ideal. Are there any structures that do allow efficient computation?

It turns out that we have already found the answer to this question through the use of factor models, which was the way we derived the Gaussian model. Suppose that the correlation matrix Σ can be expressed in factor form:

$$\Sigma_{ij} = E[Z_i Z_j]$$

$$Z_i = c_{i1}V_1 + \cdots + c_{im}V_m + \sqrt{1 - (c_{i1}^2 + \cdots + c_{im}^2)}U_i$$

Here we have transformed the uniform variables U_i to Gaussians Z_i, i.e. $Z_i = \Phi^{-1}(U_i)$, and re-expressed the correlation matrix in terms of the Z's. (This puts us on the same course as we were on in Chapter 2[25].) Then we have performed factor analysis (principal components analysis) on the Z's in the usual way, identifying a set of independent factor variables (V's) and residuals (U's) that are also Normally distributed. For the one-factor model the correlation matrix is given as

$$\Sigma_{ij} = c_i c_j \quad (i \neq j).$$

To evaluate the required multivariate probabilities we invoke the "conditioning trick": conditionally on the V's the Z's are independent, and then all we have to do is integrate out the V's, which involves a much lower dimension of integration. For a one-factor model only a single integral has to be done, regardless of the size of the portfolio. Of course this turns a virtually intractable problem into a simple one.

We have arrived at the conclusion that in the Gaussian case the copula framework does not give us anything that the factor approach doesn't. In fact all "workable" copulas have this property.

Student-t copula

The Student-t copula is best understood as a generalization of the Gaussian copula, so we assume that we already have the Gaussian model set up: $Z_i \sim N(0, \Sigma)$ with Σ denoting the correlation matrix. To get from this to a Student-t model we allow the volatility (standard deviation) of Z, which at present is 1, to be stochastic. To be precise, the reciprocal of the variance is assumed to follow a chi-squared distribution with ν degrees of freedom, divided by ν. This recipe makes the variable $\tilde{Z}_i = Z_i / (\sqrt{W/\nu})$. Student-t distributed with ν degrees of freedom. In the limit $\nu \to \infty$ we end up with the Gaussian model again, because the volatility becomes deterministic. The effect of the variable W, which is chi-squared distributed, is to introduce stochastic volatility and it is fairly clear that this has an appealing financial interpretation. Asset returns tend to exhibit 'bursty' behavior in which there are periods where returns are highly volatile, and other periods where the volatility is low. The variable W affects all asset returns equally by scaling their volatilities. To obtain an expression for the copula, we take uniformly distributed variables and transform them so that they become Student-t distributed:

---

[25] The reason the Gaussian copula came up is that it is natural to assume the existence of "asset returns" that are Normally distributed, with default events occurring when the asset return is below a certain threshold. The natural choice of multivariate distribution for these asset returns is multivariate Normal. The copula approach is less elegant because it involves going via a uniform distribution, which is an extra step.

$$C(u_1,\ldots,u_n) = P(U_1 < u_1,\ldots,U_n < u_n) = P(\tilde{Z}_1 < T_\nu^{-1}(u_1),\ldots,\tilde{Z}_n < T_\nu^{-1}(u_n))$$
$$= T_{\nu,\Sigma}\left(T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_n)\right)$$

with T_ν denoting the cumulative distribution function of the univariate Student-t distribution and $T_{\nu,\Sigma}$ denoting the cumulative distribution function of the multivariate Student-t distribution. Calculation of this latter function is not always straightforward, and it may be easier to do it from the multivariate Normal instead:

$$P(\tilde{Z}_1 < T_\nu^{-1}(u_1),\ldots,\tilde{Z}_n < T_\nu^{-1}(u_n)) = P(Z_1 < \sqrt{W/\nu}\,T_\nu^{-1}(u_1),\ldots,Z_n < \sqrt{W/\nu}\,T_\nu^{-1}(u_n))$$
$$= \mathrm{E}^W\left[\Phi_\Sigma\left(\sqrt{W/\nu}\,T_\nu^{-1}(u_1),\ldots,\sqrt{W/\nu}\,T_\nu^{-1}(u_n)\right)\right]$$

with the last line being done by numerical integration (over the distribution of W).

We can also write the Student-t model in factor form. Here we must remember that there are two factors: the common part(s) of the asset returns (previously called V), and the scaling variable W. Thus

$$Z_i = c_{i1}V_1 + \cdots + c_{im}V_m + \sqrt{1 - (c_{i1}^2 + \cdots + c_{im}^2)}\,U_i$$

$$P(\tilde{Z}_i < \xi_i \mid V, W) = P(Z_i < \sqrt{W/\nu}\,\xi_i \mid V, W) = \Phi\left(\frac{\sqrt{W/\nu}\,\xi_i - (c_{i1}V_1 + \cdots + c_{im}V_m)}{\sqrt{1 - (c_{i1}^2 + \cdots + c_{im}^2)}}\right)$$

We can then make this into a default/no-default model by thresholding: default occurs when $\tilde{Z}_i < \xi_i$, with $\xi_i = T_\nu^{-1}(p_i)$.

There are two 'advantages' that the Student-t model has over the Gaussian:

- The Student-t copula exhibits tail dependence, which we shall discuss shortly.

- If we are modeling a phenomenon that closely relates to 'asset returns' (credit risk being a good example) then the Student-t model is a natural one to choose because it incorporates stochastic volatility, which is seen in real life.

In arguing these we have been careful to distinguish a copula from a model. The second point is somewhat delicate and we shall return to it later (see 'Copulas vs. fundamental models').

It is also worth noting a disadvantage of the Student-t approach, which is the extra level of computational complexity.

Archimedean copulas

Archimedean copulas are at the moment regarded as more of a mathematical curiosity and the Gaussian copula has become a *de facto* standard. However, they do offer a completely different approach to modeling correlation and are worth discussing for that reason alone.

Their construction arises directly from a factor model. In other words one specifies the factor variable and the distribution of each asset conditionally on it. That means that the Archimedean copulas fit very easily into the top-down approach, which we discussed at the outset.

The construction involves Laplace transforms. Let V be a strictly positive random variable (always takes values >0) and define the function

$$\varphi(s) = E[\exp(-sV)] \quad \text{(the Laplace transform)}.$$

It is not difficult to show that, regardless of the distribution of V, the function φ decreases as s increases, with $\varphi(0)=1$ and $\varphi(+\infty)=0$. Then we define the conditional distribution of an (unconditionally) uniform variable U to be

$$P(U < u \mid V) = \exp(-\varphi^{-1}(u)V)$$

To show that this prescription does indeed give rise to uniform marginals, we compute

$$P(U < u) = E[\exp(-\varphi^{-1}(u)V)] = \varphi(\varphi^{-1}(u)) = u$$

which gives us what we want. By conditional independence we have the full distribution conditionally on V:

$$P(U_1 < u_1, \ldots, U_n < u_n \mid V) = \exp(-(\varphi^{-1}(u_1) + \cdots + \varphi^{-1}(u_n))V)$$

and taking the expectation over V gives

$$P(U_1 < u_1, \ldots, U_n < u_n) = \varphi\big(\varphi^{-1}(u_1) + \cdots + \varphi^{-1}(u_n)\big).$$

It is convenient if φ^{-1} can be computed in closed form[26], though not essential if φ is known, as root-finding can always be used. Here are some examples.

Name	Distribution of V	Param	Laplace transform, φ	Inverse function
Clayton	Gamma, $e^{-x} x^{1/\theta - 1} / \Gamma(1/\theta)$	$\theta > 0$	$\varphi(s) = (1+s)^{-1/\theta}$	$\varphi^{-1}(u) = u^{-\theta} - 1$
Gumbel	No closed form	$\theta > 1$	$\varphi(s) = \exp(-s^{-1/\theta})$	$\varphi^{-1}(u) = (-\ln u)^{\theta}$
Frank	Geometric; for k=1,2,... $P(V=k) \propto (1-e^{-\theta})^k / k$	$\theta > 0$	$\varphi(s) = -\dfrac{1}{\theta} \ln[1 - e^{-s}(1-e$	$\varphi^{-1}(u) = -\ln \dfrac{e^{-\theta u} - 1}{e^{-\theta} - 1}$

Source: Credit Suisse First Boston

It is quite easy to relate this construction to default/no-default models. Each obligor is assigned a uniform random variable U_i and default occurs when $U_i < p_i$, with p_i being as usual the average default rate (marginal default probability). We can now use conditional independence: the conditional default probability is

$$P(U_i < p_i \mid V) = \exp(-\varphi^{-1}(p_i)V).$$

---

[26] Note: φ^{-1} is **not** the "inverse Laplace transform".

Looking back over this construction, where is the correlation coefficient? What distinguishes the Archimedean copulas from the Gaussian and Student-t is that in Archimedean copulas the correlation, and hence the full distribution, is specified by the distribution of the latent variable. This is therefore quite a strange construction. In particular, it means that one cannot have some assets more strongly correlated than others. Consequently one cannot make assets have the same mean default rate but different default rate volatilities. In other words it is a sort of "one size fits all" setup. To an extent this can be relaxed by introducing more factor variables, but the constructions are rather difficult. This has made them less attractive than the standard Gaussian copula, which uses a correlation matrix and therefore has a certain familiarity.

Properties of copulas

In this section we are going to talk about general properties of copulas and discuss implementation and modeling issues.

Factor representation

We taken pains to emphasize that copulas are best understood in the conditional independence framework. This means that we specify the conditional distribution given the latent factor(s) of a uniform variable U that unconditionally has a uniform distribution. The construction makes sure that the unconditional distribution of U is indeed uniform.

Copula	Factor(s)	$P(U < u \mid factors)$
Gaussian	*V*, **Gaussian (possibly vector)**	$\Phi\left(\dfrac{\Phi^{-1}(u) - c \cdot V}{\sqrt{1 - \mid c \mid^2}} \right)$
Student-t	*V*, **Gaussian (possibly vector)** *W*, **chi-squared**	$\Phi\left(\dfrac{\sqrt{W/\nu}\,T_\nu^{-1}(u) - c \cdot V}{\sqrt{1 - \mid c \mid^2}} \right)$
(Any) Archimedean	*V*, **positive with Laplace transform φ**	$\exp(-\varphi^{-1}(u)V)$

Source: Credit Suisse First Boston

To apply these in a default/no-default context, simply substitute the average default rate (unconditional default probability) for u; the expression for "$P(U < u \mid factors)$" is now simply the conditional default probability.

Implementation of a Monte Carlo simulator is now very easy:

1. Draw the random variable V having the right distribution.

2. For each asset in the portfolio, decide whether it defaults or not:

- compute the conditional default probability

- draw a uniform random variable; default occurs if it is lower than the conditional default probability. (As losses are conditionally independent, the draws are also independent and can be done sequentially.)

3. Add up the losses and record the result (portfolio loss).

4. Repeat many times!

Archimedean copulas are asymmetrical

We said that to model defaults we can just threshold a uniform variable, by making default occur when *U<p*. However we could set it up differently, so that default occurs when *U>1−p*. Either way the unconditional default probability will still be *p*. For the Gaussian and Student-t copulas there is a symmetry, which in the notation we have been using can be written *Z→−Z* or *U→1−U*. So it does not matter whether the threshold of default is placed in the upper or lower tail. For Archimedean copulas this is not so and one obtains a different model by using the upper tail from the lower tail. In particular this means that the coefficients of upper and lower tail dependence are different (see next).

Tail dependence

Tail dependence measures how likely it is that extreme (tail) events occur together. This is distinct from correlation, which refers to the whole distribution. The formal definition of the coefficient of tail dependence is

$$h_- = \lim_{p \to 0} P(Y < y_p \mid X < x_p) \quad \text{(lower tail dependence)}$$

$$h_+ = \lim_{p \to 1} P(Y > y_p \mid X > x_p) \quad \text{(upper tail dependence)}.$$

Here x_p denotes the lower *p*-quantile of *X*: that is, $P(X<x_p)=p$. Similarly $P(Y<y_p)=p$.

Suppose first that *X* and *Y* are "perfectly correlated", in the sense that *X* and *Y* "move up and down together". (More formally they are said to be comonotonic.) Then they take extreme values together, so whenever *X* is less than x_p, *Y* is also less than y_p. Hence $h_-=1$, and by the same argument $h_+=1$ also. (See top diagram of Figure 4.2a.)

When *X* and *Y* are independent the coefficients of tail dependence are zero. This is because $P(Y < y_p \mid X < x_p) = P(Y < y_p) = p$ which vanishes as *p*→0. (See Figure 4.2b.)

If *X* and *Y* are correlated then there may or may not be tail dependence. The next two figures show the difference. In the third, the extreme events do not seem to be clustered together and the tail dependence is zero or close to zero, whereas in the last figure they do seem to be clustered. For example if *p*=0.01 and the proportion of points in the bottom-left quadrant is 0.0004, then there is dependence (if not, the proportion would be 0.01²=0.0001), but the probability of *X* and *Y* both taking extreme values given that one does is 0.0004/0.01=0.04. The coefficient of tail dependence is taken as the limit of this quantity in the limit *p*→0, so although no precise deduction can be made on the basis of this information, it seems reasonable to suppose that there is very little evidence of tail dependence. But if the proportion of points in the bottom left-hand quadrant is 0.005, then now the probability of two extreme events given one is 0.005/0.01=0.5, so there is evidence of tail dependence.

By construction, this measure of tail dependence does not depend on the marginal distributions of *X* and *Y*, and is instead a property only of the copula. This means that it can be expressed in terms of the copula function *C*:

$$h_- = \lim_{u \to 0} \frac{C(u,u)}{u} \quad \text{and} \quad h_+ = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1 - u}.$$

It is also important to understand that **it is possible for variables to be correlated but not have tail dependence**. The most celebrated example is the Gaussian copula. To give an illustration, let the copula correlation be 30% (recalling the construction of the Gaussian copula, each variable is associated with a Normally distributed variable, and these are then correlated; we are saying that the correlation between them is 30%). Let p=0.01; then the probability of two extreme events is

$$\Phi_2(-2.326, -2.326; 0.3) = 0.000556$$

(this computation requires the bivariate Normal integral, for which standard routines are available). So the probability of two extreme events given one is quite low (0.000556/0.01 = 0.0556) and in fact this quantity will vanish as p is made smaller. This effect is seen regardless of the correlation coefficient (unless it is 1), so even with 90% correlation there is no tail dependence.

The Student-t copula has positive tail dependence whenever the correlation is positive. The coefficients of upper and lower tail dependence are equal, by symmetry.

The Clayton copula has positive lower tail dependence ($2^{-1/\theta}$) but no upper tail dependence.

The Gumbel copula has positive upper tail dependence ($2-2^{-1/\theta}$), but no lower tail dependence.

Figure 4.2(a) *X* and *Y* comonotonic. Tail dependence = 1
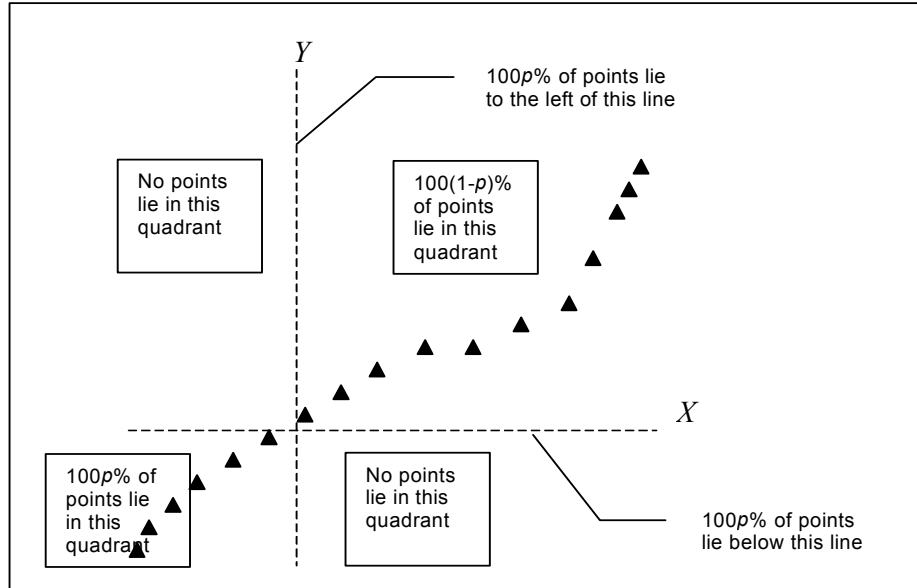
Source: Credit Suisse First Boston

Figure 4.2(b) *X* and *Y* independent. Tail dependence = 0
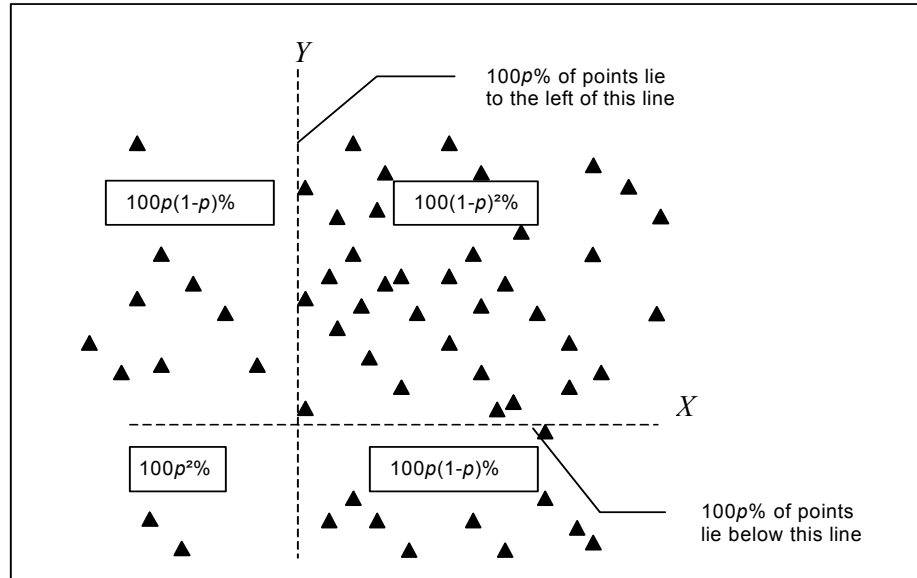
Source: Credit Suisse First Boston

Figure 4.2(c) *X* and *Y* correlated but with little or no tail dependence
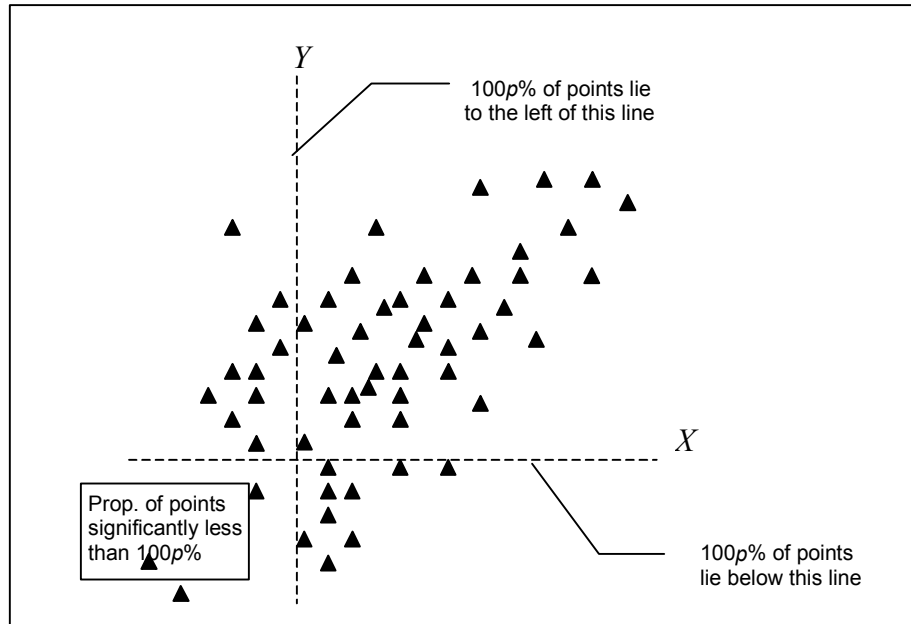
Source: Credit Suisse First Boston

Figure 4.2(d) Positive tail dependence
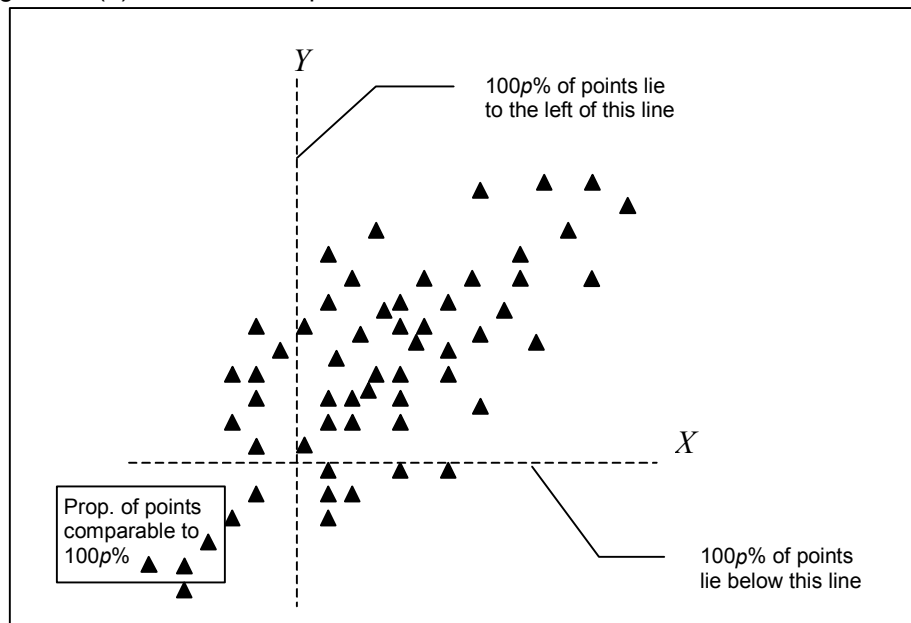
Source: Credit Suisse First Boston

Copulas vs. fundamental models

We said a little while ago that the Student-t *model*, rather than the *copula*, has an advantage over the Gaussian because, at least in situations where asset returns are involved, the Student-t distribution is more realistic. In saying this we have to be careful.

Remember that the copula approach starts by separating the marginal distribution from the correlation structure. So if we observe that the marginal distributions of asset returns happen to be Student-t distributed rather than Normally distributed, a purist would argue that we cannot say that that justifies the use of a Student-t copula—the choice of copula has nothing to do with the marginal distributions. Hence the purist would argue that we can model the asset returns marginally as being Student-t distributed, and then correlate them using a Gaussian copula, or Clayton copula, or whatever. But by doing that the purist is not attempting to come up with a model for what is causing the observed behavior. Let us argue as follows. A natural model for asset returns is that they are Normally distributed[27], and the multivariate analogue of it gives a multivariate Normal model. In observing that returns are non-Normal, and in seeing that volatility appears to be stochastic[28], we are led to modify the basic model. One way to do this is to scale the volatility in a random way: if a reciprocal-Gamma distribution is used for the variance then we end up with a Student-t model. But doing this affects both the marginals and the correlation, so not only do the marginals become Student-t, but so does the multivariate distribution, so that we naturally end up with a Student-t copula.

So an approach that involves fundamental modeling is most unlikely to treat the marginal distribution separately from the correlation structure. The purist might be right at a purely mathematical level, but he is not producing a natural model. The unnatural aspect of the copula approach is the distortion of each component to turn its marginal distribution into a uniform distribution. There is unlikely to be any physical mechanism that would effect such a change.

Portfolio analysis with different copulas

Owing to the relative inflexibility of the Archimedean copulas, it is fairly difficult to construct realistic examples of how they may be compared with the Gaussian copula "on an equal footing". However, after a mainly theoretical exposition we need to give some demonstration. We therefore take a very simple test case[29] in which we assume that the portfolio is large and homogeneous, consisting of defaultable bonds with default probability 5%. We also set the pairwise default probability to be 0.00725, which corresponds to a default event correlation of 10%. This allows us to see what the various copulas have to say about the higher-order correlations, and therefore the probabilities of multiple events occurring. From the point of view of the loss distribution, the various copulas will produce loss distributions with the same mean and variance, but will differ in their tail shapes. In particular, Value-at-Risk (VaR) will not be the same for them all. This implies the following calibration:

---

[27] Namely that each asset experiences a large number of independent shocks very rapidly, and by the Central Limit Theorem the distribution of the sum of those shocks is roughly Normal. We shall soon find out that the CLT does not work very well in practice, but that is another matter!
[28] The basis of GARCH modeling (Engle et al.)
[29] This test is identical to one carried out by Philipp Schonbucher ("Taken to the limit: simple and not-so-simple loan loss distributions", Working paper, Bonn University, 2002), though we do the Student-t as well.

Gaussian: pairwise default probability is $C(p,p) = \Phi_2(\Phi^{-1}(p), \Phi^{-1}(p); \rho)$ which requires $\rho = 30.55\%$ (this must be obtained numerically via the bivariate Normal integral).

- Student-t with $\nu = 4$ degrees of freedom: pairwise default probability is $C(p,p) = E\left[\Phi_2(\sqrt{W/\nu}T_\nu^{-1}(p), \sqrt{W/\nu}T_\nu^{-1}(p); \rho)\right]$, where W follows a Gamma distribution with mean ν and variance 2ν. This has to be evaluated by numerical integration, integrating the bivariate Normal integral over a Gamma distribution[30]. For $\nu = 4$, we require $\rho = 5.91\%$. Note how much lower the correlation is in this example as it was in the Gaussian copula, where it had to be 30%. This can be attributed to the high level of tail dependence in the Student t_4-copula.

- Clayton: pairwise default probability is $C(p,p) = (1 + 2(p^{-\theta} - 1))^{-1/\theta}$ which requires $\theta = 0.1817$ (this must be obtained numerically).

- Gumbel: pairwise default probability is $C(p,p) = p^{\wedge}(2^{1/\theta})$ which requires $\theta = 1.393$ (this can be obtained directly).

Now that we have calibrated the models, it is a question of finding the loss distribution of a large portfolio. As we are considering a homogeneous portfolio it is sufficient to work out the distribution of the proportion of defaults in the portfolio. If there were no correlation, the proportion of defaults would always be 5%, because that is the mean default rate we are using.

In Chapter 2 we used the fact that the portfolio loss is just the conditional expected loss given the risk factor, and is therefore a simple transformation of it. The transformation is given for each type of copula in the table of factor representations earlier. For the Gaussian and Clayton copulas everything is easy. The Gumbel presents a computational difficulty because the factor does not have a closed form expression for its density or tail probability[31]. The Student-t causes a different type of problem because it is a two-factor model: it is not possible to ascribe a particular level of loss to one particular value of the latent factor: there are many different combinations of factor values that give rise to the same conditional expected loss. This problem can be tackled by calculating the loss distribution for many different values of W and integrating, in the same way that the other Student-t calculations have been done.

The results are shown in Figure 4.3. What is very odd at first sight is that the copulas with tail dependence give lower VaR at (say) 99%-99.9% confidence (excluding the Gumbel, which gives a rather strange-looking distribution). Indeed at 99.5% the Gaussian gives the highest VaR, and it has no tail dependence! The main reason for this is the comparison in Figure 4.3 assumes equal *pairwise default probability* from model to model. For the Gaussian copula, there is no tail dependence, so to get a high level of pairwise default probability it is necessary to put in a high correlation parameter; but that causes the probability of large numbers of default events to significantly increase too. Had Student t and Gaussian been compared with equal correlation parameters, the Student t would obviously have had a fatter tail. The subject is quite a trappy one.

---

[30] TINV in MS Excel takes the two-tail probability as its argument, so to find $T_4^{-1}(0.05)$ we need to do
x = -TINV(0.10) (minus sign because the lower tail is required). The integration was done by 20-point Gauss-Laguerre quadrature.
[31] These must be calculated by inversion of the Laplace transform, which is done by the Fast Fourier transform algorithm.

Figure 4.3. Loan Portfolio Tests for Four Different Copulas
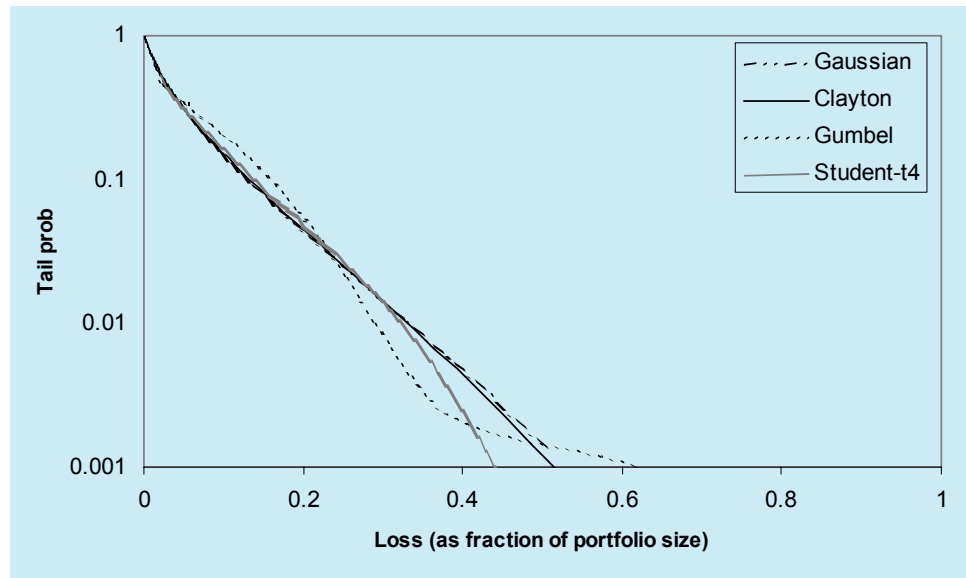
Conclusions

We have discussed copulas, related them to conditional independence models and discussed tail dependence in some detail, showing by way of one example that the effects of introducing a tail-dependent copula to model the correlation may be hard to "predict".

In the next chapter we shall talk about unsystematic risk and how to analyze concentration risk in portfolios.

5

Thinking unsystematically

In the previous chapters we have discussed the modeling of systematic risk and shown how the various approaches fit into one and the same framework. We have shown that if a one-factor model is used, and unsystematic risk can be ignored, then the portfolio loss is a simple transformation of the factor variable: this allows its distribution to be easily calculated. In reality, however, unsystematic risk cannot be ignored, as discussed in the following pages.

Résumé

In Chapter 2 we talked around Figure 5.1a that illustrates the loss distribution of a credit portfolio using the following simple model:

- One discrete risk factor is needed to model systematic risk, and it takes five discrete values ❶❷❸❹❺ (see top of diagram);

- At the time horizon in question, each obligor is either in default or not, and the loss per unit exposure is 1 or 0 respectively, with the result that asset distributions are completely specified by the conditional default probabilities (next row);

- Conditional on the factor the defaults are independent.

Figure 5.1a shows the conditional loss distributions of some of the assets and the loss distributions of three different portfolios. We have argued that for a large enough portfolio the loss is entirely driven by the value of the latent factor, and is equal to the conditional expected loss (first portfolio).

In the next two pictures we see the situation for increasing amounts of unsystematic risk. Now there is some uncertainty in portfolio loss even when we know the state of the world. So each of the spikes in the infinitely fine-grained case becomes fattened, though its mean position is the same. The objective of the next few chapters will be to show how each of these conditional distributions can be analyzed, or, more loosely, to quantify the "fattening".

Figure 5.1a. Systematic and unsystematic risk, default/no-default model with discrete latent factor
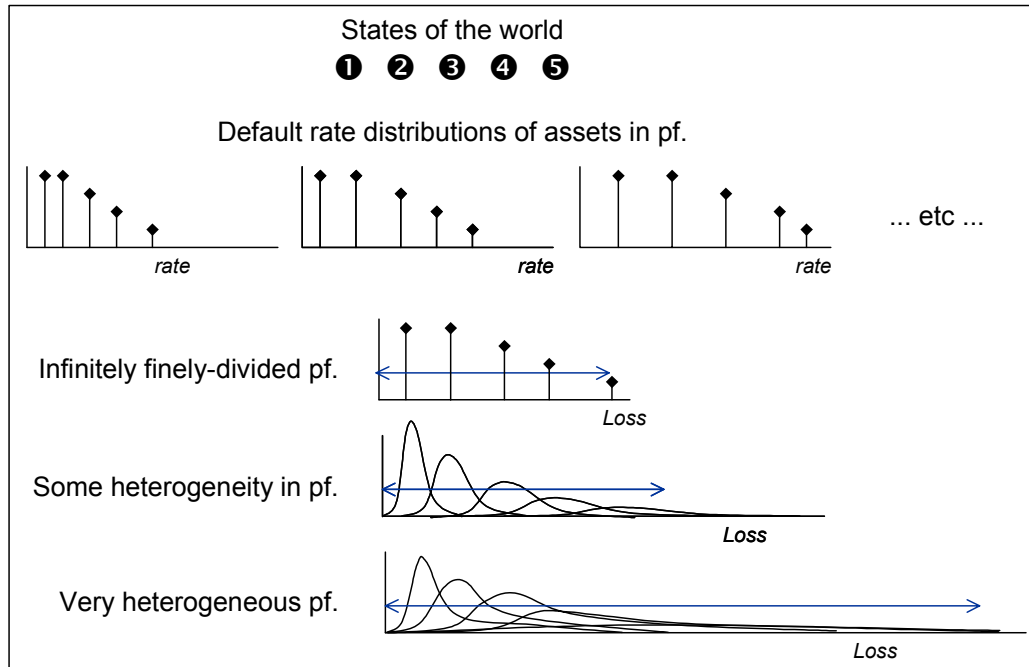
Source: CSFB

Figure 5.1b. Systematic and unsystematic risk, general case for continuous latent factor
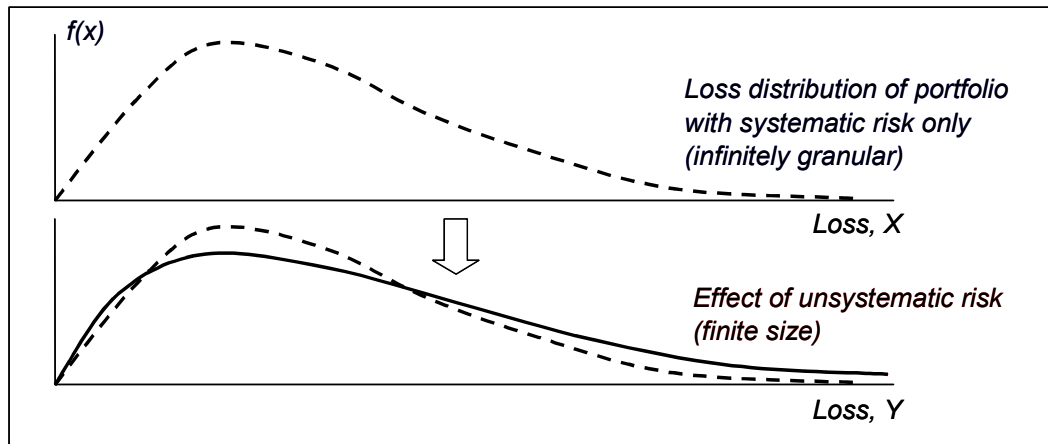
Source: CSFB

To take a slightly different view of the same problem, Figure 5.1b shows the picture when the distribution of the factor variable is continuous; models of this sort are more common than the discrete-variable ones, but they are not fundamentally different. (Recall the example of the Gaussian copula, or quasi-Merton model, which we considered before: the underlying factor is Normally distributed (possibly multivariate) and the expected loss of any asset conditional on the factor is given by a transformation involving the cumulative Normal distribution function, representing the probability of a firm's assets being insufficient to repay its debt at maturity.)

In the top picture we have the distribution of the infinitely fine-grained portfolio, which we have already shown how to obtain[32]. In the lower picture we have replotted the infinitely granular distribution (dashed) and shown as a solid line the distribution of a real portfolio that does have unsystematic risk. The tail of the distribution is fattened (there is a greater chance of big losses), however, the means of the two distributions are the same: unsystematic risk does not alter the expected loss.

As with the lower two graphs in Figure 5.1a, we have not described how to construct the solid line in Figure 5.1b. To do so requires several techniques. First, we must learn how to analyze the unsystematic risk distribution, i.e. the extra risk that remains even when we have conditioned on the factor variable; then, we must work out how to combine this with the systematic risk model. In this chapter, we begin on the first of these tasks, avoiding a mathematical construction, but illustrating with some examples and showing how mean and variance may be calculated.

Independent random variables, pictorially

In the examples below, we shall investigate the subdivision of a position in one asset into smaller positions in similarly distributed independent assets: commonly known as diversification. In the first two examples, which were chosen for ease of analysis, we show that, as the degree of subdivision is increased, the risk reduces to zero and the distribution contracts around the mean loss. However, there is a surprise in store, in the shape of the third example.

Binomial distribution

Figure 5.2a shows the picture for a Binomial distribution. This would be the situation for a default/no-default model of a loan portfolio. For the sake of argument, we set the default probability to 5% and, for different portfolio sizes, plot the distribution of the proportion of defaults in the portfolio[33]. For increasing portfolio size, the tail of the distribution reduces and the distribution becomes more symmetrical, resembling a Normal distribution by the time the size reaches about 200. In fact, it can be shown that the limiting distribution is indeed Normal in this case. By the time the size exceeds a few thousand, there is little uncertainty in the loss rate of the portfolio, which tends towards 5%. In Chapter 2, we said that it was exactly this property that allowed us to define the term default probability (default rate) in the first place, viz. the expected proportion of losses in a large portfolio of "similar" assets.

---

[32] In the one-factor case only it is considerably more difficult in multifactor models (i.e. where the factor is multivariate), though in Chapter 12 we mentioned the Student-*t* model, where there is more than one factor.
[33] In the top picture no probability is plotted for proportions of 0.01, 0.02, 0.03, 0.05, ..., and similarly in the second none is plotted for a proportion of 0.01, 0.03, ... This is because (taking the top example) in a portfolio of 25 assets the number of defaults must be a whole number, and so the proportion must be a multiple of 0.04.

Gamma distribution

Figure 5.2b shows the picture for a Gamma distribution. The basic asset is assumed to have a loss that follows a Gamma distribution with mean 1 and variance 10 (shape parameter 0.1 and size parameter 10: see note[34]). The loss distribution from one asset is extremely long-tailed (not shown). This sort of distribution would be a reasonable model for the loss in value of a bond at some particular horizon: the value cannot increase very much, but it can downgrade all the way to zero, so the losses are more extreme than the gains. In the figure, the results are shown for portfolios of 10 assets of 1/10 the size, 50 assets of 1/50 the size, and so on. Again the loss distribution starts to become more Normal-looking and contracts around the mean loss, which we have maintained at 1. In fact this example is fairly easy to construct because each portfolio has a Gamma distribution. More precisely, the sum of independent Gamma-distributed variables, that have the same size parameter, is also Gamma distributed[35]. Again, it can also be shown that the limiting distribution is indeed Normal in this case.

Does this always work?

More formally we are asking the following.

Suppose that (X_i) follow some distribution and are independent; does the distribution of $\frac{1}{n}(X_1 + \cdots + X_n)$, that is, the average of *n* of them, converge around its mean as $n \to \infty$?

The answer to this is: Not always. Some distributions are so fat-tailed that this subdivision, or averaging, process, fails to attenuate the extreme moves. The simplest example is the Cauchy distribution, which is in the family of *t*-distributions and also in the family of α-stable distributions. Its density function is $1/[\pi(1 + x^2)]$. Remarkably, averaging Cauchy distributions has no effect at all! Regardless of *n*, the distribution is exactly the same[36]. See Figure 5.2b. If in the financial world everything was Cauchy distributed, there would be no point in trying to diversify!

---

[34] The shape (α) and size (β) parameters of a Gamma distribution are related to the mean (μ) and variance (σ²) by μ=αβ and σ²=αβ². The (α,β) parametrization is used in Excel.
[35] If the shape and size parameters for X_1 are α_1 and β then those for λX_1 are $\lambda\alpha_1$ and β. Now let X_2 also be Gamma distributed, with shape α_2 and size β. Then the shape and size parameters for (X_1+X_2) are $(\alpha_1+\alpha_2)$ and β. Combining these results, we find the parameters for $(X_1+X_2)/2$ are $(\alpha_1+\alpha_2)$ and β/2. The same result holds however many variables we add or average, provided that the size parameters are all the same. Most statistical distributions do not have an additive property like this.
[36] In essence, this is what the α-stable property is.

*Figure 5.2a. Proportion of defaults in homogeneous portfolios of "loans" of default probability 0.05.
The portfolio sizes are 25, 50, 200, 1000, 5000. Notice the "contraction" of the distribution around
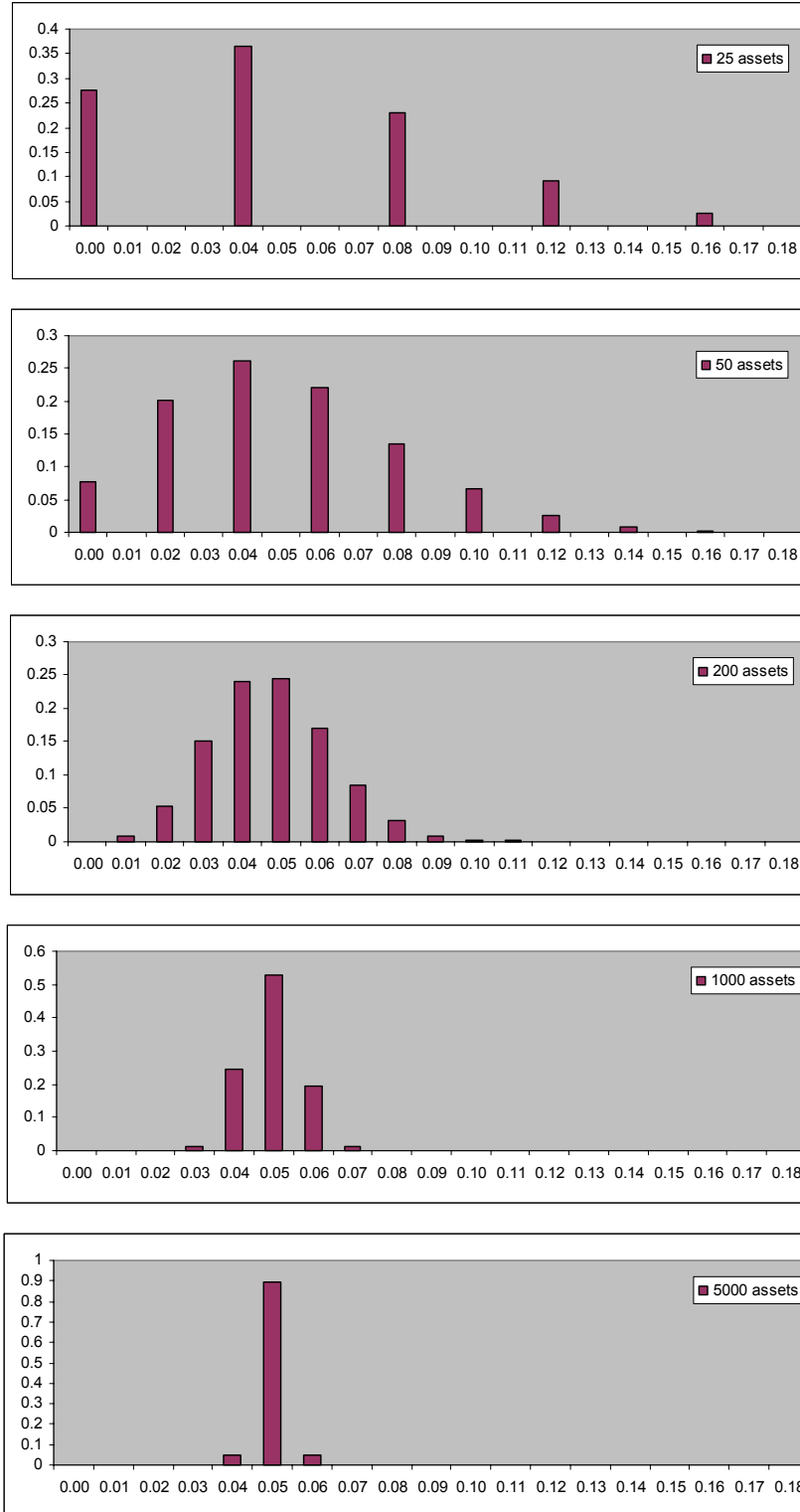the mean (0.05) as the portfolio size increases.*

Figure 5.2b. Loss density function of portfolios of Gamma-distributed assets, with increasing size (10, 50, 200, 1000, 5000) and hence increasing degrees of averaging (granulation). As before, notice the contraction of the distribution around the mean loss (1.0).
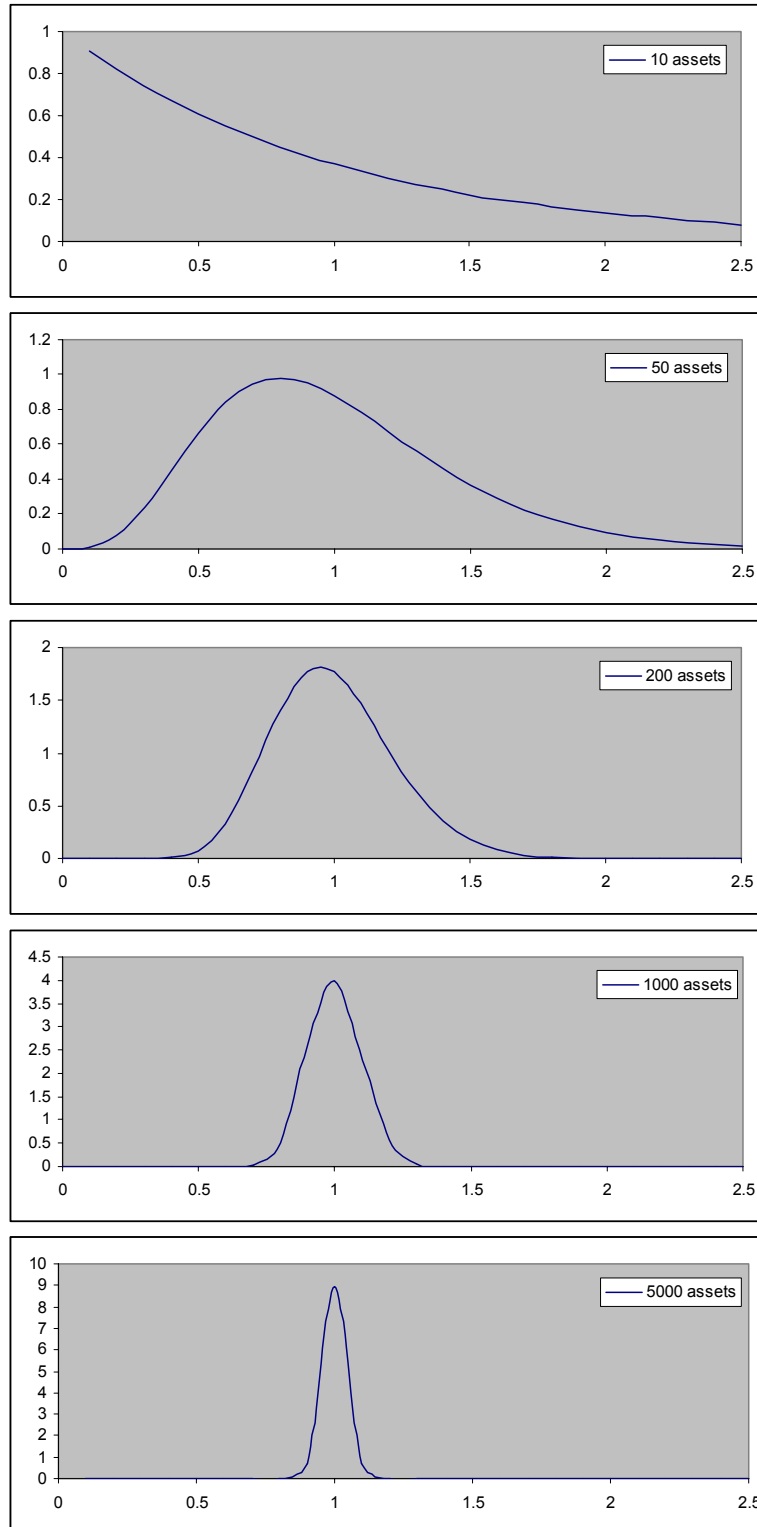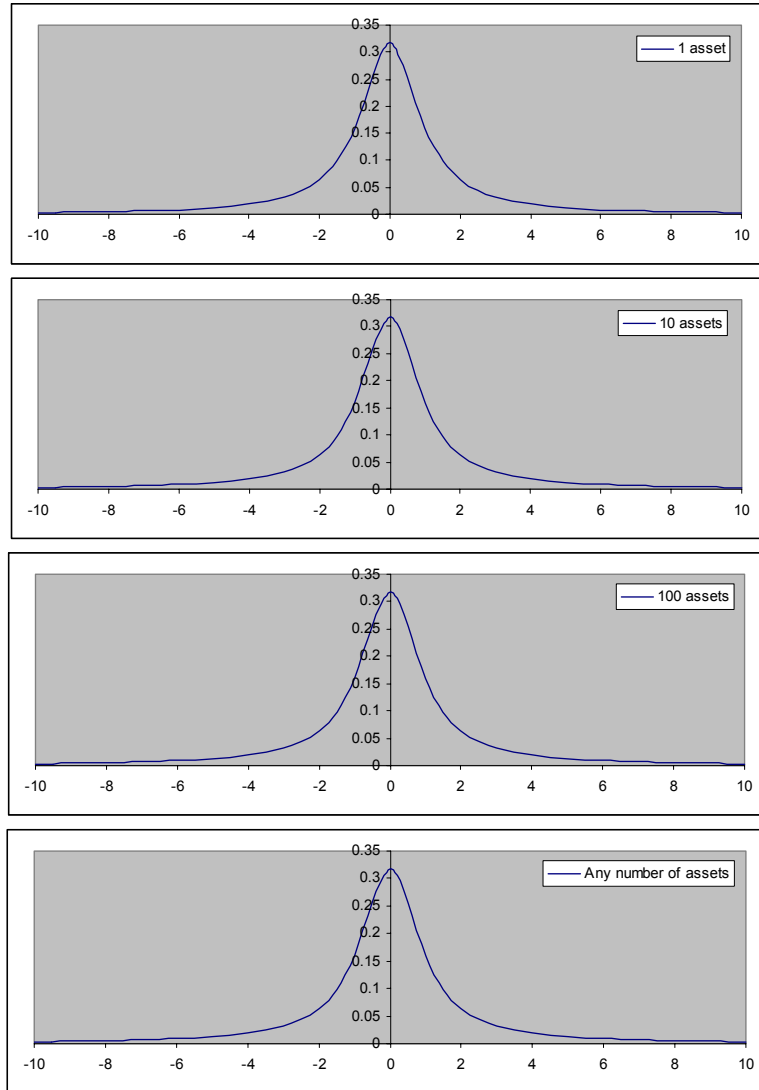
Figure 5.2c. Loss density function for portfolios of Cauchy-distributed assets. Unlike in the previous examples, diversification shows no effect. The risk does not decrease as the degree of averaging is increased. Fortunately this case "does not occur in practice".

Source for all Figures: CSFB

Independent random variables, mean-variance

As a starting point for the mathematical analysis it is reasonable to consider the mean and variance of a sum or average of independent random variables. The mean is easily dealt with: the mean of the sum is the sum of the means, and the mean of the average is the average of the means, regardless of whether the random variables are independent or not. What this implies, of course, is that the mean (expected loss) is of no use in trying to quantify diversification.

The variance is more interesting. The variance of the sum of random variables is only equal to the sum of their variances when they are uncorrelated. Independence, which we are assuming, implies uncorrelatedness, so letting V denote variance we may deduce

$$V\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{1}{n^2} V[X_1 + \cdots + X_n] = \frac{1}{n^2}\left(V[X_1] + \cdots + V[X_n]\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

with σ^2 denoting the variance of one asset. Hence the variance of the average is inversely proportional to portfolio size. Considering instead the standard deviation, we conclude that the standard deviation is inversely proportional to the square root of the portfolio size. To reduce unsystematic risk (as measured by standard deviation) by a factor of two, we must subdivide each asset by a factor of four, i.e. chop each asset into four independent assets, thereby ending up with four times the number of different obligors as we had before. This means that, in a theoretical and in a practical sense, **unsystematic risk decreases only slowly with portfolio size**. (Incidentally the exact same argument explains why Monte Carlo estimates converge slowly as the number of simulations is increased.)

From Figure 5.2b one can see by eye that in increasing the number of assets from 200 to 1000 to 5000 the "width" of the distribution (as some sort of proxy for the standard deviation) is reduced by a little more than a factor of 2 each time. Of course, the reduction in standard deviation is exactly $\sqrt{5} \approx 2.24$.) In Figure 5.2a a similar argument applies but it is harder to measure by eye.

One might enquire why in Figure 5.2c the distribution is the same regardless of the number of assets, when we have just argued that the variance is supposed to be decreasing as the number of assets increases. The answer is that the Cauchy distribution has infinite variance. Hence the mean-variance analysis does not apply: or, if you like, half infinity is still infinity.

Conclusions

We have begun to address the issue of unsystematic (diversifiable) risk in portfolios and have shown, somewhat loosely and without recourse to mathematical techniques, what happens when independent losses are added. In particular, we have observed that when we diversify a portfolio by chopping up the exposure into small pieces and distributing these across independent risks, the standard deviation of the loss distribution decreases as the square root of the number of pieces.

To be able to develop this subject matter in greater depth, and with greater applicability beyond the elementary examples that we have considered, we will need to introduce a variety of pieces of mathematical machinery. In particular we shall show a neat approximation, known as the granularity adjustment, for performing the calculation implied in Figure 1b.

6

Characteristically elegant

In the first three chapters we discussed the modeling of systematic risk and demonstrated how the various approaches fit into the same framework. Since then we have moved on to unsystematic risk, for which we gave an informal discussion of what happens when independent risks are added or averaged. In the next few chapters, we will look at some of the machinery that can be used to give quantitative answers to this question.

Résumé

We began this primer by describing a framework in which the existence of explanatory variables or factors is supposed, and conditionally where the values of assets are independent. This means that we must work out how to find the distribution of the sum of independent risks. It is this question that we will address in this and the next couple of issues. We will introduce some mathematical techniques for solving the problem and show them in action.

Whenever we talk about a default probability we are assuming that we have already conditioned on the latent factors; hence, we are talking about *conditional* default probabilities.

The characteristic function (Fourier transform)

Much of the theory that we will develop in our later work is based on properties of the characteristic function. As it will become apparent, the characteristic function is very suitable for analyzing sums of independent random variables. One can use this function to obtain numerical and analytical approximations that would be almost impossible by any other means.

Definition and properties

There are many similar conventions in the definition of the Fourier Transform but they all boil down to the following definition of the characteristic function of a random variable X by

$$C_X(\omega) = E[e^{i\omega X}] = \int_{-\infty}^{\infty} e^{i\omega x} f(x) dx$$

where f is its probability density function, E denotes expectation, and i is the square root of -1 (we need to use complex numbers). Any random variable has a characteristic function.

Here are its most important properties:

- Uniqueness. It is not possible for two different distributions to have the same characteristic function. Hence the characteristic function contains all the information about the distribution. Moreover, there is an inversion formula which looks very much like the Fourier Transform itself:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x} C_X(\omega) d\omega \,.$$

- Linear transformation. If we know the characteristic function of X then we can find the characteristic function of $aX+b$ (if a, b are constants) without any extra work. This means that we can easily deal with scaling a risk by some factor a (e.g. an asset allocation) or adding on some constant amount b (which in our situations would be a risk-free cash flow):

$$C_{aX+b}(\omega) = E[e^{i\omega(aX+b)}] = e^{i\omega b} E[e^{i(a\omega)X}] = e^{i\omega b} C_X(a\omega)$$

- Multiplication rule. If X and Y are two *independent* random variables then the characteristic function of $X+Y$ is the product of the characteristic function of X and the characteristic function of Y. One of the steps requires independence (marked "ind"):

$$C_{X+Y}(\omega) = E[e^{i\omega(X+Y)}] = E[e^{i\omega X} e^{i\omega Y}] \overset{\text{ind}}{=} E[e^{i\omega X}]E[e^{i\omega Y}] = C_X(\omega) C_Y(\omega) \,.$$

It is fairly obvious that this is going to be quite a useful tool for portfolio problems, because we can deal with weighted sums of independent random variables. The basic strategy is:

- Find the characteristic function of each asset's distribution.

- Multiply to get the portfolio characteristic function.

- Invert the characteristic function to obtain the portfolio distribution.

The middle step is trivial so we will deal with the first and last in more detail.

Examples of characteristic functions

For many well-known distributions the characteristic function can be expressed in closed form. Here are some examples of distributions that we have already encountered, with their corresponding characteristic functions:

Name	Distribution Function	Characteristic Function
Binomial	1 (prob p), 0 (prob $1-p$)	$1 - p + pe^{i\omega}$
Poisson	k with prob $e^{-\mu}\mu^k / k!$, $k=0,1,2,...$	$e^{\mu(e^{i\omega}-1)}$
Normal	Density $e^{-(x-\mu)^2/2\sigma^2} / \sigma\sqrt{2\pi}$	$e^{i\mu\omega - \sigma^2\omega^2/2}$
Gamma	Density $x^{\alpha-1}\beta^{-\alpha}e^{-x/\beta} / \Gamma(\alpha)$, $x>0$	$(1 - i\beta\omega)^{-\alpha}$
Cauchy	Density $1/[\pi\theta(1+(x/\theta)^2)]$	$e^{-\theta|\omega|}$

Source: Credit Suisse First Boston

Most of these results are easily derived by integration. For the Binomial the easiest route is to use the fact that the characteristic function is an expectation: $1-p$ of the time X is 0 and p of the time it is 1, so the characteristic function is $(1-p)e^{i\omega\times 0} + pe^{i\omega\times 1} = 1 - p + pe^{i\omega}$.

Families

By a family of distributions we mean a collection of distributions that are all mathematically similar. For example, all Normal distributions could reasonably be described as belonging to a family, and similarly, all Gamma distributions could be as well. Often the construction of families is done with reference to scaling and addition. By this we mean that

- if X is in the family, then aX is in the family (a = constant).

- if X_1 and X_2 are in the family and are independent then X_1+X_2 is in the family (sometimes restrictions have to be imposed).

As it is easy to write down the characteristic function of $X+Y$, construction of families is more easily done using characteristic functions. Some examples:

- Normal distribution. If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then their characteristic functions are $e^{i\mu_1\omega - \sigma_1^2\omega^2/2}$ and $e^{i\mu_2\omega - \sigma_2^2\omega^2/2}$ and the product of these is of the same form: $e^{i(\mu_1+\mu_2)\omega - (\sigma_1^2+\sigma_2^2)\omega^2/2}$. This can be recognized as the characteristic function of the Normal distribution with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$, as expected. It is important to note that it is the variances that we must add, not the standard deviations.

- Gamma distribution. If $X_1 \sim Ga(\alpha_1, \beta_1)$ and $X_2 \sim Ga(\alpha_2, \beta_2)$ then their characteristic functions are $(1-i\beta_1\omega)^{-\alpha_1}$ and $(1-i\beta_2\omega)^{-\alpha_2}$. Now the product of these is only of the same form when $\beta_1=\beta_2=\beta$ (i.e. the scale parameters are equal). Then the product of the characteristic functions is $(1-i\beta\omega)^{-(\alpha_1+\alpha_2)}$ which is the characteristic function of a variable with distribution $Ga(\alpha_1+\alpha_2, \beta)$. So the Gamma family is "partly closed under addition".

- Poisson distribution. As is also well known, f $X_1 \sim Poi(\mu_1)$ and $X_2 \sim Poi(\mu_2)$ then their characteristic functions are $e^{\mu_1(e^{i\omega}-1)}$ and $e^{\mu_2(e^{i\omega}-1)}$ and the product of these is of the same form: $e^{(\mu_1+\mu_2)(e^{i\omega}-1)}$. So $X_1+X_2 \sim Poi(\mu_1+\mu_2)$.

- Cauchy distribution. If $X_1 \sim Cau(\theta_1)$ and $X_2 \sim Cau(\theta_2)$ then their characteristic functions are $e^{-\theta_1|\omega|}$ and $e^{-\theta_2|\omega|}$ and the product of these is of the same form: $e^{-(\theta_1+\theta_2)|\omega|}$. So $X_1+X_2 \sim Cau(\theta_1+\theta_2)$.

In fact, all the examples we have shown have the same "shape": the dependence of the characteristic function is exponential in the relevant parameters, and so the parameters are added when the characteristic functions are multiplied (or the random variables are added). The families that are thereby generated are called *exponential families*.

Name	Parameters that generate the exponential family
Poisson	μ
Normal	μ , σ^2 (not σ)
Gamma	α (not β)
Cauchy	θ

Source: Credit Suisse First Boston

It is worth extending the argument a little for the Normal and Cauchy distributions to illustrate an important difference. Suppose that $Y=(X_1+\dots+X_n)/n$. Now:

- If all the $X_i \sim$ N(0,1) then $Y \sim$ N(0,1/n),   BUT

- If all the $X_i \sim$ Cau(1) then $Y \sim$ Cau(1), not Cau(1/n).

In other words, diversification smoothes out the extreme variations for Normally distributed variables, but has no effect if they are Cauchy distributed (when extreme events are much more common owing to the fat-tailed nature of the distribution). To give a technical reason for this effect, it is easiest by far to look at the characteristic function. For the Normal distribution the form is

$$\exp(\text{parameter} \times \omega^2)$$

whereas for the Cauchy it is

$$\exp(\text{parameter} \times \omega^1).$$

Adding n independent identically distributed random variables causes the parameter to be multiplied by n (multiplication rule). Dividing by n to get the average causes ω to be replaced by ω/n (linear transformation rule). In the Normal case the net effect is to divide the term in the exponential by n, which is synonymous with reducing the variance by a factor of n (the usual diversification law). In the Cauchy case the net effect is to do nothing: hence the lack of a diversifying effect. There is a range of intermediate cases too, and they are called "α-stable" distributions. Only the Normal and Cauchy cases have a closed form expression for the density function.

Inversion: some first thoughts

Once the portfolio characteristic function has been obtained, it is necessary to "invert" it to obtain the density function. In the examples of exponential families, we were able to identify the portfolio density function by inspection because its characteristic function was in a recognizable form. In practice, though, this is too restrictive an approach. We cannot expect to recognize the density function in, for example, the following situations:

- Adding Normal and Poisson variables

- Adding two Gamma variables whose scale parameters (β) are different

- ... the list is endless.

So we need to think of a better approach to inversion. One possibility that has some attraction is the Central Limit Theorem, which we turn to next.

Central Limit Theorem

The Central Limit Theorem states that, under appropriate conditions, the distribution of the sum (or average) of a large number of independent random variables is roughly Normally distributed. This is clearly quite a remarkable result as it makes no assumption about the precise nature of the distributions being added. We will not go into too many of the technicalities but we will sketch the underlying ideas. First we recall last chapter's "experiment" in which large numbers of independent identically distributed variables were averaged (reproduced in Fig 6.1a, 6.1b later on).

It is worth being a little more precise about what we mean by "is roughly Normally distributed". For one thing, we have to remember that (again assuming the unstated "appropriate conditions") the variance of the average tends to zero as the number of variables is made large. We do not want a theorem that just states that the distribution of the average tends towards a "spike" of probability mass located at the distribution's mean: that does not give us enough information. Consequently we standardize the distribution by subtracting its mean and dividing by the square root of its variance. The standardized distribution, therefore, has mean 0 and variance 1. In detail, when we standardize the sum we consider

$$\frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n\sigma^2}}$$

whereas, when we standardize the average we consider

$$\frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{\sqrt{\sigma^2 / n}}$$

which is identical. Hence it no longer matters whether we talk about the sum or the average. Here we have assumed that the X_j's are identically distributed; if they are not, we substitute $\frac{1}{n}\sum_{j=1}^{n}\mu_j$ for μ and $\frac{1}{n}\sum_{j=1}^{n}\sigma_j^2$ for σ^2, where μ_j and σ_j^2 are the mean and variance of X_j.

Now let us compute the characteristic function of

$$Y_n = \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n\sigma^2}}\,.$$

We hope to find that it will look like the characteristic function of the Normal distribution when n is large, under appropriate conditions. Following the same working as previously, the characteristic function of Y_n (assuming that the X_j's are identically distributed) is:

$$C_{Y_n}(\omega) = C_X\left(\frac{\omega}{\sqrt{n\sigma^2}}\right)^n e^{-i\omega\mu\sqrt{n/\sigma^2}}\,.$$

As $n \to \infty$, $\omega/\sqrt{n\sigma^2} \to 0$, and so we need to understand how $C_X(z)$ behaves when z is small. If (and this requires justification) $\log C_X(z)$ can be expanded in a Taylor series,

$$C_X(\omega) = \exp\left(a\omega + b\omega^2 + o(\omega^2)\right) \tag{*}$$

where $o(\omega^2)$ means a term smaller than ω^2 when ω is small, then it can be shown that:

$$a = i\mu_X \quad b = -\sigma_X^2/2 \,.$$

This is an important consequence that we will exploit in later work: the moments of the distribution are related to the behavior of $C(\omega)$ for small ω.

Substituting, we find

$$C_{Y_n}(\omega) = e^{\,n\left(\dfrac{i\omega\mu}{\sqrt{n\sigma^2}} - \dfrac{\omega^2\sigma^2/2}{n\sigma^2} + o(1/n)\right)} e^{-i\omega\mu\sqrt{n/\sigma^2}} \tag{**}$$

and letting $n \to \infty$ we arrive at:

$$C_{Y_n}(\omega) \to e^{-\dfrac{\omega^2}{2}}$$

which is the characteristic function of the Normal distribution of mean 0 and variance 1. This is what we set out to show. The crucial step is (*), and it is not always possible to perform the Taylor series expansion. A reasonable rule of thumb is that if the distribution has finite variance then expansion is possible. For the Cauchy distribution it is not possible because $\exp(-|\omega|)$ misbehaves at the origin (it is not smooth) and indeed the variance is not finite in that case.

We have therefore established the Central Limit Theorem for independent identically distributed variables. For non-identical variables the analysis of (**) is a little more complicated: essentially one has to derive a condition that ensures that the term marked "$o(1/n)$" does disappear faster than $1/n$ in the limit $n \to \infty$.

The Central Limit Theorem is a simple result and its use depends only on being able to work out mean and variance. However it is quite a poor approximation in practice, particularly for credit loss distributions, which are by their nature asymmetric (downside much more severe than upside). One can see this from Both Fig 6.1a and Fig 6.1b, where the number of assets needs to be fairly large (>50) before the distribution starts to look symmetrical.

In fact, the Central Limit Theorem is particularly bad at estimating the tail of the distribution, that is, the region of high losses. However, the graphs in Figure 6.1 do not show this very well because one cannot see the shape of the tail. We have therefore redrawn some of these figures in Figure 6.2, with the following modifications: we plot the tail probability rather than the probability density, and we make the vertical axis logarithmic. Figure 6.2 shows that the discrepancy can be quite significant: the CLT underestimates the tail probability by over a factor of ten in the top figure (25 defaultable bonds, conditional default probability 5%). The error is smaller for a larger portfolio (200 assets). On the other hand, reducing the conditional default probability to 0.5% causes

the error to increase again. Even for a portfolio of 200 assets the error is quite substantial, and for only 25 assets the approximation is of little use. (Admittedly much of the error arises from trying to approximate a discrete distribution by a continuous one, but even so the CLT gives a very poor result.)

Note again that Figures 6.1 and 6.2 are not "real" loss distributions (unless by some stroke of luck the losses happen to be uncorrelated): they are the conditional loss distributions for one particular value of the risk factor. In large, well-balanced, or highly correlated portfolios there is not much unsystematic risk anyway, so underestimating it is unlikely to be serious, but in small or unbalanced portfolios, or where the correlations are low, underestimating the unsystematic risk will have a much more severe impact. We will reintroduce correlation in Chapter 8.

Fig. 6.1a. Proportion of defaults in homogeneous portfolios of "loans" of default probability 0.05. The portfolio sizes are 25, 50, 200, 1000, 5000. Notice the "contraction" of the distribution around the mean (0.05) as the portfolio size increases.   Source: CSFB
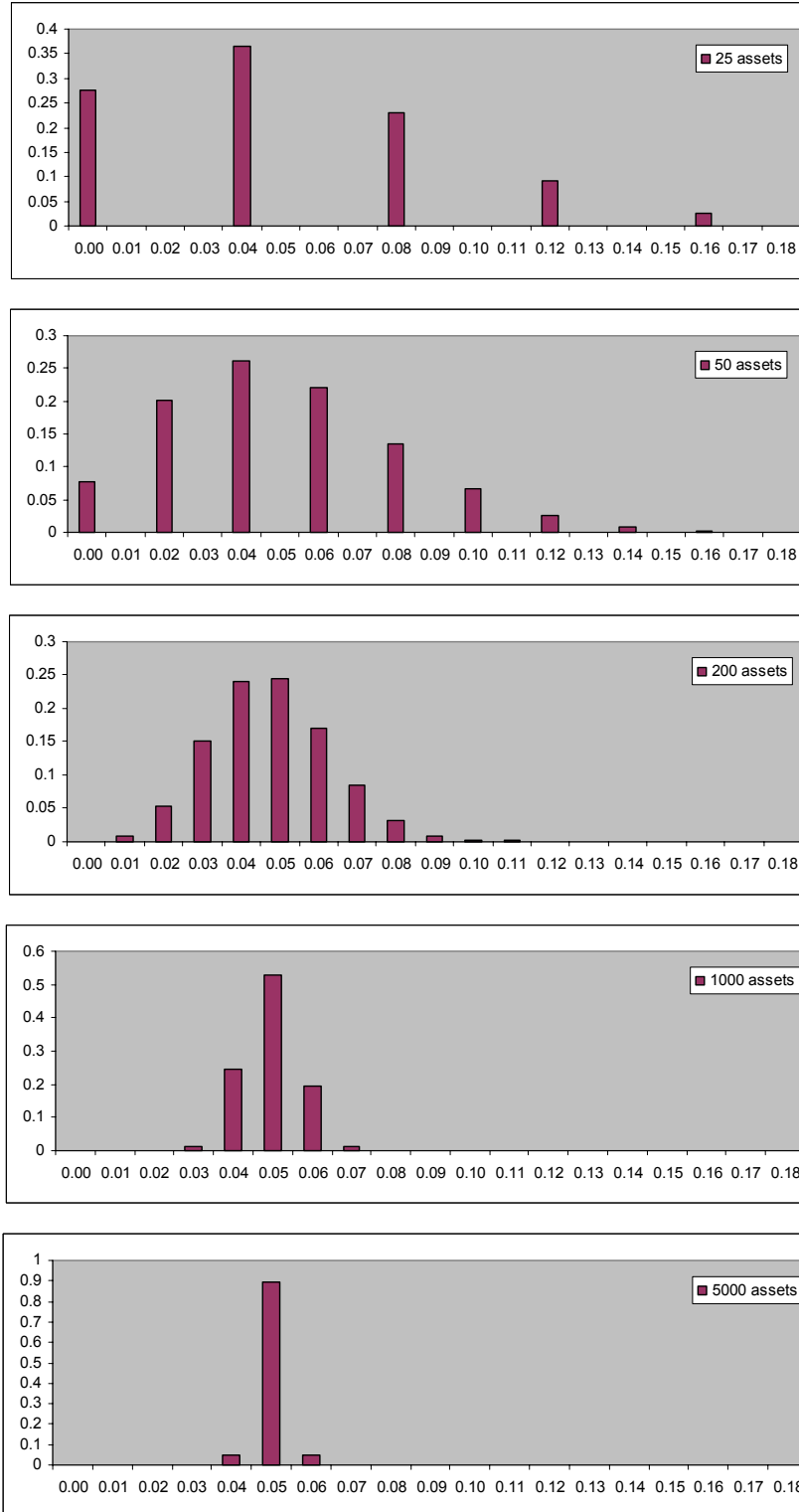
Fig. 6.1b. Loss density function of portfolios of Gamma-distributed assets, with increasing size (10, 50, 200, 1000, 5000) and hence increasing degrees of averaging (granulation). As before, notice the contraction of the distribution around the mean loss (1.0). Source: CSFB
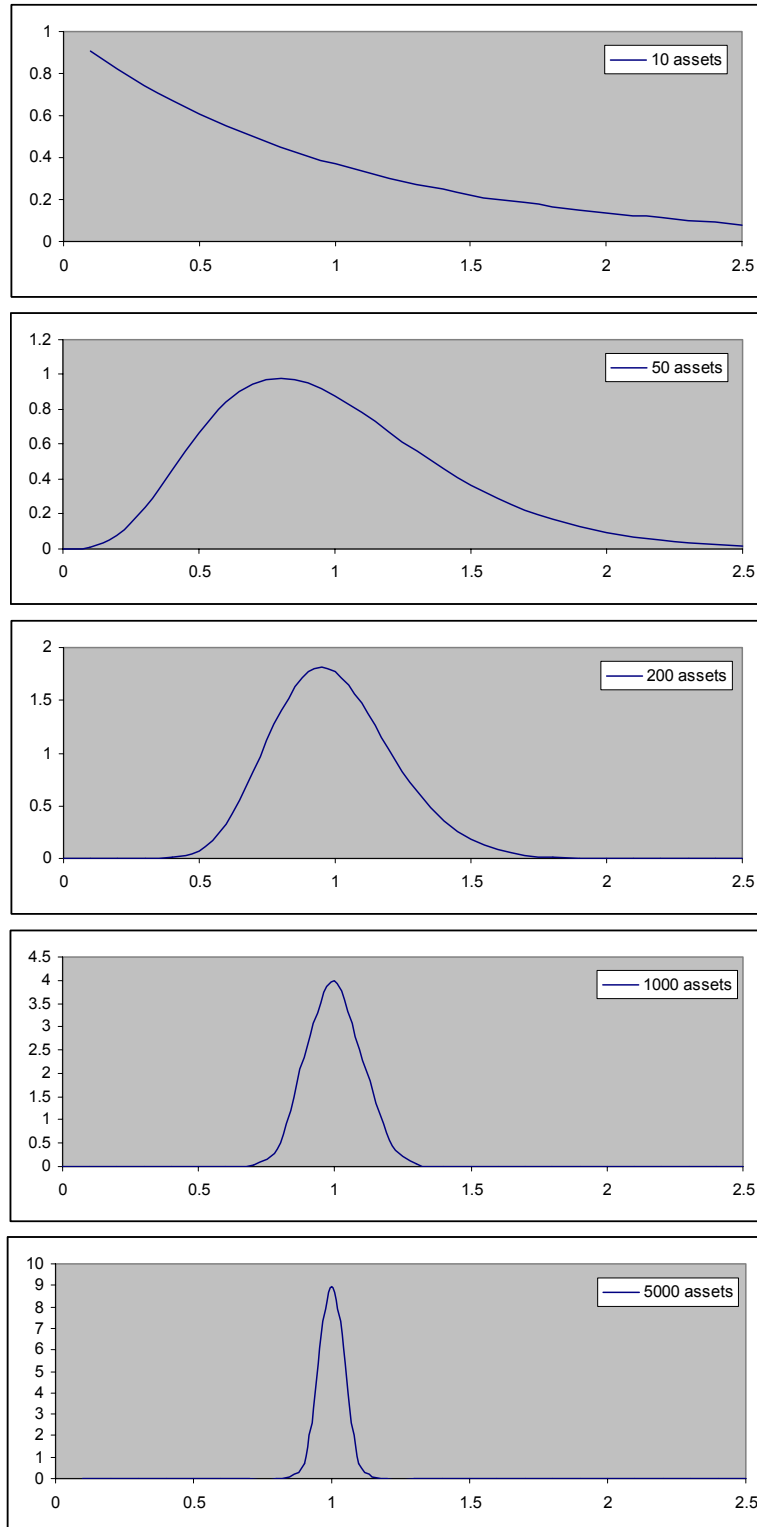
Fig. 6.2. Following Fig. 6.1a, this shows that the Central Limit Theorem (Normal approximation) underestimates the tail risk, though the error decreases with increasing portfolio size. The first two graphs reproduce two of the cases shown in Fig. 6.1a. The bottom two are for a lower default probability: the error is much bigger, as the distribution is more skewed then. Source: CSFB
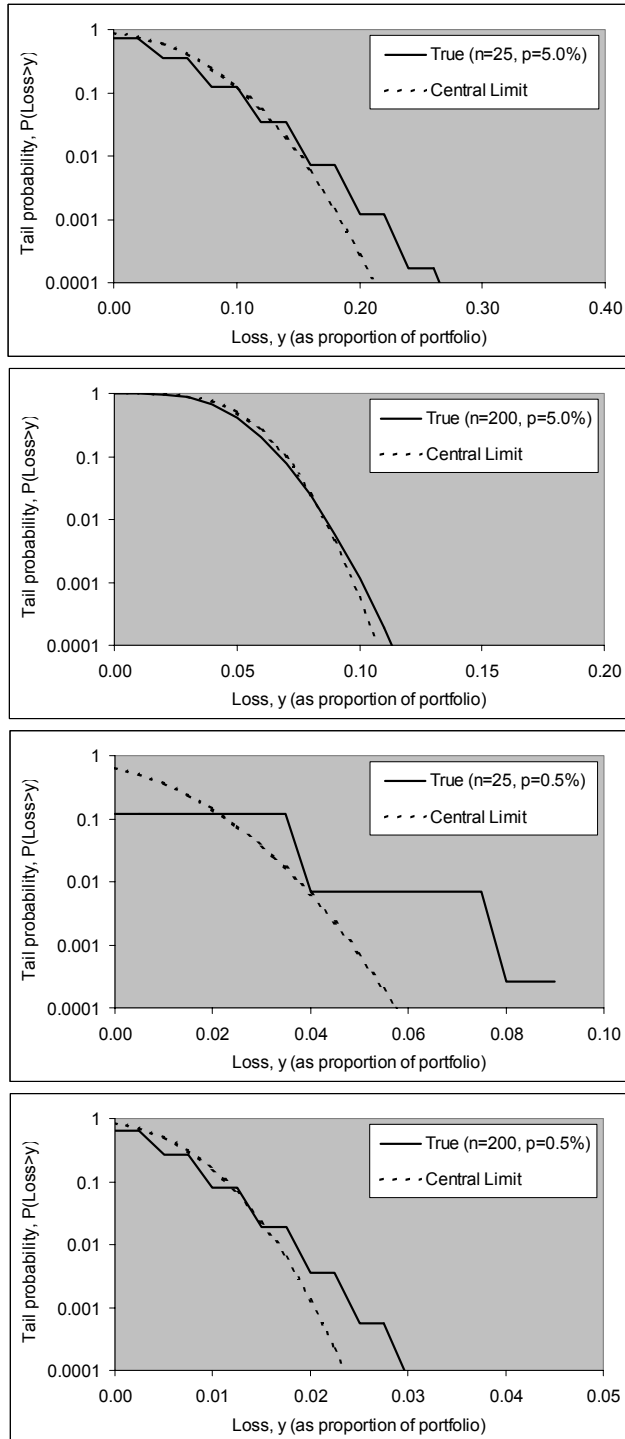
Numerical inversion

Having found the Central Limit Theorem to be rather disappointing in credit portfolio applications, we have two options: develop the analytics further (which we shall do in the next chapter) or try to invert the Fourier Transform (characteristic function) numerically.

When approximating an integral, a natural idea is to try and represent it as a finite (discrete) sum. This is the basis of the Discrete Fourier Transform. Recalling that the continuous formulation is:

$$C_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)dx , \qquad f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x} C_X(\omega)d\omega$$

it is natural (by discretising x and ω) to make the discrete analogue:

$$C_j = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} \zeta^{jk} f_k , \qquad f_j = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} \zeta^{-jk} C_k$$

with $\zeta = e^{i(\Delta\omega)(\Delta x)/N}$ (Δx and $\Delta\omega$ denoting the discretization steps in x- and ω-space). We will not pursue the technical details further, but suffice it so say that Δx and $\Delta\omega$ are related[37] by $\Delta x.\Delta\omega = 2\pi$). We have then arrived at the conventional definition of the Discrete Fourier Transform (DFT). This is an ideal tool for analyzing the distribution of random variables that only take values on a discrete grid of points (spaced by Δx). To find the distribution of the sum of several independent random variables, one performs the following sequence of operations:

- DFT the distribution of each
- Multiply the DFTs together
- Invert the DFT

At a first glance it appears that the DFT could be quite computationally intensive because the calculation of each of the N C_j's requires an N-length summation to be performed, which would mean a computational load of order N^2. Fortunately the DFT has the remarkable property that it can be computed much more rapidly if N can be factorized. The algorithm (known as the Fast Fourier Transform Algorithm) works fastest when N is an exact power of 2, and then the computational load is roughly $N\log_2 N$, which is much less than N^2 (e.g. for $N=1024$, compare 10,240 with 1,048,576). Hence the three-stage scheme described above is quite satisfactory in practice, provided that one chooses Δx (and hence $\Delta\omega$) correctly and takes enough points.

To give an illustration of how this works in practice, we consider a portfolio of 5 defaultable loans with the following characteristics (recovery is supposed to be zero):

Exposure	Default prob
4	5%
5	2%
3	10%
6	2%
2	4%

Source: Credit Suisse First Boston

---

[37] This is known as time-frequency reciprocity and is fundamental in the theory of signal processing.

The loss distribution is as follows:

Figure 6.3. Loss distribution for test portfolio, obtained numerically
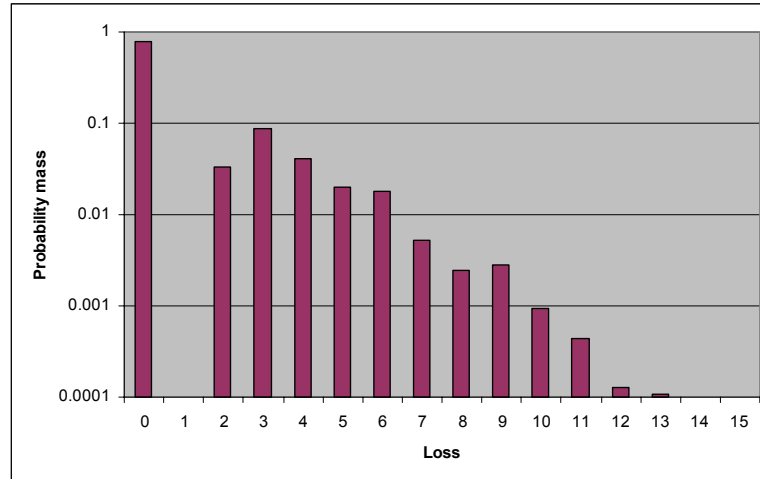
Source: CSFB

This could have been obtained long-hand, by working out each possible combination of defaults. While this is practicable for very small portfolios, it is impracticable for large ones because when the exposures are different there are 2^n different loss combinations.

This technique is particularly well-suited to the analysis of distributions that are discrete. One situation in which these often arise is in the pricing of synthetic CDOs and baskets where it is desired to find the distribution of the number of default events. The implementation is more fiddly when the distribution is continuous; this situation arises when recovery rates are made stochastic or because it is mark-to-market risk that is being modeled rather than just default/no-default.

Conclusions

We have introduced the characteristic function and derived some useful results from it: the Central Limit Theorem, which is a Normal approximation to the distribution of a large uncorrelated portfolio, and a discrete form which gives a numerical expression for the distribution via the Discrete Fourier Transform. We have seen that the Central Limit Theorem is of limited applicability in credit unless the portfolio is large (more than a few hundred), because of the asymmetric nature of the risk. In the next chapter we shall develop a more sophisticated analytical approximation, the saddle-point approximation, that works better on smaller portfolios.

7

Posing on the saddle: the cowboys of portfolio theory

One of the techniques that we think gives CSFB a significant advantage is the saddle-point method. This technique was introduced a few years ago to quantitative finance, but it is not particularly well understood. To give a complete exposition would be quite technical and there are also several different ways of using saddle-point methods, which give rise to essentially different types of approximation. The purpose of this chapter is to explain some of the basic ideas and issues and to give a few simple examples.

Résumé

In this chapter we pick up from the last one. The issue is the derivation of an analytical approximation to a sum of independent random variables. We have already found that:

- For "most" distributions, if many of them are averaged (and if they are independent), then the so-called "law of large numbers" applies and the variability becomes progressively less (diversifies) as the portfolio size increases. In the limit, only the conditional expected loss remains;

- For many distributions, the Central Limit Theorem says that if enough independent variables are added then their sum is roughly Normally distributed;

- But in practice the Central Limit Theorem does not always work very well, and it underestimates the probability of large losses (losses that are several standard deviations from the mean).

Obviously the last point is a cause for concern. But how much? We can give a qualitative answer to this question immediately. The extra risk that occurs even when the values of the risk factors are known is unsystematic (or specific) risk, and it is significant when any or all of the following occur:

- The portfolio is small;

- The correlations are low (high correlation means that most of the risk comes from variation of the risk factors);

- The distributions are highly skewed, as tail risk diversifies quite slowly.

The moment-generating function (MGF)

When we discussed the Central Limit Theorem in the last chapter we made considerable use of a device called the *characteristic function*, which for a random variable *X* is defined as

$$C_X(\omega) = E[e^{i\omega X}] = \int_{-\infty}^{\infty} e^{i\omega x} f(x) dx$$

where *f* is its probability density function, *E* denotes expectation, and i is the square root of –1. For the purposes of this discussion, we slightly modify the discussion by defining instead the moment generating function (MGF),

$$M_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} f(x) dx \ .$$

Although the difference between these may seem trivial, there is an important distinction. Whereas every distribution has a characteristic function, it is not always the case that the moment-generating function will exist for all real values of the variable *s*: and it is real values that we will be most interested in. This is because the exponential function e^{sx} increases rapidly (unlike the function $e^{i\omega x}$, which never gets bigger than 1 in absolute value, and simply oscillates), and so the integral does not necessarily converge. This makes the MGF a more specialized tool than the characteristic function, but when it exists, more powerful results can be derived.

Here are examples of MGFs:

Name	Distribution function	MGF
Binomial	1 (prob *p*), 0 (prob 1–*p*)	$1 - p + pe^s$
Poisson	*k* with prob $e^{-\mu}\mu^k / k!$, *k*=0,1,2,...	$e^{\mu(e^s-1)}$
Normal	Density $e^{-(x-\mu)^2/2\sigma^2} / \sigma\sqrt{2\pi}$	$e^{\mu s + \sigma^2 s^2/2}$
Gamma	Density $x^{\alpha-1}\beta^{-\alpha}e^{-x/\beta} / \Gamma(\alpha)$, *x*>0	$(1-\beta s)^{-\alpha}$ $(s < 1/\beta)$
Cauchy	Density $1/[\pi\theta(1+(x/\theta)^2)]$	**Does not exist**

Source: Credit Suisse First Boston

As the Gamma density decays only at an exponential rate, the MGF "blows up" if *s* is too large, hence the restriction $(s < 1/\beta)$. The Cauchy distribution is so fat-tailed that the integral always "blows up."

All the properties of the characteristic function carry over to the MGF, so we repeat them:

- Uniqueness. It is not possible for two different distributions to have the same MGF. Hence, the MGF contains all the information about the distribution. The inversion formula is similar to that for the characteristic function and it requires $M(s)$ to be evaluated for purely imaginary s:

$$f(x) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{-sx} M_X(s)\,ds \; .$$

- Linear transformation. If the MGF of X is known, then the MGF of $aX+b$ (if a, b are constants) can be found without any extra work. This means that we can deal with the following transformations: scaling a risk by some factor a (e.g., an asset allocation), and adding on some constant amount b (which, in our situations, would be a risk-free cash flow):

$$M_{aX+b}(s) = E[e^{s(aX+b)}] = e^{bs} E[e^{(as)X}] = e^{bs} M_X(as) \; .$$

- Multiplication rule. If X and Y are two *independent* random variables then the MGF of $X+Y$ is the product of the MGF of X and the MGF of Y. One of the steps requires independence (marked "ind"):

$$M_{X+Y}(s) = E[e^{s(X+Y)}] = E[e^{sX}e^{sY}] \overset{\text{ind}}{=} E[e^{sX}]E[e^{sY}] = M_X(s)M_Y(s) \; .$$

So far we have not in effect done anything beyond the theory of characteristic functions. The following properties are therefore deeper and exploit the fact that $M(s)$ is assumed to exist for *real s*:

- Differentiability. The moment-generating function is a complex-differentiable (analytic) function, and this endows it with a variety of useful properties. One of these is given below, and a more far-reaching consequence will be discussed later.

- Extraction of moments. Courtesy of the differentiability property, M possesses a Taylor series expansion around the origin ($s=0$) and the moments equate to the derivatives of M there (hence the name moment-generating function):

$$M_X(s) = \sum_{r=0}^{\infty} \frac{E[X^r]s^r}{r!} \quad \text{so} \quad E[X^r] = M_X^{(r)}(0) \; .$$

- Cumulant-generating function. It is convenient to define the cumulant-generating function by the log of the MGF, and this is also analytic. When independent random variables are added, their cumulant-generating functions add:

$$K(s) = \log M(s)$$

$$K_{X+Y}(s) = K_X(s) + K_Y(s) \qquad (X, Y \text{ independent})$$

- Cumulants. The cumulants are the derivatives of K evaluated at the origin, in the same way that the moments are the derivatives of M evaluated at the origin.

$$K_X(s) = \sum_{r=1}^{\infty} \frac{\kappa_r s^r}{r!} \quad \text{so} \quad \kappa_r = K_X^{(r)}(0).$$

The two are closely related, but the cumulants are generally more useful. Their interpretation is as follows:

The 3rd and higher cumulants vanish for Normal distributions, so they can be regarded as a measure of non-Normality, with κ_3 giving a measure of asymmetry

Cumulant	Interpretation
κ_1	mean, $E[X]$
κ_2	variance, $E[X^2] - E[X]^2$
κ_3	skewness is $\kappa_3/\kappa_2^{3/2}$
κ_4	(excess) kurtosis is κ_4/κ_2^2

Source: Credit Suisse First Boston

and κ_4 a measure of tail fatness. There is no zeroth cumulant because $K(0)$ is always 0.

- Convexity. Of crucial importance to the development of the theory, K is convex[38]:

$$K''(s) > 0.$$

Review of Central Limit Theorem

We recall that the Central Limit Theorem (CLT) states that, under appropriate conditions, the distribution of the sum (or average) of a large number of independent random variables is roughly Normally distributed. The proof of this depends on investigating the characteristic function of such a sum, in the limit of infinitely many variables being added, i.e., an infinitely large portfolio. Provided certain technical conditions are met, that characteristic function converges to the characteristic function of the Normal distribution. When those technical conditions are not met, as we saw in fact for the Cauchy distribution, other limiting distributions are possible, but we shall not explore any further in that direction.

What is of most importance to us at the moment is that the CLT does not always work very well in practice. This is because a limit theorem just says, "for a large enough portfolio it will work", but doesn't tell us whether the portfolio is large enough. In fact, there couldn't be a result that just gave a simple statement such as "it works well for portfolios >100 in size", because logically the underlying distributions must have *some* influence on the rate of convergence. We saw this in our earlier discussion of the CLT, and we shall reproduce some of the results for convenience (please refer forward to Figures 7.2 and 7.3). Figure 7.2 is an example of a default/no-default model (such as might be used for analyzing a CDO, or an accrual-accounting model of a loan portfolio). If the default probabilities are assumed to be 5%, the CLT becomes reasonably accurate once >200 assets are added, but if they are 0.5%, then the accuracy is not very good, even for that number. The reason for this is that the events that are being

---

[38] There are a large number of convexity results that surround this, and they all stem from the following basic idea. Define $Q(t) = E[(X+t)^2 e^{sX}]$. Then $Q(t)>0$ for all t, because it is the expectation of something positive. Hence its minimum is also positive, and that can be found by "completing the square". This when rearranged states that $M''(s)M(s) > M'(s)^2$, or $K''(s)>0$, as required.

added are highly asymmetric: the downside is very rare and large, and the upside is minimal (here, the upside is that there is no downside!). The position is even worse if the exposure amounts are not all the same, because in that case there is concentration risk to diversify away and this requires an even larger portfolio size.

Looking at Figure 7.1, it seems reasonable to hypothesize that the CLT is doing better near the mean, or in the 'middle of the distribution', where the tail probability is around 0.5, than in the right-hand tail (tail probability <1%, say). This is an important observation, because the first derivation of the saddle-point approximation that we are going to show makes explicit use of this idea. In fact, it is sufficiently important that we are going to garner some more evidence in its favor.

What makes Figure 7.1 a little awkward to deal with is that the true portfolio distribution is discrete, as opposed to a continuum of possible outcomes: the true distributions are 'steppy', whereas the CLT approximations are smooth. So part of the error simply comes from the approximation of a discrete distribution by a continuous one, and this is not really a comment on the substance of the CLT. Accordingly, it might be a good idea to try the CLT out on a continuous distribution, where this discrete-vs.-continuous issue does not arise. A useful test is to look at Gamma distributions (which we also considered previously). Recall that Gamma distributions form a *family*, which means that when independent Gamma-variables are added[39], the result is another Gamma-variable: this is useful because we know what the exact answer is. (If we did not know the exact answer, we would have to find some other technique for testing our analytical approximations, which would probably be Monte Carlo simulation. There is nothing wrong with this, but it introduces unwelcome distractions, such as what the Monte Carlo error is.) In fact, the Gamma distribution is not chosen as a purely academic exercise. In the structural model of credit risk, the credit risk in a bond is a put option on the underlying firm's assets, and so the price difference between a risky and risk-free bond of the same maturity is the price of this embedded put option. Options are nonlinear instruments, so the distribution of the value of a put option is non-Gaussian: indeed, common sense tells us it must be, as a short put position has a small upside (spread tightens) and a big downside (spread widens all the way out to default). It turns out— though we ask the reader to take this on trust—that the Gamma distribution is a reasonable approximation to this, though there are better ones.

A test for Gamma distributions is shown in Figure 7.2. By plotting the results on both linear and logarithmic vertical scales, we get a good appreciation of the general shape (from the former) and the tail (from the latter). We see that indeed the CLT is much better in the middle than it is in the tail.

This characteristic is known as 'non-uniform convergence'. The purpose of saddle-point approximations is to make the convergence more uniform, and we can now introduce them.

---

[39] Scale parameters must be equal. See Chapter 6 for details.

Enter the saddle

The following derivation of the saddle-point approximation is quite daring in its simplicity. Suppose that the mean (expected loss) is \$10M and we are interested in losses around the \$70M mark. Noting that the CLT works quite well in the middle of the distribution, try this idea:

> "I shall deform the distributions of all my loss variables in such a way that the total mean loss is now \$70M, and then apply the CLT, and because the CLT works well near the mean, I shall get a good answer."

What does 'deform' mean? For our purposes, the most analytically convenient way of deforming a distribution is to multiply its density by an exponential 'tilting' function[40]. First, let us see what happens when we do this to the portfolio loss variable $Y = \sum_j X_j$

directly: then we shall see what has to be done to the distributions of the individual loss variables X_j to achieve this.

If we change the density of Y by multiplying by $e^{\lambda y}$, we get a new density

$$\widetilde{f}_Y(y) \propto f_Y(y)e^{\lambda y} .$$

The effect of this operation is to shift the bulk of the density over to the right if $\lambda > 0$, or the left if $\lambda < 0$. We have written 'proportional to' because the tilted density does not integrate to 1, and we now fix that problem by dividing by the integral of the RHS, which is just the MGF of Y evaluated at λ:

$$\widetilde{f}_Y(y) = \frac{f_Y(y)e^{\lambda y}}{M_Y(\lambda)} . \tag{1}$$

It is not difficult to show that using the tilted probability distribution, the mean of Y is now $K'(\lambda)$ and the variance is $K''(\lambda)$, where K denotes the log of the MGF (as introduced earlier) and $'$ denotes differentiation. More substantially, the MGF of Y using the tilted distribution is

$$\widetilde{M}_Y(s) = \frac{M_Y(s + \lambda)}{M_Y(\lambda)} . \tag{1*}$$

We choose λ so that the 'tilted mean' is the loss level in which we are interested—say y^* (= \$70M in our discussion just now):

$$y^* = K'_Y(\lambda^*). \tag{2}$$

As $K''(\lambda) > 0$ (which we said earlier was most important), for each y there is only one λ that will do this. If y is equal to the mean, $E[Y]$, then $\lambda = 0$, which means that no tilting is being done.

So we now have a new probability distribution that has the desired mean loss. However, we need to do a bit more than this: the distributions of the individual loss variables X_j need to be altered to make sure that their sum does indeed follow this new distribution.

---

[40] This, the 'Esscher tilt', is also used in importance sampling. Like the CLT, but for completely different reasons, Monte Carlo estimates best in the middle of the distribution.

In other words, we cannot just pluck a distribution for Y out of thin air. To do this, we make the tilted density of X_j

$$\widetilde{f}_{X_j}(x) = \frac{f_{X_j}(x)e^{\lambda^* x}}{M_{X_j}(\lambda^*)} ;$$

then the moment-generating function of each X_j under the new measure is

$$\widetilde{M}_{X_j}(s) = \frac{M_{X_j}(s + \lambda^*)}{M_{X_j}(\lambda^*)}$$

and the MGF of the portfolio loss is obtained by multiplying them together (they are still independent):

$$\widetilde{M}_Y(s) = \frac{M_Y(s + \lambda^*)}{M_Y(\lambda^*)} ,$$

which is what it is supposed to be (see (1*)).

Now the CLT, applied in the new measure, states that the distribution of Y is roughly Normal and it is likely to be a good approximation near the mean, which we have made equal to y^*. And the density of any Normal distribution at its mean is given by $1/\sqrt{2\pi\sigma^2}$ (where σ^2 is the variance), so in the tilted measure we must have

$$\widetilde{f}(y^*) \approx \frac{1}{\sqrt{2\pi K''(\lambda^*)}}$$

and 'un-tilting' by (1) gives us the basic **saddle-point approximation**

$$f(y^*) \approx \frac{e^{K(\lambda^*)-\lambda^* y^*}}{\sqrt{2\pi K''(\lambda^*)}} . \tag{3}$$

In evaluating this, we must remember that y^* and λ^* are linked by equation (2).

Demonstration

Figures 7.1 and 7.2 also show the saddle-point results. Figure 7.2a, which shows the density, uses equation (3). Figures 7.1 and 7.2b, rather than integrating (3) to get the tail probability (which would work quite well), instead use a better and more convenient result. We shall talk about that later, as we want to discuss the test results, keeping the focus on examples rather than theory. The figures provide an easy comparison of the true distribution, CLT and saddle-point approximation. Clearly the results are very good.

What is quite remarkable is that the saddle-point approximation works well even on distributions that are not at all Normal—and yet the derivation we have given makes the saddle-point approximation look like a variant of the CLT. The top graph in Figure 7.3a shows a very good approximation to an exponential distribution (10 independent assets with Gamma (0.1,10) distribution, when added, give an exponential distribution). The true probability of the loss exceeding 4 is about 0.018, the saddle-point approximation is almost exact, but the CLT makes it 0.00135, which is a factor of >10 too small.

Figure 7.1. Risk aggregation of Binomial assets (single default events, as in a simple model of a loan portfolio or CDO). Here the true loss distribution is shown as the solid, stepped line: the distribution is discrete because a whole number of events must occur. If one imagines smoothing out the true distribution, it is apparent that the CLT underestimates the probability of large losses, and the saddle-point approximation does a reasonable job of "following a median path". The CLT performs worst for a small portfolio with low default probabilities, as that combination produces the highest skew (deviation from non-Normality).
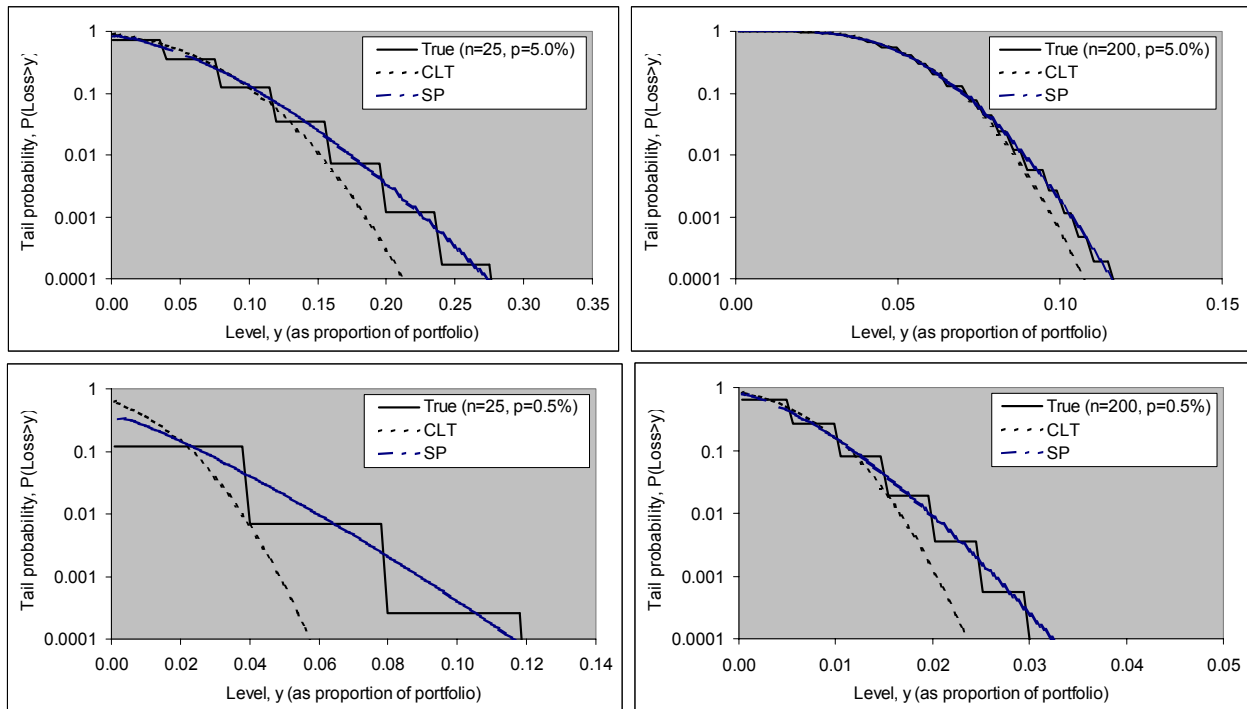
Source: Credit Suisse First Boston

Figure 7.2a. Risk aggregation of Gamma-distributed assets using CLT (dotted line) and saddle-point approximation (dashed line), compared with the correct distribution (solid line). Here the density is being shown. The CLT underestimates the probability of large losses and the saddle-point method does very well, even on the smallest portfolio where the distribution is highly non-Normal.
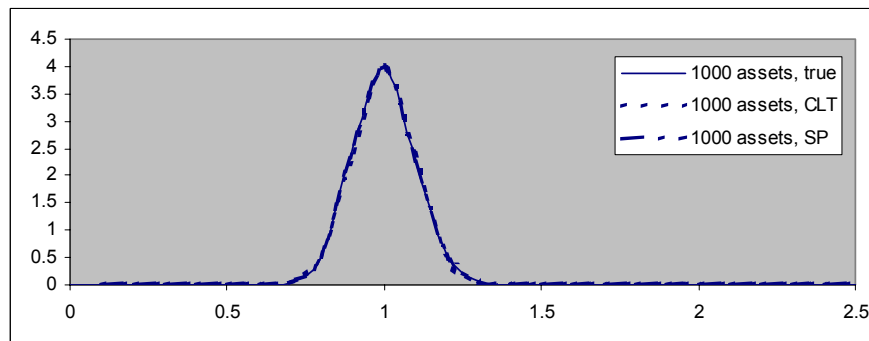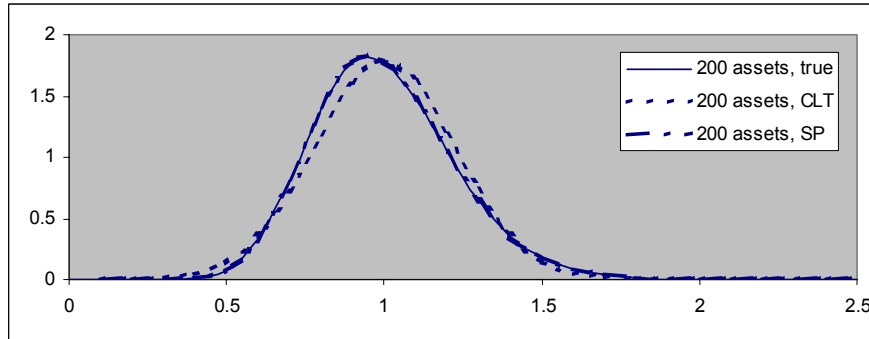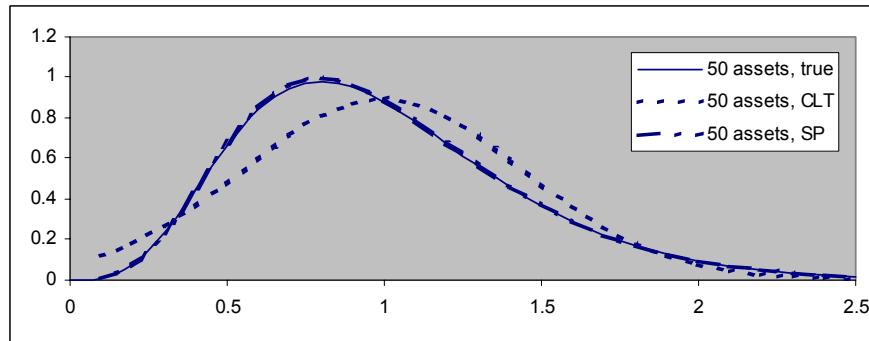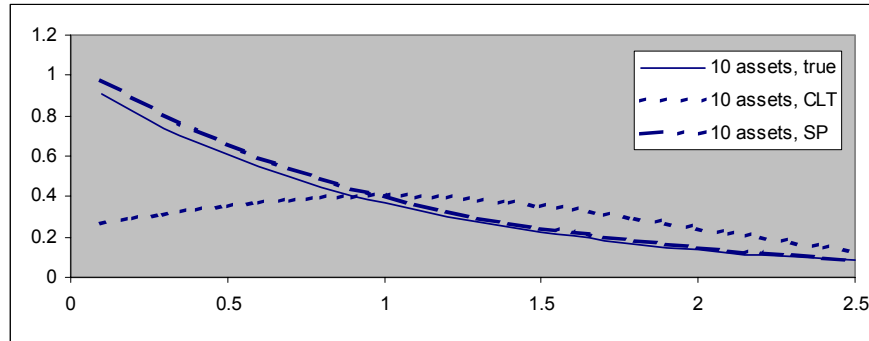
Source: Credit Suisse First Boston

Figure 7.2b. As Figure 7.2a, but showing the tail probability on a logarithmic scale, to "amplify" the tail. As is apparent, the saddle-point approximation is indistinguishable from the true result, while the CLT underestimates the tail risk.
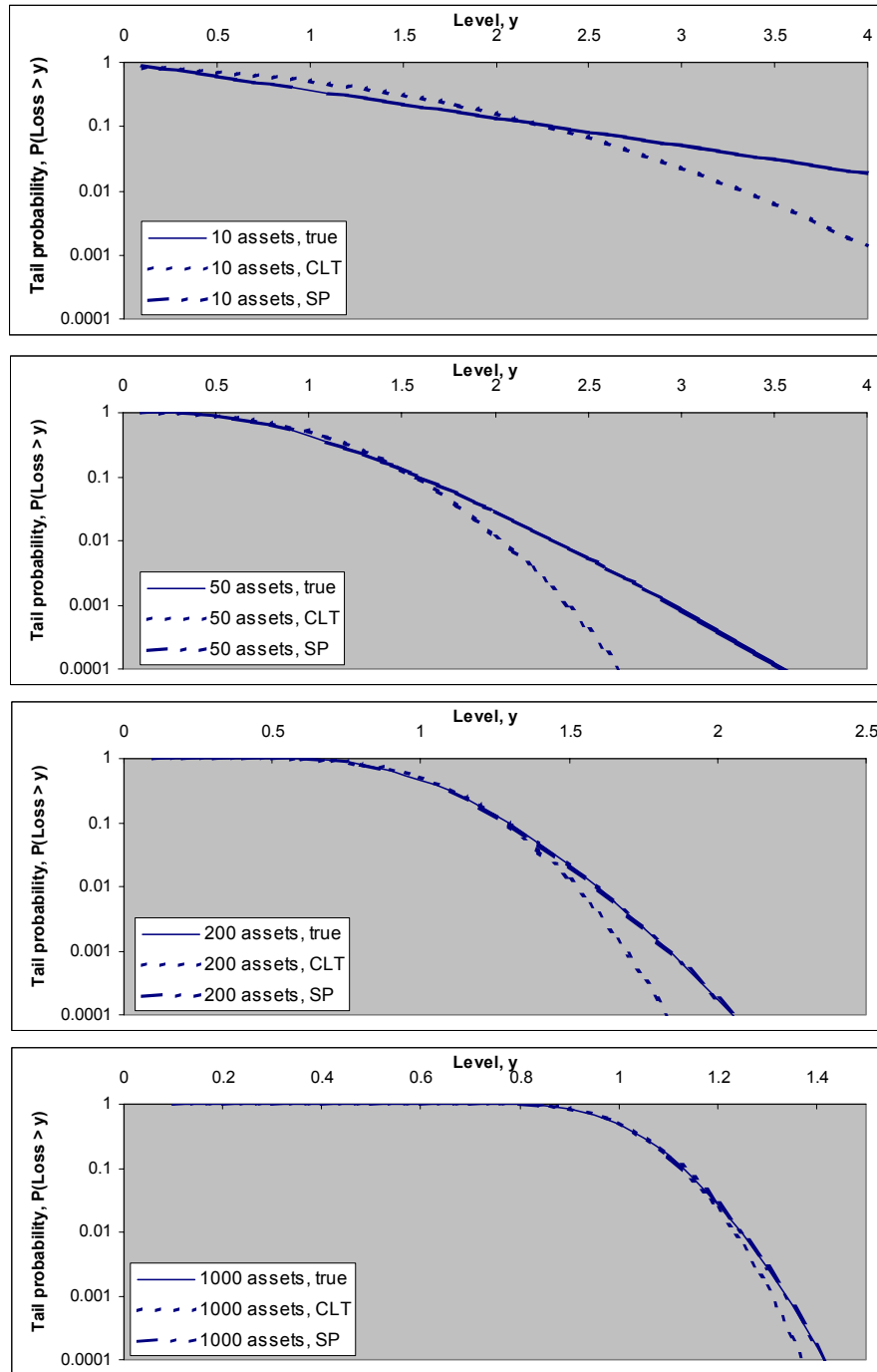
Source: Credit Suisse First Boston

Uniform approximation property

A general rule of saddle-point approximations when applied to a sum of independent random variables is that they generally exhibit uniform approximation in the limit of a large portfolio. By uniform we mean that the relative error is roughly the same across the distribution. So, for example, if the true density at one point is 0.57 and the approximation is 0.62, and at some other point the true density is 0.0071 and the approximation is 0.0075, we say that the relative error is roughly the same (a few percent), even though the *absolute* error is much larger in the middle (0.05 cf. 0.0004). The relative error tends to be more important than the absolute.

Second derivation of the saddle-point method

An alternative, and considerably more flexible, derivation of the saddle-point method consists in returning to the original Fourier inversion formula for the density:

$$f_Y(y) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{-sy} M_Y(s)\,ds = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{K_Y(s)-sy}\,ds \;.$$

Assume that Y arose as a sum of n independent identically distributed random variables, each of which has cumulant-generating function K_X. Then $K_Y(s) = nK_X(s)$ and we arrive at

$$f_Y(y) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{n(K_X(s)-s\bar{y})}\,ds$$

where $\bar{y}$ is short for y/n (and hence is the average rather than the total loss).

The question to resolve now is what happens when n is made large. The basic idea with integrals of this sort is that in this limit the integrand becomes 'concentrated' around one or more points and negligible elsewhere: so the integral is approximated by adding together the contributions from these points. On account of the fact that K_X is convex, the term in the exponential is 'bowl-shaped', as s varies and is real, and assumes its minimum at the value of s satisfying[41]

$$K_X'(s) = \bar{y} \quad \text{(equivalently, } K_Y'(s) = y \text{).}$$

In the orthogonal direction (as the imaginary part of s varies), the term in the exponential has instead a local *maximum*[42]. But there is a minor problem: that point does not lie on the path of integration! (The saddle-point is a real value of s, whereas the path of integration is over imaginary values.) It is at this point that we can wheel out an important result from analytic function theory, which states that it is OK to deform the path of integration so that it passes through the saddle-point, and the value of the integral will be unaffected[43]. This can only be invoked because the MGF is an analytic function, which is the 'far-reaching consequence' we alluded to earlier when introducing the MGF.

---

[41] differentiate w.r.t. s and set the derivative to zero
[42] This is where the term saddle-point comes from: analytic functions are not permitted to have local maxima, in the way that many real functions do. For a complex function, both the real and the imaginary parts can be altered, and an analytic function must have 'a local maximum in one direction and a local minimum in the other'.
[43] (provided there are no singularities to negotiate) – Cauchy's theorem.

By expanding the integrand in a Taylor series as far as the quadratic term, and doing the integration, one ends up with

$$f(y) \approx \frac{e^{K(\hat{s}) - \hat{s}y}}{\sqrt{2\pi K''(\hat{s})}}$$

which is (3).

This method draws a neat comparison with the CLT. Both can be viewed as ways of approximating the integral above: but whereas the CLT always expands around the origin ($s=0$), the saddle-point method gets its information about $K_Y(s)$ from around the saddle-point, which is where the dominant contribution to the integral is (note again that its position depends on the point on the loss distribution at which one is seeking to find the density). This is probably the cleanest explanation of why saddle-point methods work well.

We mentioned earlier that in approximating the tail probability there are better methods available than just integrating the density. The upper tail probability is given by

$$P(Y > y) = \frac{1}{2\pi i} \int_{-i\infty,(0+)}^{i\infty} e^{K_Y(s) - sy} \frac{ds}{s}$$

in which the notation (0+) indicates that a 'certain avoidance' has to be taken owing to the fact that the integrand "blows up" at the origin. In fact, this singular behavior causes some delicate problems and the most satisfactory approach (best discussed by Daniels (1987)[44]) is the method of Lugannani & Rice. This consists in splitting out the singular part of the integral into one that can be done exactly and approximating the rest using the saddle-point method. The first (singular) part is fiddled with so that the term in the exponential becomes exactly quadratic as opposed to only approximately so. This is done by changing the variable to z given by

$$K(s) - sy = K(\hat{s}) - \hat{s}y + \frac{1}{2}(z - \hat{z})^2$$

$$0 = K(\hat{s}) - \hat{s}y + \frac{1}{2}\hat{z}^2$$

so that when $s=0$, $z=0$, and also when $s = \hat{s}$, $z = \hat{z}$. The development is then

$$P(Y > y) = \frac{1}{2\pi i} \int_{-i\infty,(0+)}^{i\infty} e^{\frac{1}{2}z^2 - z\hat{z}} \frac{1}{z} dz + \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{\frac{1}{2}z^2 - z\hat{z}} \left(\frac{1}{s}\frac{ds}{dz} - \frac{1}{z} \right) dz$$

and the first part is recognizable as $\Phi(-\hat{z})$, while the second is approximated as

$\phi(\hat{z}) \left(\dfrac{1}{\hat{s}\sqrt{K_Y''(\hat{s})}} - \dfrac{1}{\hat{z}} \right)$. The complete result is then

$$P(Y > y) \approx \Phi(-\hat{z}) + \phi(\hat{z}) \left(\frac{1}{\hat{s}\sqrt{K_Y''(\hat{s})}} - \frac{1}{\hat{z}} \right)$$

---

[44] H E Daniels, "Tail Probability Approximations", International Statistical Review 55(1):37-48 (1987).

which is known as the *Lugannani-Rice formula*. PR+2 uses a variant of this and also has an analogous formula for the CVaR, which requires the treatment of the following integral:

$$E[(Y - y)^+] = \frac{1}{2\pi i} \int_{-i\infty,(0+)}^{i\infty} e^{K_Y(s)-sy} \frac{ds}{s^2}.$$

This completes our introduction to the 'classical' saddle-point method, which is employed by PR+2. We finish with a discussion of other implementations of the saddle-point method that have been used in related applications.

Alternative uses of saddle-point approximations

The line of argument we have pursued here—using the saddle-point method to approximate the sum of independent random variables—is the basis of the PR+2 methodology. It also corresponds very closely to the classical development of the subject (for which see, for example, Jensen (1995)[45] or Daniels (loc. cit.)). However, it is not the way in which the subject was introduced into finance. That approach took the view that the saddle-point method often works well when applied to all sorts of distributions, not just those that arise from sums of independent random variables[46]. It is important to bear in mind that when used like this the uniform approximation property is no longer guaranteed. Hence, the results may not be very good.

Examples include:

- Feuerverger & Wong (2000)[47] used it to approximate the distribution of a quadratically transformed multivariate Gaussian variable, which they used as a general model for market risk. Here the idea is that there are a variety of underlyings, all Normally distributed (e.g., FX rates, interest rate tenors, etc.) and the values of the instruments indexed on them (options, etc.) are roughly quadratic. In option-pricing parlance, this is referred to as the delta-gamma approximation. The resulting model has the desirable property of having a moment-generating function that is known in closed form. The saddle-point method can then be used to obtain the density and tail probability.

- Gordy (1998)[48] used it to approximate the loss distribution from CSFB's CreditRisk+ model of credit portfolios. As one of the main planks of CreditRisk+ is that the portfolio MGF be known in closed form, the saddle-point inversion is particularly easy to implement.

- Martin (1998)[49] applied it to a reliability problem, which can also be applied in insurance: losses of random severity arrive as a Poisson process, and it is desired to find the distribution of total loss in some time period. Again the model was chosen so that the MGF could be calculated in closed form (coincidentally, it bears a strong resemblance to CreditRisk+).

---

[45] J L Jensen, "Saddlepoint Approximations", Oxford (Clarendon Press), 1995.
[46] R J Martin et al, "Taking to the saddle", RISK 14(6):91-94 (2001).
[47] A Feuerverger and A C M Wong, Computation of VaR for nonlinear portfolios, J of Risk 3(1):37-55 (2000). Beware that there are a fair number of minor errors in this paper, particularly in the equations.
[48] M B Gordy, "Saddlepoint approximation to CreditRisk+", J. of Banking and Finance 26(7):1337-1355 (1998).
[49] Jensen (1995) has a good discussion on this sort of thing. Martin's original paper is hard to get.

The point about these examples is that they were all implemented with the 'direct method', as described next.

Direct method

- Find the unconditional moment-generating function. If the model is a conditional-independence model (Credit Risk+ very obviously is, for example), find the conditional portfolio moment generating function $E[e^{sY}|V]$ for each value of the risk factor V, and then integrate over V to get $E[e^{sY}]$.

- Apply the saddle-point method to $E[e^{sY}]$.

But the approach we are using in PR+2 is different, and we call it the 'indirect method':

Indirect method

- Condition on the risk factor V.

- For each value that V takes, find the conditional portfolio moment-generating function $E[e^{sY}|V]$ and apply the saddle-point method to that, to get the probability density, tail probability or whatever.

- 'Integrate out' over V: this usually has to be done numerically by finding a representative set of 'states' (see earlier chapter[50]) and averaging over these.)

What makes these two methods analytically distinct is that for the indirect method to work it is only necessary that the saddle-point method be a satisfactory approximator for a sum of **independent** random variables (it is being applied to the conditional MGF, for each value of V). For the direct method, one is using the saddle-point method as a black box (once, to approximate the unconditional MGF). In the examples listed above, there are technical reasons why the direct method is likely to work very well[51]. However, there is no general result that says that it must, and one can easily find examples in which is does not work very well at all[52]: on these, the indirect method is much better. Unfortunately, the indirect method is more difficult to implement, but the problems are not insuperable and indeed they have been overcome in PR+2.

Conclusions

We have discussed the basics of the saddle-point method and have shown how and why it is an improvement over the Central Limit Theorem.

Incidentally, it is an important consequence of the saddle-point theory that the risk contributions, i.e., sensitivities of VaR and CVaR to asset allocations, can be calculated analytically at the same time. This makes for an efficient implementation of a significant amount of portfolio analysis.

---

[50] "Getting the full picture", Chapter 8.
[51] The distributions involved are Poisson and Gamma, and both are very close to their saddle-point approximations, at least for reasonable values of the parameters.
[52] One of the most common models in credit portfolio analysis, the case of default/no-default models using a Gaussian copula, is a good example of this. Results on this will be published shortly.

8

Getting the full picture

In the previous chapters we have discussed the modeling of systematic risk and showed how the various approaches fit into a risk-factor framework. We then proceeded to show what happens when independent variables are added as a route to understanding idiosyncratic (unsystematic) risk. It is now time to put these two concepts together to build up the whole picture of a loss distribution.

Résumé

In our previous chapters on credit portfolio modeling, we built up the following picture:

- In the case where a single risk factor drives the systematic risk, and when the portfolio is very large so that unsystematic risk can be ignored, the portfolio loss distribution can easily be obtained. To work out Value-at-Risk (VaR) is very easy: we take the appropriate percentile of the factor distribution, and work out the expected loss of the portfolio when the factor takes that particular value[53].

- An example of such a model, for defaultable bonds that either default (loss=1) or do not default (loss=0), a model can be constructed by assuming that default occurs when the (unobserved) value of an obligor's assets falls below the level of its liabilities. Assume that the obligor's assets are after simple transformation Normally distributed, and perform a linear regression to correlate all firms' assets to a single factor V. The expression for the conditional default probability, and hence the conditional expected loss of a particular obligor, is

$$\Phi\left(\frac{\zeta_j - c_j V}{\sqrt{1 - c_j^2}}\right) \qquad (1)$$

Here ζ_j and c_j are parameters that respectively describe the credit quality (the average default probability is $\Phi(\zeta_j)$) and the level of correlation of the obligor.

As a matter of fact, this model is used by Basel II Regulatory Capital Proposals. If a more complicated model than default/no-default is required, the same arguments apply, but formula (1) has to be modified.

---

[53] There is a technicality: the relation between the factor and the portfolio expected loss must be "one-to-one", i.e., either the expected loss always increases as the value of the factor increases, or the expected loss always increases as the value of the factor decreases. But if two different values of the factor give rise to the same expected loss, then we have a problem.

- We have thus obtained the distribution of the "infinitely granular" portfolio (see dotted distribution in Chart 8.1). The effect of unsystematic risk, which arises because the portfolio is only of finite size and may not be very homogeneous, is to increase the risk and cause the loss distribution to "spread out" (solid distribution in Chart 8.1).

- We have considered how this "spreading-out" works in theory, by arguing as follows. When a large number of independent[54] random variables are added, the distribution of their sum (i.e., the portfolio distribution) can be approximated as Normal[55], with standard deviation proportional to $1/\sqrt{n}$, where n is the portfolio size. Under more restrictive assumptions we can invoke the saddle-point approximation, which as we have seen does a better job.

Figure 8.1: *Systematic and unsystematic risk, general case for continuous risk factor*
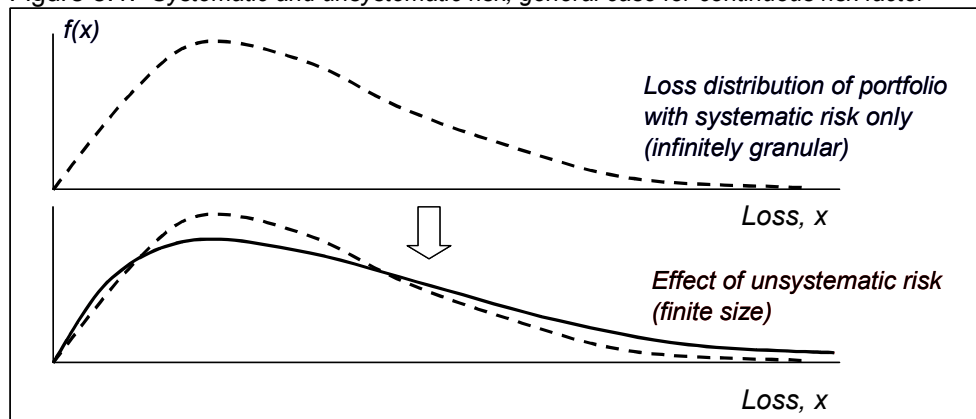
Source: Credit Suisse First Boston

In this chapter, we combine the elements of systematic and unsystematic risk. In other words, we are going to construct the solid line in Chart 8.1. We can do this in at least two ways. One is an explicit construction that needs implementing in a computer program. The other seeks to make an analytical approximation of the VaR of the solid distribution given the VaR of the dotted one, i.e., it is an adjustment to the infinitely granular portfolio. It is known as the "granularity adjustment" and is a useful technique that is a current area of research.

---

[54] They are independent because we are considering their distribution conditionally on the systematic factors.
[55] Subject to technical conditions that we shall assume are met.

Combining systematic and unsystematic risk

We assume that conditionally on the risk factor, which we denote V, the distribution of portfolio loss is Normally distributed with mean $\mu(V)$ and variance $\sigma^2(V)$; the latter of these will be quite small when the portfolio is large and diversified. Then the probability of the loss exceeding y is obtained by integrating over V:

$$P(Y > y) = E\left[\Phi\left(\frac{\mu(V) - y}{\sqrt{\sigma^2(V)}}\right)\right] = \int_{-\infty}^{+\infty} \Phi\left(\frac{\mu(V) - y}{\sqrt{\sigma^2(v)}}\right) f_V(v)dv \qquad (2)$$

with $f_V(v)$ the density function of V. Hence all we have to do is perform the integration by some numerical method. If the distribution of V is something quite simple, such as a Normal distribution (as we have been dealing with), then the method of Gaussian quadratures can be used. This seeks to write the integral as a weighted sum of samples taken at discrete points. The Gauss-Hermite formula is for integrals of the form $\int_{-\infty}^{\infty} g(v)\phi(v)dv$, where $\phi(v) = e^{-x^2/2}/\sqrt{2\pi}$ is the density of the Normal distribution; the formula is

$$\int_{-\infty}^{\infty} g(v)\phi(v)dv \approx \sum_{k=1}^{m} h_k g(v_k)$$

where m is known as the order, (h_k) the weights and (v_k) the abscissas. These are chosen in such a way that the formula is exact if g is a polynomial of degree less than or equal to $2m$, and there are standard ways of determining them. For example, the formula of order 11 is shown in Table 8.1. Incidentally, the abscissas are spaced roughly by $\pi/\sqrt{m}$; this property is particular to the Normal distribution (i.e., the Gaussian quadrature formulae for other distributions do not have this property).

In statistical language, we have replaced the continuous distribution of V with a discrete one. Monte Carlo simulation also does this, but by choosing random points instead of deterministic ones. The Gaussian quadrature formula has the convenient property of matching the first, second and so on up to the $(2m-1)$th moments of the distribution. Provided g is quite smooth (and hence can be approximated quite well by a polynomial), the method works well. Note that we don't have to work out what the polynomial approximation is to perform the integral, though if it were required, it could be found fairly easily.

Table 8.1 Abscissas and weights of the 11-point Gauss-Hermite quadrature formula

Abscissa	Weight (%)
5.188	8.12E-07
3.936	1.96E-04
2.865	0.0067
1.876	0.0661
0.929	0.2422
0.000	0.3694
-0.929	0.2422
-1.876	0.0661
-2.865	0.0067
-3.936	1.96E-04
-5.188	8.12E-07

Source: Credit Suisse First Boston

Chart 8.2a Normal distribution

Source: Credit Suisse First Boston

Chart 8.2b 11-point Gauss-Hermite quadrature rule (outermost two points not shown)

Source: Credit Suisse First Boston
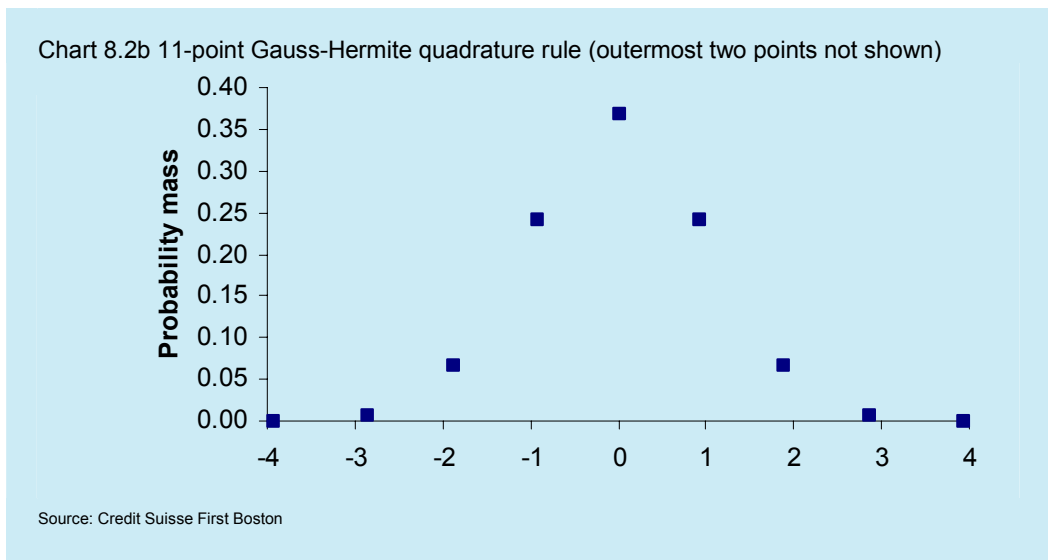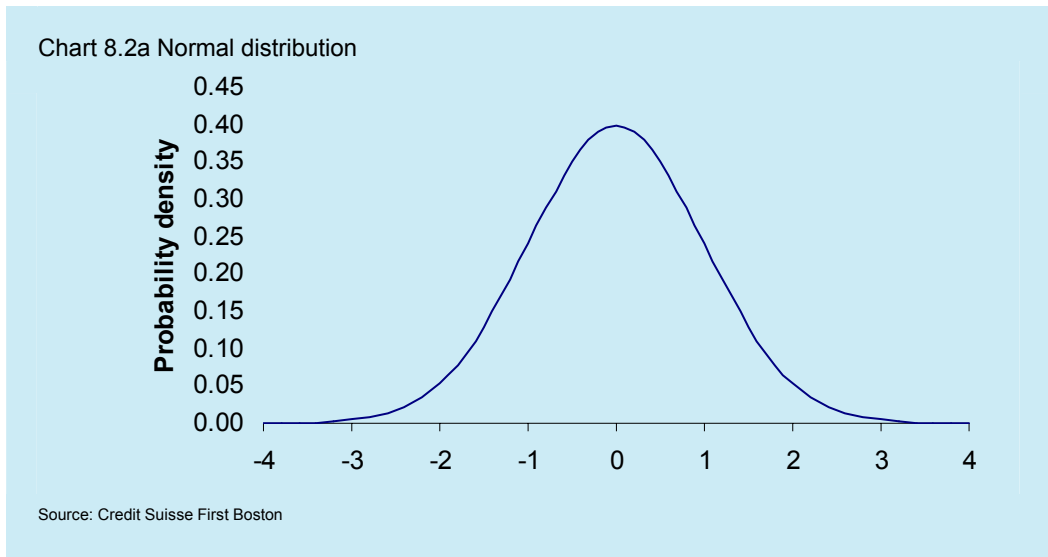
To return to the problem at hand, we can use the Gaussian quadrature formula without any further ado, because we just recognize that

$$\Phi\left(\frac{\mu(v) - y}{\sqrt{\sigma^2(v)}}\right) = g(v) \quad \text{and} \quad f_V(v) = \phi(v) .$$

Hence our end result is

$$P(Y > y) \approx \sum_{k=1}^{m} h_k \Phi\left(\frac{\mu(v_k) - y}{\sqrt{\sigma^2(v_k)}}\right) \tag{3}$$

which not too difficult to implement, for example in Excel.

This describes the method for single-factor models. When the number of factors is large, this kind of scheme is inefficient because of the large number of points that would be required. We have implemented methods that combine analytical integration, Monte Carlo integration and Gaussian quadrature for this purpose, but the details are outside the scope of this chapter.

Example: Gaussian copula model

For the Gaussian copula model in a default/no-default framework, the conditional expected loss is

$$\mu(V) = \sum_{j=1}^{n} a_j \Phi\left(\frac{\Phi^{-1}(p_j) - c_j V}{\sqrt{1 - c_j^2}}\right)$$

(as previously shown) and the conditional variance is

$$\sigma^2(V) = \sum_{j=1}^{n} a_j^2 \Phi\left(\frac{\Phi^{-1}(p_j) - c_j V}{\sqrt{1 - c_j^2}}\right) \overline{\Phi}\left(\frac{\Phi^{-1}(p_j) - c_j V}{\sqrt{1 - c_j^2}}\right)$$

because for a single loss event with probability p and loss a, the variance is $a^2 p(1-p)$, and for independent losses one adds the variances. [Here $\overline{\Phi}$ denotes the upper tail probability of the Normal distribution: $\overline{\Phi}(x) = 1 - \Phi(x) = \Phi(-x)$.] These expressions can then be substituted into (3).

Granularity adjustment

An alternative treatment avoids the use of numerical integration by assuming that the conditional variance is small and using an analytical approximation. The objective is to argue that a small amount of unsystematic risk causes the VaR to increase by a small amount, and find the first-order sensitivity. Its advantage is that it can be calculated easily, and is simpler to implement than a numerical method. This is an interesting and active area of current research. The result is

$$\Delta VaR \approx -\frac{1}{2f(y)} \frac{\partial}{\partial y}\left[\sigma^2(v^*)f(y)\right]$$

where $f(y)$ is the density function of the infinitely granular portfolio's loss, and $\sigma^2(v^*)$ is the variance of unsystematic risk for the value of the risk factor (v^*) that corresponds to a portfolio loss of y (in the infinitely granular portfolio).

The proof is technical and given in the Appendix.

Numerical example 1

We consider the following hypothetical example of a portfolio of defaultable loans, the details of which are presented in Table 8.2.

Table 8.2 Test portfolio

	Exposure	Default prob	Correl		Exposure	Default prob	Correl
1	2.27	0.0044	0.43	26	0.23	0.0169	0.35
2	1.19	0.0081	0.42	27	0.42	0.0037	0.57
3	0.10	0.0095	0.33	28	1.02	0.0247	0.39
4	0.28	0.0075	0.39	29	2.73	0.0032	0.34
5	0.49	0.0068	0.52	30	0.61	0.0002	0.32
6	0.18	0.0108	0.46	31	1.10	0.0237	0.50
7	0.75	0.0121	0.58	32	5.85	0.0003	0.59
8	0.54	0.0014	0.56	33	0.50	0.0517	0.41
9	0.00	0.0051	0.53	34	0.60	0.0011	0.33
10	3.91	0.0013	0.31	35	0.93	0.0015	0.46
11	0.10	0.0258	0.48	36	0.77	0.0089	0.52
12	3.01	0.0092	0.36	37	0.70	0.0300	0.47
13	3.00	0.0010	0.56	38	0.13	0.0114	0.52
14	0.77	0.0050	0.32	39	0.36	0.0313	0.58
15	0.57	0.0047	0.32	40	1.08	0.0402	0.58
16	0.76	0.0042	0.34	41	2.06	0.0115	0.50
17	1.17	0.0059	0.46	42	1.26	0.0048	0.43
18	2.16	0.0062	0.51	43	0.13	0.0017	0.46
19	0.99	0.0070	0.52	44	0.21	0.0071	0.34
20	3.17	0.0000	0.45	45	0.62	0.0094	0.45
21	0.82	0.0053	0.54	46	0.28	0.0032	0.36
22	0.30	0.0094	0.58	47	0.14	0.0022	0.42
23	1.39	0.0020	0.54	48	1.62	0.0094	0.54
24	0.43	0.0012	0.33	49	3.20	0.0090	0.32
25	0.88	0.0005	0.54	50	0.82	0.0157	0.32

Source: Credit Suisse First Boston

The exposures were chosen by randomly drawing from an exponential distribution with mean 1 (see note[56]), the default probabilities $(\Phi(\zeta_j))$ by randomly drawing from a exponential distribution with mean 0.01, and the correlations (c_j) by drawing from a uniform distribution between 0.3 and 0.6. The recovery rates are assumed to be zero.

We are going to analyze the portfolio in several different ways:

- Monte Carlo simulation. This consists in simulating the factor (V), then simulating independent defaults conditionally upon it, and repeating many times. This shows the "true" loss distribution, but no particular insight is gained from it.

  Infinitely granular portfolio. This is the graph of VaR against tail probability, obtained by plotting the conditional expected loss $\mu(V)$ as a function of the factor V. (We showed this in Chapter 2 for a similar test portfolio.) Specifically, $\mu(V)$ is plotted on the horizontal axis and $\Phi(V)$ on the vertical axis, because the tail probability is just $\Phi(V)$. Note that more losses occur when V is negative, so it is the lower tail of V that is associated with the upper tail of the loss distribution.

- Introducing unsystematic risk by assuming that conditionally on V the portfolio loss is Normally distributed, and integrating over V as described earlier.

- Using the granularity adjustment method.

---

[56] Excel: =-LN(RAND()).

Chart 8.3 Loss distributions estimated by various methods
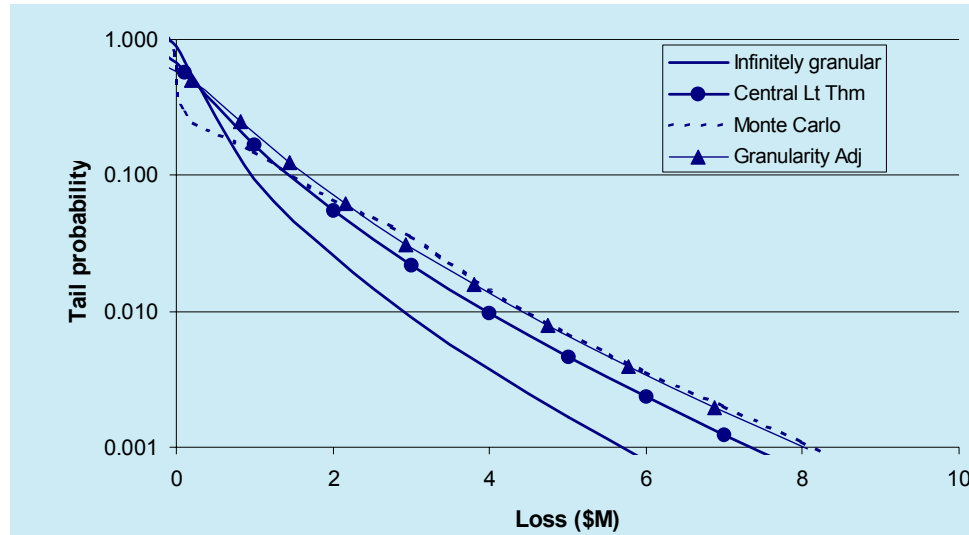
Source: Credit Suisse First Boston

The infinitely granular approximation and Central Limit Theorem both underestimate the true risk.

In discussion of these results:

- The infinitely granular distribution shows the least risk. This is because no unsystematic risk is taken into account.

- The Central Limit Theorem underestimates the true risk (Monte Carlo) because the Normal approximation does not capture the skewness (asymmetry): the downside is much bigger than the upside. We saw this in our previous discussions of unsystematic risk: when the default probabilities are low, a large number of independent losses need to be added before the distribution becomes close to normal.

- The granularity adjustment produces a very good approximation to the true distribution, but this is due to two canceling errors. The method is supposed to approximate the Central Limit Theorem result, but it overestimates. On the other hand, the Central Limit Theorem underestimates the true risk. As can be seen from this example, the errors almost exactly cancel.

By graphing the true loss distribution and the infinitely granular loss distribution, one obtains a measure of the proportion of systematic and unsystematic risk in a portfolio. For example, if we use VaR(99%) as a risk measure, then we have the true risk at about 4.4M$ but the systematic-only risk at about 2.9M$. So about two-thirds of the risk is systematic. The point about this is that the unsystematic risk is diversifiable. If the single-name risk is concentrated in a few names, as seems likely (though we haven't yet talked about how to analyze that), we can surmise that a couple of well-chosen trades might be able to reduce the portfolio risk while keeping the return about the same.

Numerical example 2

This is a second example of a default/no-default model with the Gaussian copula model for correlation. The rough details of the portfolio are shown in the table in Figure 8.4. The correlation model is the PR+ 18-sector correlation model described in Chapter 11.

As is apparent, using the Central Limit Theorem to approximate the unsystematic risk causes the tail risk to be underestimated. Using the saddle-point method removes nearly all the error, though there is a slight overestimation of risk in the $0-10M loss region. This can be ascribed to an appreciable probability of no loss, which causes a spike of probability at the origin: the saddle-point method cannot pick this up[57] and so it tries to smooth the distribution out, causing the overestimation near the origin. In practice this is not something to worry about.

Conclusions

We have shown how to combine systematic and unsystematic risk to build up the full portfolio loss distribution. The previous few chapters have raised a few issues:

- Can we improve on the Central Limit Theorem?

- How can we break down the pie chart in Chart 8.4 further and show risk contributions of individual obligors?

We shall begin to address these questions in the next chapters.

---

[57] At least, the version described here. Improvements that take this into account can be made.

Figure 8.4. Test portfolio contents and results. Top graph shows the effect of using Central Limit Theorem to approximate the unsystematic risk; the bottom figure shows the saddle-point approximation, which is seen to be substantially better and indeed is roughly uniformly accurate across the distribution.

Portfolio size	50
Asset-sector correlation	**0.4 - 0.6**
Average default probability (EDF)	**0.3% - 2%**
Exposure	**Average 3M, with a few very much higher exposures (10M - 20M) to better-grade credits**

Source: Credit Suisse First Boston
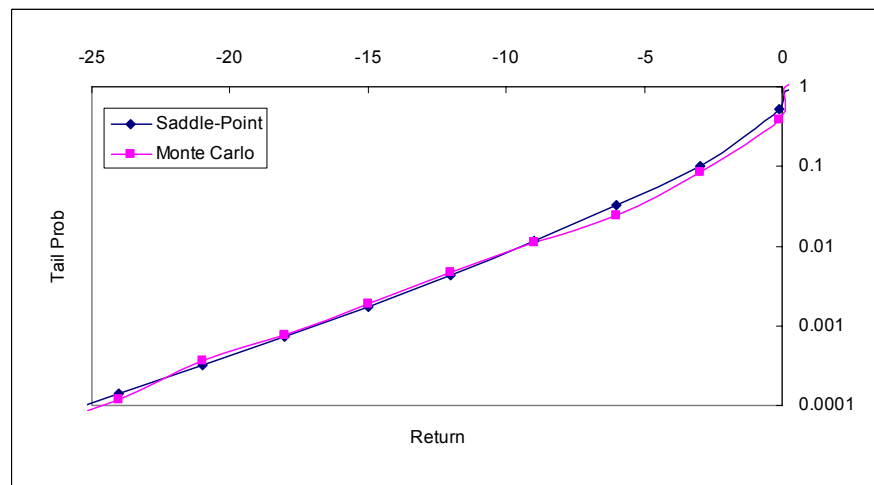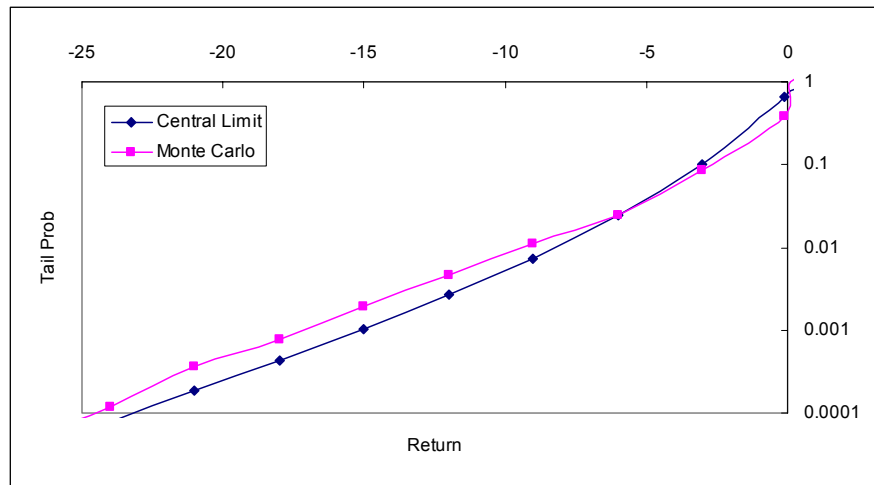
Source: Credit Suisse First Boston

Appendix: derivation of the granularity adjustment

We have

$$P(Y > y) = \int_{-\infty}^{\infty} \Phi\left(\frac{\mu(v) - y}{\sqrt{\sigma^2(v)}}\right) f_V(v) dv$$

and seek an approximation when $\sigma^2(v)$ is small. To this end, expand the $\Phi()$ term as a Taylor series[58]:

$$P(Y > y) \approx \int_{-\infty}^{\infty} \left\{ H(\mu(v) - y) + U\left[\frac{\partial}{\partial U} \Phi\left(\frac{\mu(v) - y}{\sqrt{U}}\right)\right]_{U=0} + \cdots \right\} f_V(v) dv$$

where we have adopted the following notation. $H(x)$ is the step function (0 if $x<0$, 1 if $x>0$). When there is no unsystematic risk, only this term comes into play because the portfolio loss is exactly $\mu(V)$: hence it exceeds y if and only if $\mu(V)$ does. In the other part we have written U for $\sigma^2(v)$. Tidying-up we have

$$P(Y > y) - P(\mu(V) > y) \approx \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial U} \Phi\left(\frac{\mu(v) - y}{\sqrt{U}}\right)\right]_{U=0} U f_V(v) dv.$$

The next step is to notice that $\Phi(...)$ is a solution of the diffusion equation:

$$\frac{\partial}{\partial U} \Phi\left(\frac{X}{\sqrt{U}}\right) = \frac{1}{2} \frac{\partial^2}{\partial X^2} \Phi\left(\frac{X}{\sqrt{U}}\right).$$

Hence

$$P(Y > y) - P(\mu(V) > y) \approx \frac{1}{2} \int_{-\infty}^{\infty} \left[\left(\frac{\partial}{\partial \mu(v)}\right)^2 \Phi\left(\frac{\mu(v) - y}{\sigma(v)}\right)\right]_{\sigma(v)=0} \sigma^2(v) f_V(v) dv$$

or equivalently

$$P(Y > y) - P(\mu(V) > y) \approx \frac{1}{2} \int_{-\infty}^{\infty} \frac{1}{\mu'(v)} \frac{\partial}{\partial v} \left[\frac{\partial}{\partial \mu(v)} \Phi\left(\frac{\mu(v) - y}{\sigma(v)}\right)\right]_{\sigma(v)=0} \sigma^2(v) f_V(v) dv$$

If we now integrate by parts the derivative gets transferred to the undifferentiated part of the integrand:

$$P(Y > y) - P(\mu(V) > y) \approx -\frac{1}{2} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \mu(v)} \Phi\left(\frac{\mu(v) - y}{\sigma(v)}\right)\right]_{\sigma(v)=0} \frac{d}{dv}\left[\frac{\sigma^2(v) f_V(v)}{\mu'(v)}\right] dv$$

---

[58] This step is a little dubious because of the misbehavior when σ=0.

The derivative of $\Phi()$ can be expressed in terms of the Normal density (which is the density function of the portfolio loss conditional on V):

$$P(Y > y) - P(\mu(V) > y) \approx -\frac{1}{2} \int_{-\infty}^{\infty} \left[\frac{1}{\sigma(v)} \phi\left(\frac{\mu(v) - y}{\sigma(v)} \right) \right]_{\sigma(v)=0} \frac{d}{d\mu(v)} \left[\frac{\sigma^2(v)f_V(v)}{\mu'(v)} \right] d\mu(v)$$

Note that we must assume a single risk factor model, so that there is a one-to-one correspondence between factor values V and expected losses $\mu(V)$: without this assumption, many different values of V correspond to one particular level of loss, and the argument fails. We now let $\sigma(v) \to 0$, and the left-hand part pulls out the contribution from $y=\mu(v)$: this is again because in that limit there is no difference between Y and $\mu(V)$. Also, we can identify $f_V(v)/\mu'(v)$ as the density of the infinitely granular portfolio's loss distribution, and write it $f_\infty(y)$ (this follows from the transformation formula for a density function, i.e., what is the density of $\mu(V)$, given the density of V). We have arrived at

$$P(Y > y) - P(\mu(V) > y) \approx -\frac{1}{2} \frac{\partial}{\partial y} \left[\sigma^2(v)f_\infty(y) \right]\Big|_{v=v^*, \mu(v)=y}$$

which is one form of the granularity adjustment. Here v^* is the value of V that "corresponds to" the VaR, i.e. $\mu(v^*) = y$.

We now let $\sigma(v) \to 0$, and the left-hand part pulls out the contribution from $y=\mu(v)$: this is again because in that limit there is no difference between Y and $\mu(V)$. We have arrived at

$$P(Y > y) - P(\mu(V) > y) \approx -\frac{1}{2} \frac{\partial}{\partial y} \left[\sigma^2(v)f_\infty(y) \right]\Big|_{v=v^*, y=\mu(v)}$$

which is one form of the granularity adjustment. Here v^* is the value of V that "corresponds to" the VaR, i.e. $\mu(v^*) = y$. Note that we must assume a single risk factor model: without this assumption, many different values of V correspond to one particular level of loss, and the argument fails.

The more common form is to express it in terms of Value at Risk rather than tail probability, which requires another step. Write $\overline{F}(y) = P(Y > y)$ for the upper tail probability of Y, and Q_p for the VaR for tail probability p ('Q' for quantile). Then $\overline{F}(Q_p) = p$. If some parameter θ say (here the level of unsystematic risk) is allowed to vary while keeping the tail probability constant, then by the chain rule of differentiation we have

$$\frac{\partial \overline{F}}{\partial \theta} + \overline{F}'(Q_p)\frac{dQ_p}{d\theta} = 0 .$$

The first term is the sensitivity of the tail probability; the second is the product of (minus the density) and the sensitivity of the VaR. In other words:

$$\Delta(\text{VaR}) = \Delta(\text{Upper tail prob}) \div \text{Density}.$$

So the adjustment in VaR is

$$Q - \mu(V^*) \approx -\frac{1}{2f_\infty(y)} \frac{\partial}{\partial y}\left[\sigma^2(v)f_\infty(y)\right]\Bigg|_{y=\mu(v)=VaR}$$

with Q denoting the VaR of the real portfolio and $\mu(V^*)$ the VaR of the infinitely granular portfolio.

A better and more versatile proof uses characteristic functions.

The granularity adjustment method can also be used in a more general context where a small perturbation is made to a distribution. For example, one may be in the situation where there is more than one risk factor in a model but a one-factor approximation works fairly well. In that case, the contributions from the remaining factors can be analyzed approximately using the method.

CREDIT SUISSE | FIRST BOSTON

The Quantitative Credit Strategist
Risk measures: how long is a risky piece of string?

9

Risk measures: how long is a risky piece of string?

In the previous chapters we have described most of the machinery necessary for the construction of the loss distribution for a portfolio of correlated credit-risky instruments. We are now going describe some of the calculations that can be performed once the distribution has been constructed. We begin by looking at a problem that is actively discussed in the risk management community: *risk measures*, or how to ascribe a single "risk figure" to a distribution. We conclude, on the basis of our tests of a few common risk measures, that "conditional Value at Risk", or CVaR, also known as Tail VaR and Expected Shortfall, is the only method with all the right properties.

Why think about risk measures?

The basic problem is this. When we model a risky security, or a portfolio of such securities, we do not know what the value will be at the horizon in question and therefore construct a distribution that shows the probability of any possible gain or loss. Of course, a distribution contains a vast amount of information, for example:

- mean (expected) loss,

- standard deviation of loss,

- probability of losing more than $A,

- probability of losing between $A and $B,

- expected payout in a layer of insurance (e.g., a CDO tranche) where the attachment and exhaustion points are $A and $B,

and so it cannot be represented as a single number. But at the end of the day one needs to come up with some "dollar measure" of an uncertain outcome as a method of communicating riskiness to regulators, senior management, investors, the press, or whoever. So, given that we are to come up with a single figure:

- What are the desirable properties that such a measure should have?

- Do the simple well-known measures satisfy these?

- Are there any better measures?

Examples of risk measures

We list below some of the commonly adopted methods of measuring risk.

Note: In what follows, *X*, *Y* and so on denote the gain or loss ("P&L"). We adopt the convention that losses are **negative**. Hence one thinks of the distribution (density function) as looking like this:

Loss ←——————————————————→ Profit
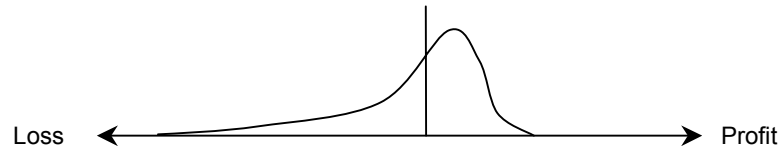
In the previous chapters we have concentrated on loss, and most of the graphs of loss distributions have the horizontal axis transposed. On the other hand we "generally" want risk to be positive, and to become more positive when the possibilities of downside loss increase, so there will be a few minus signs in the equations to effect this conversion.

- mean (expected) loss.

- standard deviation of loss.

- VaR (Value at Risk). We find the value *Q* such that $P(X<Q) = p$, for some given tail probability (such as 0.1% or 1%). The VaR (risk) is then $-Q$.

- CVaR (Conditional VaR; also known as Tail VaR and Expected Shortfall). As with VaR this requires a tail probability to be specified. One first finds the VaR (Q, say); then the CVaR is the expectation conditional on the loss exceeding *Q*, again with a minus sign put in: $CVaR = -E[X \mid X < Q]$. It is fairly clear that CVaR > VaR because all the events that give worse outcomes than a loss of *Q* are being averaged when the CVaR is computed.

Artzner's theory

In a ground-breaking paper[59], P. Artzner and co-authors came up with several axioms that, they stated, a sensible, or "coherent", measure of risk should satisfy. These are listed below. We have added an extended discussion of convexity and what seems to be a new notion of risk sensitivity (which is not in Artzner's paper, and is marked †).

Homogeneity
If the distribution is scaled by some multiple θ (assumed positive), then the risk is also scaled in proportion. In other words,

$$Risk(\theta X) = \theta \times Risk(X). \tag{1}$$

Virtually "any sensible" construction satisfies this, but notice that the variance does not: it multiplies by θ², not θ.

---

[59] P Artzner, F Delbaen, J-M Eber and D Heath (1999), "Coherent Measures of Risk," Mathematical Finance, 9, 203-228. See also article "Thinking coherently", RISK 10(11):68-71 (1997).

CREDIT SUISSE | FIRST BOSTON

The Quantitative Credit Strategist
Risk measures: how long is a risky piece of string?

Translation

If a riskless gain or loss is added to the portfolio, thereby shifting the distribution to the left or right but not altering its shape, then the risk shifts by the same amount. In other words,

$$\text{Risk}(X+A) = \text{Risk}(X) - A \qquad (A \text{ constant}). \qquad (2)$$

To some extent this is at variance (unintentional pun) with a notion of risk: risk is about uncertainty, so why does the equation not read

$$\text{Risk}(X+A) = \text{Risk}(X) \quad ? \qquad (2?)$$

The reason is that the original Artzner paper was about capital and capital allocation, which is more a measure of "downside" than risk *per se*. In any case the distinction is not very important: one can go from (2) to (2?) and back again by adding or subtracting the mean loss.

Monotonicity

If X and Y are such that $X \geq Y$ in all states of the world then $\text{Risk}(X) \leq \text{Risk}(Y)$. Again, we really have to think of risk as "downside", rather than uncertainty. One can easily construct theoretical examples (see later) in which X always outperforms Y, but is nevertheless much more volatile.

Subadditivity

For any two distributions X, Y, which we do NOT assume to be independent, the risk satisfies the inequality $\text{Risk}(X+Y) \leq \text{Risk}(X) + \text{Risk}(Y)$. This means that the risk of a portfolio is always less risky than the sum of the risks of its constituents. From the regulatory or risk management perspective this is important because a subadditive measure makes it impossible to "hide" risk by disaggregating it.

Much importance has been attached to this property, and many similar examples have been given showing that VaR does not satisfy the relation. For large, correlated portfolios the problem is unlikely to occur.

A corollary of subadditivity is **convexity**. The idea here is that the risk of a portfolio should be a convex function of asset allocation. This is the picture to bear in mind:
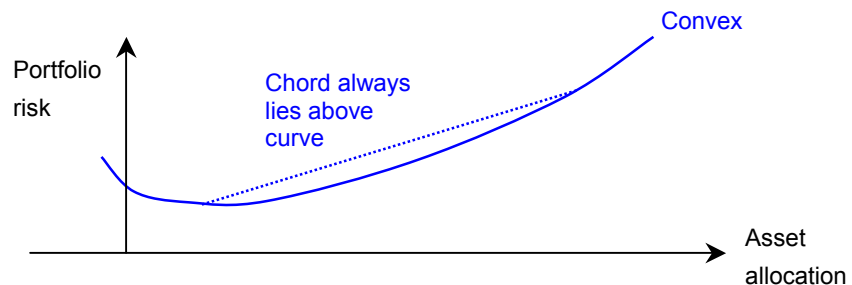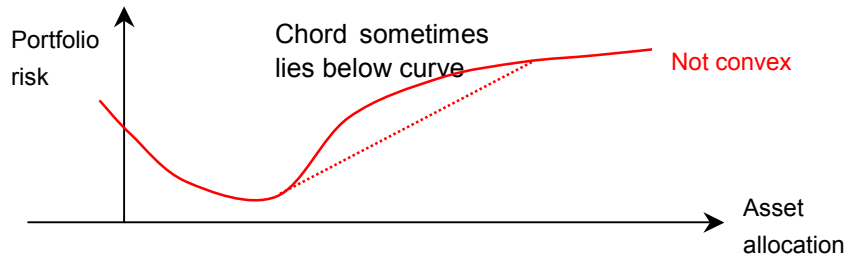
The graphs shown are for one particular asset allocation. Of course there are many assets to allocate in practice. For a subadditive risk measure we will have

$$\text{Risk}(\theta X + (1-\theta)Y) \le \theta\text{Risk}(X) + (1-\theta)\text{Risk}(Y) \tag{3}$$

If, as is often the case, the risk is a smooth function of the asset allocations, then the convexity condition can be written alternatively as a condition on the matrix of second partial derivatives of risk with respect to asset allocations[60]. This is usually the easiest way to show convexity.

The reason that convexity is important is that eq.(3) tells us that it is beneficial to diversify: a weighted sum of two portfolios is less risky than the weighted sum of their risks. It is then apparent that there is a unique optimal portfolio. To see this, suppose that there are two optimal portfolios X, and Y and consider their weighted average, $(\theta X + (1-\theta)Y$. Now the risk of that portfolio is given by (3), and the return is just the weighted average of the returns:

$$\text{Return}(\theta X + (1-\theta)Y) = \theta\text{Return}(X) + (1-\theta)\text{Return}(Y)$$

So

$$\frac{\text{Return}[\theta X + (1-\theta)Y]}{\text{Risk}[\theta X + (1-\theta)Y]} \ge \frac{\theta\,\text{Return}[X] + (1-\theta)\text{Return}[Y]}{\theta\,\text{Risk}[X] + (1-\theta)\text{Risk}[Y]} = \frac{\text{Return}[X]}{\text{Risk}[X]} = \frac{\text{Return}[Y]}{\text{Risk}[Y]},$$

the last equalities follow from the presumed optimality of X and Y. But then the return on risk of the diversified portfolio is higher than that of either X or Y, which is a contradiction. So there can only be one optimal portfolio. It also follows that for a subadditive risk measure the graph of Risk and against Return must be convex:
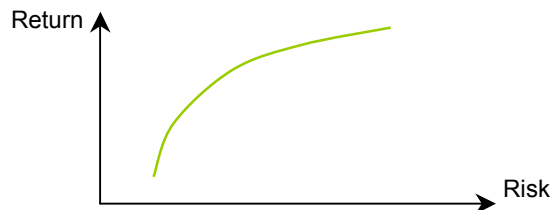
The green line is known as the *efficient frontier*, and (loosely speaking) all portfolios lie on or below it. The main thrust behind portfolio optimization is to find this frontier, which is why portfolio analytics are so important.

---

[60] The matrix must be positive semidefinite.

CREDIT SUISSE | FIRST BOSTON

The Quantitative Credit Strategist
Risk measures: how long is a risky piece of string?

† Risk Sensitivity

The notion of risk sensitivity comes from the conditional independence framework we have constructed in the previous chapters. Suppose that, conditionally on some factor V (say) the distribution of P&L has mean $\mu(V)$ and variance $\sigma^2(V)$. The variation in $\mu(V)$ as V varies is the systematic risk, and $\sigma^2(V)$ is the unsystematic risk. We would like the risk measure to be an increasing function of both of these risks, i.e., decreasing in $\mu(V)$ (remember $\mu<0$ means loss) and increasing in $\sigma(V)$. If we write the change in portfolio risk consequent on a small change in $\mu(V)$ and $\sigma(V)$ as

$$d(\text{Risk}) = S(V)\,d\mu(V) + U(V)\,d\sigma(V) \qquad (4)$$

then we say that the risk measure is risk-sensitive if

$$S(V) < 0 \text{ and } U(V) > 0.$$

We shall discuss more fully which measures are risk-sensitive and which are not, mainly with the aid of diagrams. Fig. 9.1 shows what we have in mind, with V discrete and taking four states (but the principle does not rely on V being discrete, so it could for example be multivariate Normal, as in the Gaussian models that we have already talked about). In essence we want the risk to increase if any of the conditional distributions are shifted to the left (μ dependence) or stretched out (σ dependence). There are some surprises in store!

Fig. 9.1. Diagram of loss distribution for discussion of risk sensitivity.

CREDIT SUISSE | FIRST BOSTON

The Quantitative Credit Strategist
Risk measures: how long is a risky piece of string?
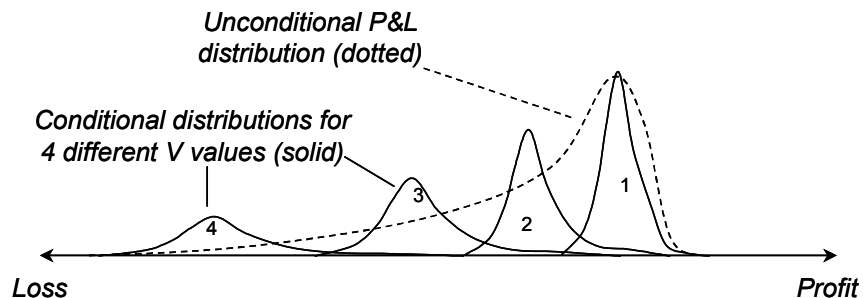
Summary table

The table shows which of the common risk measures satisfy which properties.

Notes (Subadditivity and Risk sensitivity are discussed in detail later)

- The standard deviation is unaltered by the addition of a risk-free asset, but as we said before this can be fixed by adding the expected (mean) loss.

- Fig. 9.2 below gives an example in which X always gives a more favorable result than Y, but is

	Homogeneity	Translation	Monotonicity	Subadditivity	Risk sensitivity
Mean	✔	✔	✔	✔[3]	✘
Std dev	✔	(✔)[1]	✘[2]	✔	✘
VaR	✔	✔	✔	✘	✘
CVaR	✔	✔	✔	✔	✔

Source: Credit Suisse First Boston

more risky. This is not a very important issue, though: in practice most distributions are "centered," then it is not possible to construct an example in which X always outperforms Y.

- The mean is linear in the asset allocation, from which subadditivity follows. But the convexity is never strict (i.e., one always has equality in equation (3)), and hence it is useless for risk-return optimization.

Fig. 9.2. X is always more favorable than Y, but a "mean loss + some number of standard deviations" measure may say that X is riskier. This is not a very important point.

<div style="text-align: right">

CREDIT SUISSE | FIRST BOSTON

The Quantitative Credit Strategist
Risk measures: how long is a risky piece of string?

</div>
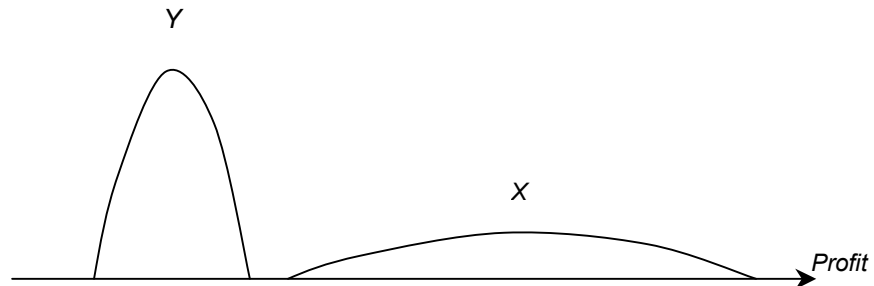
Subadditivity, convexity and risk sensitivity in more depth

In this section we explain the assertions in the table.

Standard deviation is subadditive

To show $\sigma_{X+Y} \le \sigma_X + \sigma_Y$, square both sides, then use $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$ and the fact that ρ (the correlation) cannot exceed 1.

VaR is not subadditive

Let X_1 and X_2 be single-asset portfolios in which the asset is a defaultable bond of notional \$1M and default probability 0.75%. Assume that default events are independent and consider only a default/no-default model. Assess VaR at 99% confidence (tail probability 1%). For either of the portfolios the VaR is zero because the VaR does not look at the loss beyond the 1% level, whereas if the two portfolios are combined into one the VaR is equal to \$1M (the probability of two defaults is too unlikely, and the probability of one or more defaults is roughly 2×0.75% > 1%). There is therefore a temptation to disaggregate the risk, i.e., rather than go for a diversified portfolio, to just have a large exposure to a single asset. Other examples can be constructed on the same principle, but for large correlated portfolios it is much more difficult to construct realistic examples in which subadditivity fails.

CVaR is subadditive

The proof of subadditivity of CVaR is a little fiddly, in our opinion, so we shall omit it. However, a different perspective is obtained by concentrating on convexity. The second derivative of portfolio CVaR with respect to asset allocations is given as the covariance of the assets given that the portfolio loss is equal to the VaR: $\partial^2 CVaR / \partial a_i \partial a_j = \text{Cov}[X_i, X_j \mid Y = VaR]$, where $Y = \sum a_i X_i$. Convexity then follows directly. This is quite a powerful result.

Mean is not risk-sensitive

The mean is simply the weighted average (using the state probabilities of V as weights) of the conditional means. Hence it exhibits the correct sensitivity to $\mu(V)$. However it does not depend on $\sigma(V)$. So it is insensitive to unsystematic risk, and therefore insensitive to concentration risk. This is a very serious objection.

Standard deviation is not risk-sensitive

Standard deviation has the opposite problem. If we write the variance of the portfolio through the ANOVA (Analysis of Variance) formula we have

$$\mathbf{V}[Y] = \mathbf{V}[\mu(V)] + \mathbf{E}[\sigma^2(V)]$$

($\mathbf{V}$ denotes variance, V is the systematic risk factor). So clearly the portfolio variance increases with increasing unsystematic risk ($\sigma(V)$). But reducing $\mu(V)$ does not necessarily increase the variance: in Fig.9.1, if we shift conditional distribution #4 to the left then the portfolio variance increases because the distributions are more spread out, whereas if we shift conditional distribution #1 to the left then the portfolio variance decreases because the distributions are more bunched up.

VaR is not risk-sensitive

VaR has the right sensitivity to μ(V), as it is fairly clear that shifting any distribution to the left will cause the VaR to increase (fix the tail probability; then the associated level of P&L moves to the left). But increasing σ(V) may reduce the VaR, depending on the threshold. Referring to Fig. 9.3, we can see that, for a fixed tail probability, if the threshold is at *A*, fattening conditional distribution #4 will push the threshold to the left (so the VaR increases, as desired) but if it is at *B* then the threshold will be pushed to the right (so the VaR decreases).
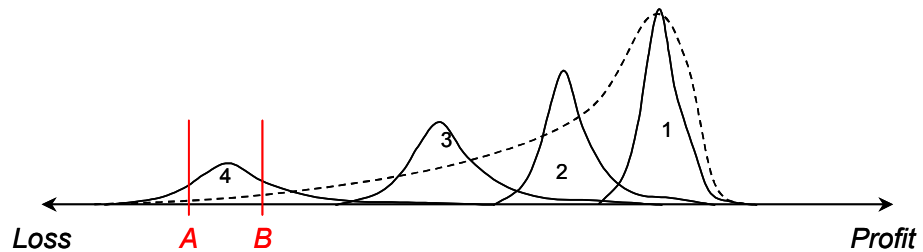
Fig. 9.3. "Proof" that VaR is not risk-sensitive.

CVaR is risk-sensitive

To our knowledge this is a new result and the proof is rather awkward, so we ask the reader to take it on trust. It is worth considering why CVaR is an improvement over VaR, however: when the threshold is at *B*, as just discussed, the VaR reduces (point *B* moves to the right) when conditional distribution #4 is fattened, but the mass of the distribution to the left of *B* increases much more. It is then a matter of proving that the increasing mass does indeed have a bigger effect, causing the CVaR to increase even though the VaR decreases.

Conclusion

We have discussed the various properties that a risk measure ought to have and found that the traditional mean-variance framework, as well as the VaR framework, have shortcomings. However it can be shown that CVaR (expected shortfall) overcomes all of these. That is not to say, of course, that it has its own problems that we have not yet identified, or that there are other coherent alternatives to CVaR. However, its properties make it an excellent risk measure.

10

Portfolio optimization: the importance of convexity

In this chapter on portfolio modeling, we embark on a study of *risk contributions*. The risk contribution of an asset in a portfolio is fundamental, because without it, one cannot perform portfolio optimization. The classical mean-variance theory was developed by Markowitz in the 1950s, and we shall consider this first before moving on to the quantile-based risk measures, such as VaR and expected shortfall, which are more suitable for the skewed loss distributions that arise in credit risk. The chapter on risk measures was slanted towards portfolio optimization, and we shall pick up there.

Risk and reward: the traditional view

It is generally appreciated that a well-diversified portfolio is, in a sense, "better" than an ill-diversified one because it offers "the same expected return for less risk." To make this a bit more precise, we need to be a bit careful about how we measure risk. Take the following hypothetical example of equities:

Asset	Expected return	Standard deviation
ABC Inc.	8%	20%
DEF Inc.	25%	50%

Source: Credit Suisse First Boston

To simplify matters, we shall assume that the equities are normally distributed and that their correlation is 25%. Now look at three different portfolios:

Portfolio	Allocation ABC	Allocation DEF	Value today	Expected return	Risk (stdev)	Return ÷ Risk
1	$20M	0	$20M	$1.6M	$4M	0.4
2	0	$20M	$20M	$5.0M	$10M	0.5
3	$10M	$10M	$20M	$3.3M	$5.8M	0.57

Source: Credit Suisse First Boston

By construction, all three portfolios have the same price. The expected returns are easy to calculate – e.g., for portfolio 3:

> 5M x $2 x 8% + 10M x $1 x 25% = $3.3M.

The risk is a little harder. When correlated assets are added, their standard deviations add according to the following rule:

$$\sigma_{X+Y} = \sqrt{\sigma_X{}^2 + \sigma_Y{}^2 + 2\rho_{XY}\sigma_X\sigma_Y}$$

with ρ denoting correlation. It is apparent that Portfolio 2 has the highest expected return, but Portfolio 1 is safest (lower risk). Meanwhile, Portfolio 3 offers the highest return on risk: its return is the weighted-average return of Portfolios 1 and 2, but its risk is less than the average risk because of diversification.

By plotting risk against each other for different portfolios (all of which have the same present value), we obtain a well-known graph. The three portfolios given above are shown:
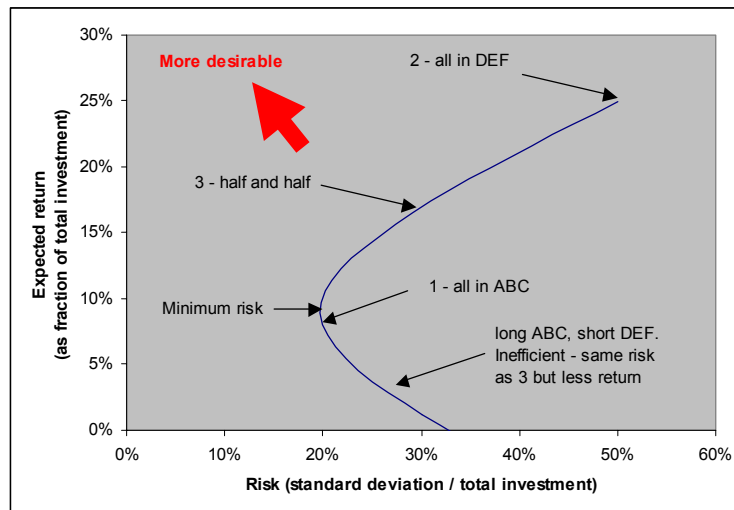
As can be seen, there is a minimum level of risk (about 20%, with a return of about 9%). This portfolio contains mainly ABC (the less risky equity), with only a small amount of DEF. To increase the expected return requires more risk to be taken on. If (as seems reasonable) we increase the allocation in DEF, we move upwards and to the right, with a riskier but better-rewarded portfolio. If, by contrast, we take a short position in DEF, we move down the lower branch of the curve and end up with a portfolio that is riskier and has a worse expected return.

Let us introduce risk-free borrowing and lending. The investment possibilities for a combination of a risk-free and a risky asset lie along a line connecting the two assets. In our figure, the frontier of most desirable investment possibilities is defined by the most "northwest" (highest expected return in relation to risk) investment opportunities that are achievable. That frontier is outlined by the most "northwest" line that connects the risk-free rate to the risky investment set. So, to find the portfolio with the highest return on risk, we draw a straight line from the vertical axis (intersecting at the risk-free rate) that touches the curve. The slope (gradient) gives the ratio of excess return to risk; this is also known as the Sharpe ratio. The point of tangency indicates the portfolio of risky assets that has the highest return to risk. For convenience, we shall call this the Sharpe

portfolio. By mixing this optimal portfolio with risk-free asset in different proportions, one ends up with a choice of portfolios with the same Sharpe ratio, and all are optimal in the sense that they dominate[61] all other achievable portfolios.
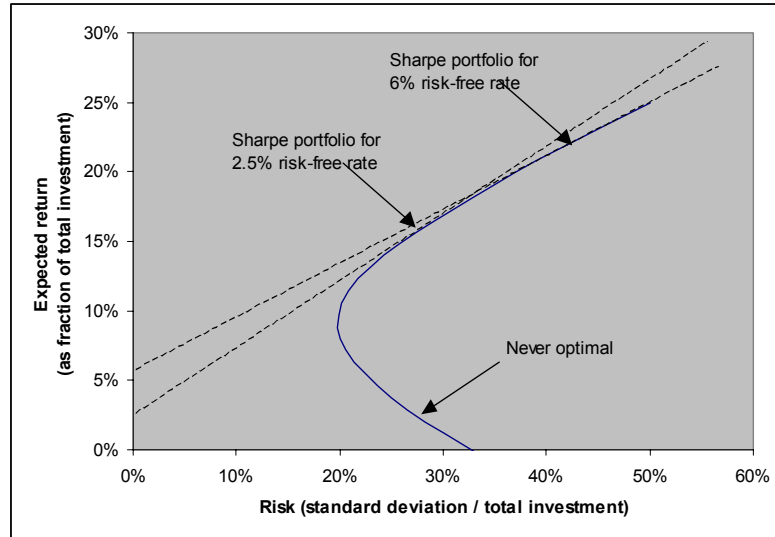
The higher the risk-free rate, the higher the proportion invested in the riskier asset (or, in more general situations, assets), because the assets with less return become relatively less attractive as the risk-free rate increases.

Does this work with other risk measures?

The construction that we have made works so nicely because of the *convexity* of the region on the risk-return plot occupied by all portfolios that contain only risky assets and have equal value. This is equivalent to saying that the upper part of its boundary, which is known as the *efficient frontier of the set of risky assets*, is convex. By a convex region or boundary, we mean that any tangent to it lies wholly on one side (the "outside")

Convex   *Not Convex*
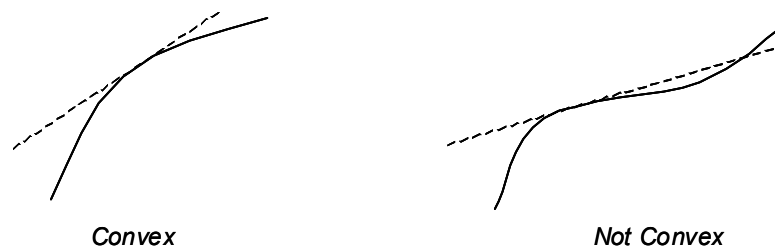
or equivalently any chord lies on the "inside".

Convex   *Not Convex*

---

[61] Have lower risk for given return, or higher return for given risk.

Convexity is important for two reasons:

- Lack of convexity implies more than one Sharpe portfolio, which is a nuisance from the point of view of computation.

- Lack of convexity implies that the risk measure is not subadditive (as per last chapter's discussion), which in turn means that the investor's view of risk is not necessarily logical.

To see the second point, study the diagram below. If the risk measure is subadditive, then a portfolio consisting of some weighted combination of A and B (e.g., 25% A and 75% B) must have less risk than the weighted average of A's risk and B's risk, so it lies to the left of the chord A-B. This gives the required convexity and uniqueness of the optimal portfolio.

Return

A

B
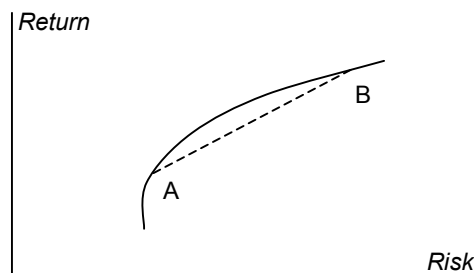
Risk

As we have seen before, VaR does not satisfy the subadditivity property, but expected shortfall (ESF, CVaR) does. Hence it is quite in order to use expected shortfall as a risk measure for optimization. That does not mean that using VaR will always cause problems: for example, if the joint distribution of the assets is multivariate normal, then VaR, standard deviation and ESF give the same results. However, VaR is not recommended in general.

Risk contributions

A further desirable property of risk measures, as indicated in Chapter 9, is 1-homogeneity. This means that if all the allocations are scaled, then the portfolio risk R is scaled by the same factor. In other words,

$$R(\theta a_1, \ldots, \theta a_n) = \theta R(a_1, \ldots, a_n).$$

By taking the total differential with respect to θ and setting $\theta=1$, we deduce

$$\sum_{j=1}^{n} a_j \frac{\partial R}{\partial a_j} = R.$$

It is fairly clear that in order to optimize a portfolio, one must know the sensitivity of risk to asset allocation: hence, the first derivative is necessary. What the above equation shows is that the sensitivities of risk to allocation (we call these the *deltas*), when multiplied by the asset allocations, give quantities (that we call *risk contributions*) that add up to the portfolio risk. This is nice because it gives a way of attributing the risk to the assets "without any risk going missing."

Now let us denote by R the risk of a portfolio, a_j the allocation to the jth asset and r_j the expected return of the jth asset. Then the Sharpe ratio is

$$\frac{\sum_{j=1}^{n} a_j r_j}{R} .$$

If we differentiate with respect to a_j and set the derivative equal to zero[62], we obtain

$$\frac{\sum_{j=1}^{n} a_j^* r_j}{R} = \frac{r_j}{\partial R / \partial a_j^*} \quad \text{or} \quad \frac{a_j^* \partial R / \partial a_j^*}{R} = \frac{a_j^* r_j}{\sum_{j=1}^{n} a_j^* r_j}$$

(stars denote optimal allocation). The first of these equations states that the returns of the assets in the optimal portfolio are in equal proportion to their deltas. The second says, equivalently, that each asset's contribution to risk (the left-hand side) is equal to the proportion of the return that it generates (right-hand side). So, for the optimal portfolio, each asset "pulls its weight" by generating enough return to compensate its risk.

In fact, the risk contributions of the *optimal* portfolio must add up to the portfolio risk whether or not the risk measure is homogeneous. This is seen by summing the second equation over all assets:

$$\sum_{j=1}^{n} a_j^* \frac{\partial R}{\partial a_j^*} = R .$$

Convexity

However, the picture is nowhere near complete. Suppose that we wish to try and optimize a portfolio knowing only the first-order sensitivities. We begin by writing the expected return (or gain, G) as a function of the asset allocations:

$$G = G_{current} + \sum_{j=1}^{n} \Delta a_j r_j$$

We do the same for the risk, although risk is not generally linear and so we can only approximate:

$$R \approx R_{current} + \sum_{j=1}^{n} \Delta a_j \frac{\partial R}{\partial a_j}$$

Now, we try to maximize G/R. Let us just concentrate on maximizing with respect to one variable. We immediately find, as G/R is of the form (linear $\div$ linear) in Δa_j, that it cannot have a maximum. This is because a function $y = \frac{ax + b}{cx + d}$ typically looks like this:

---

[62] We assume that this finds the maximum rather than the minimum Sharpe ratio!
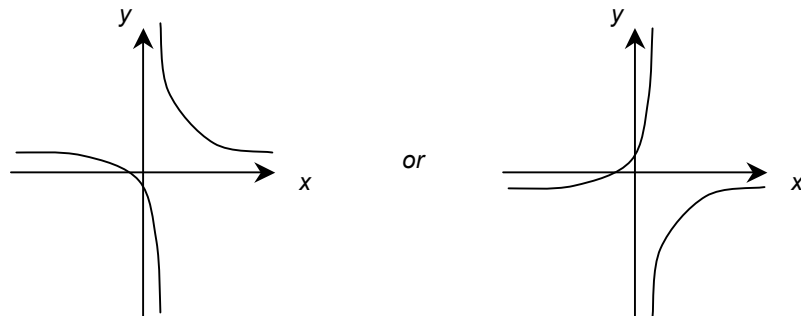
So no maximum exists, and we cannot even make a sensible guess as to what the optimal allocations are. In fact, all that the first-order sensitivities tell us is which assets should be increased and which reduced. They do not say how much to adjust the allocations. To cure this problem, we have to expand the risk to a higher order. The quadratic approximation would be

$$R \approx R_{current} + \sum_{j=1}^{n} \Delta a_j \frac{\partial R}{\partial a_j} + \tfrac{1}{2} \sum_{i,j=1}^{n} \Delta a_i \Delta a_j \frac{\partial^2 R}{\partial a_i \partial a_j},$$

but rather than going through the algebra, let us consider what happens when we confine ourselves to varying one asset only. By recognizing the curvature of the risk, we obtain a more reasonable picture of what is going on:
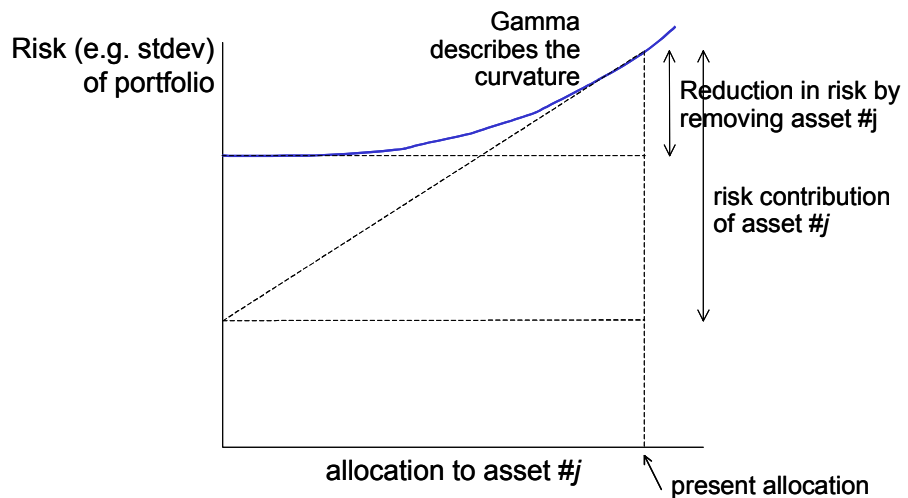
When the returns are brought into play, the Sharpe ratio can then be analyzed. Suppose that the present allocation to asset j is too high (in relation to its return). Then the situation is seen to be something like this:
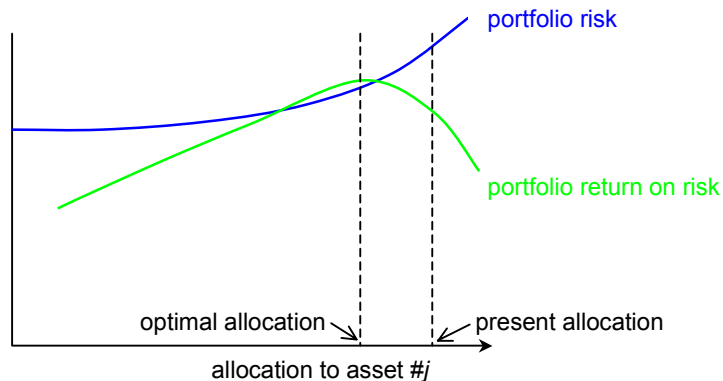
Without the second derivative $\partial^2 R / \partial a_j^2$, the line marked "portfolio return on risk" cannot have a local maximum, which is why the previous analysis failed.

Even the second derivative of portfolio risk with respect to each asset allocation is not the complete picture. First, the variation of risk with allocation is not quadratic (even in the mean-variance framework[63]). Second, there are many assets to be adjusted, and so the full matrix of partial derivatives $\partial^2 R / \partial a_i \partial a_j$ is required. This is necessary if we are to understand what happens when several assets are adjusted at the same time. (As we shall see in later work, the effects are not always easy to predict.)

An example of linear (non-convex) optimization

If we use expected loss as a risk measure, then the efficient frontier is straight rather than curved. This is because expected loss does not take diversification into account; the expected loss of a portfolio can be computed without regard to the correlations between the assets. Consequently, an optimal portfolio in the sense of return on risk cannot exist.

That said, there are some interesting problems that arise from use of expected loss as a risk measure, and they give rise to a fundamentally different type of optimization. Although it is possible to demonstrate one of these for an equity portfolio, a rather more natural construction is a default-only model of a loan or bond portfolio.

Asset	Credit rating	One-year expected default rate	Spread to LIBOR
ABC Inc.	High	20bp	50bp
DEF Inc.	Low	100bp	300bp

Source: Credit Suisse First Boston

---

[63] The risk measure is the standard deviation, not the variance. So although the variance is exactly quadratic, the standard deviation isn't.

We shall not need to bother with correlations here (the analysis is purely based on expected loss). Consider again some portfolios:

Portfolio	Allocation ABC	Allocation DEF	Value today	Expected return	Risk (EDF)	Return ÷ Risk
1	$20M	0	$20M	$60K	$40K	1.5
2	0	$20M	$20M	$400K	$200K	2.0
3	$10M	$10M	$20M	$230K	$120K	1.92

Source: Credit Suisse First Boston

Zero recovery rate is assumed. By expected return, we mean spread return minus expected loss. Hence, for ABC, the expected return is 30bp. By risk, we now just mean expected loss (expected default frequency) times asset allocation.

As before, Portfolio 1 has a lower risk and return than Portfolio 2. But Portfolio 3, despite being intermediate in terms of risk and return, does not offer the best return on risk: Portfolio 2 does that. Furthermore, we can get better and better return on risk as follows: long $30M DEF and short $10M ABC gives a return on risk of $570K/$280K=2.04, and lengthening the position in DEF while shorting ABC causes this figure to increase without limit. There is no penalty for concentration risk! To make sense, these types of optimization *must* be constrained. If we stipulate that shorting is not allowed and that the total amount for investment is <$20M, then we have the following picture:
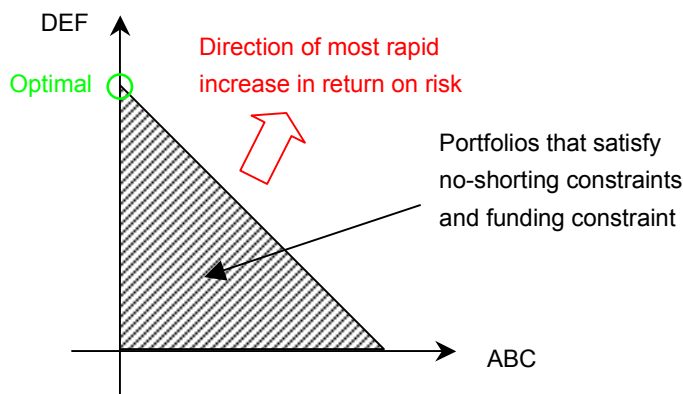
Notice that the solution is at a corner: this is always the case for this kind of linear optimization. It is also very important to notice that the solution is not stable to small variations in the model parameters. Suppose that the expected loss of ABC is adjusted to 16bp. Now, the return on risk for Portfolio 1 is 2.125, Portfolio 2's remains 2 and Portfolio 3 is still in between. So now, Portfolio 1 is the optimal allocation, and the optimal solution has jumped to the other corner. What has happened (referring to previous and next diagrams) is that the arrow indicating the "preferred direction" has rotated clockwise a little.

DEF

Direction of most rapid
increase in return on risk

ABC
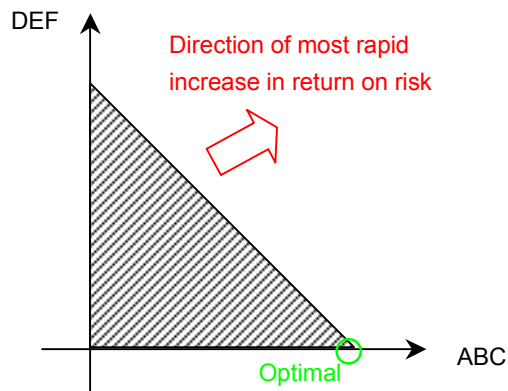
Optimal

One therefore has to be rather careful, as the optimal portfolios produced can have strange properties. They are called "corner solutions."

Conclusions

We have discussed essentially by means of diagrams what happens when a portfolio is optimized and shown that it is important to be able to obtain the first- and second-order derivatives of risk with respect to asset allocation.

One of the most important recent developments in portfolio theory has been the derivation of these sensitivities in frameworks other than the traditional mean-variance framework. However, we shall keep things simple for the moment, and in the next chapters, we shall begin with the mean-variance framework, in more mathematical depth than the discussion we have made here.

THIS PAGE IS INTENTIONALLY LEFT BLANK

11

An advanced approach to correlation

In this chapter we shall discuss methods of measuring and estimating correlation between variables. We begin with the simple Pearson correlation coefficient and show some of its deficiencies before moving on to describe more robust methods that are suitable for use on data that are non-Gaussian, noisy, or both. Next we show how to apply these techniques, combined with Bayesian inference, to the calibration of correlations in the structural (CUSP) model. We finish off with examples from back-testing this new methodology.

Correlations between pairs of random variables

Linear correlation coefficient

The traditional measure of dependence is the Pearson linear correlation coefficient. In this, the correlation between two random variables *X* and *Y* is given by

$$\rho = \frac{E[XY] - E[X]E[Y]}{\sqrt{V[X]V[Y]}}.$$

The linear correlation is the natural one to use for Gaussian data. This is because if the joint distribution of (*X*, *Y*) is bivariate Normal then its density is given by

$$\text{const} \times \exp\left\{ -\tfrac{1}{2}\left[\left(\frac{x - \mu_X}{\sigma_X}\right)^2 + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) \right] \right\}$$

with ρ defined as above. When this is the case, *Y* and *X* are linearly related in the sense that

$$\frac{Y}{\sigma_Y} = \rho\frac{X}{\sigma_X} + U \quad (U \text{ uncorrelated with } X)$$

and reciprocally

$$\frac{X}{\sigma_X} = \rho\frac{Y}{\sigma_Y} + U^* \quad (U^* \text{ uncorrelated with } Y).$$

With both these expressions the variance of *U* (or *U**) vanishes as the correlation is made stronger (closer to ± 1).

In more general applications there are several disadvantages of the linear correlation coefficient. These are:

- It is not generally true that the correlation between X and Y is the same as that between $f(X)$ and $g(Y)$, where f and g are some functions[64]. In other words, linear correlations are altered by distortion. This is because the distortion destroys linearity.

- The correlation does not exist if X or Y have infinite variance. Hence, it is not suitable for very fat-tailed distributions. Although this on its own would not be a particularly good reason to reject the linear correlation entirely, the fact that it involves squaring of the data points causes problems if there is even a small amount of contamination: outliers are greatly amplified[65].

The theory of *robust statistics* seeks to redress the latter problem, and there are other types of correlation measure that redress both. Essentially, techniques can be divided into two categories, as follows.

Parametric approaches

These seek to write down an alternative model for the density than the bivariate Normal. To start off, let us rework the Normal case. The density of the bivariate Normal distribution can be written

$$\frac{1}{\sqrt{2\pi \det C}} \exp\left\{-\tfrac{1}{2}\begin{bmatrix} x & y \end{bmatrix} C^{-1} \begin{bmatrix} x \\ y \end{bmatrix}\right\}$$

in which C is the covariance matrix. In estimating the covariance matrix from sample data $(x_i, y_i)_{i=1}^{N}$, the log-likelihood function (i.e. logarithm of the probability of observing the data given C) is

$$L = \ln\left\{\left(\frac{1}{\sqrt{2\pi \det C}}\right)^{N} \prod_{i=1}^{N} \exp\left\{-\tfrac{1}{2}\begin{bmatrix} x_i & y_i \end{bmatrix} C^{-1} \begin{bmatrix} x_i \\ y_i \end{bmatrix}\right\}\right\}$$

which simplifies to

$$L = \text{const} - \tfrac{1}{2} N \ln \det C - \tfrac{1}{2} \sum_{i=1}^{N} \begin{bmatrix} x_i & y_i \end{bmatrix} C^{-1} \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

By differentiating with respect to C and setting the derivative to zero we find the covariance matrix that "most likely" fits the observed data, and this is given by the "expected" result:

$$\hat{C} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \begin{bmatrix} x_i & y_i \end{bmatrix}.$$

---

[64] This point will be seen to be particularly relevant later, in the discussion of equity, debt and firm value returns, which are nonlinearly related. Similarly, if one models default risk by assuming that default occurs when the firm value falls below some threshold, one finds that the correlation between the firm values is unequal to the correlation between the default events. This is because the default indicator variable is a nonlinear transformation of the firm value return.

[65] See "Linear Correlation Estimation", F. Lindskog: www.math.ethz.ch/~lindskog. We demonstrate this in a moment.

As we said before it is the squaring that causes problems when the data are non-Gaussian. From the point of view of statistical theory, a large deviation in a Gaussian model is so unlikely that it forces the estimator to "sit up and take notice", thereby throwing the estimation away from the true result.

It follows that one remedy is to choose a more fat-tailed distribution than the Gaussian, as this will be more tolerant to outliers (it says that they are not so unlikely). An obvious candidate is the Student t distribution, which has density

$$\frac{1}{\sqrt{2\pi \det C}}\left(1 + \frac{1}{\nu}\begin{bmatrix} x & y \end{bmatrix} C^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right)^{-\frac{\nu+2}{2}}$$

Here ν is the number of "degrees of freedom" (in the limit $\nu \to \infty$ the Gaussian distribution is recovered). The log-likelihood function is now

$$L = \text{const} - \tfrac{1}{2}N \ln \det C - \tfrac{\nu+2}{2}\sum_{i=1}^{N} \ln\left(1 + \frac{1}{\nu}\begin{bmatrix} x_i & y_i \end{bmatrix} C^{-1}\begin{bmatrix} x_i \\ y_i \end{bmatrix}\right)$$

and on setting the derivative to zero we find the maximum likelihood estimate to be

$$\hat{C} = \frac{1}{N}\frac{\nu+2}{\nu}\sum_{i=1}^{N}\frac{\begin{bmatrix} x_i \\ y_i \end{bmatrix}\begin{bmatrix} x_i & y_i \end{bmatrix}}{1 + \frac{1}{\nu}\begin{bmatrix} x_i & y_i \end{bmatrix} \hat{C}^{-1}\begin{bmatrix} x_i \\ y_i \end{bmatrix}}$$

This equation is awkward to solve because C (the solution) appears on both sides. However it can be solved by iteration. Although we cannot write down an explicit answer, we can see the effect of outlying contaminating points: they occur on both the bottom of the fraction as well as the top, and hence their effect is attenuated. The attenuation is strongest for small ν, which is not surprising since that is when the distribution is most fat-tailed. When $\nu \to \infty$ we are back with the Gaussian result.

This method is known as *M-estimation*.

Non-parametric approaches

These seek to make robust correlation estimates without having to specify a distribution. The first possibility that we shall discuss is to use the Spearman's rank correlation coefficient. The idea behind this is that rather than working with the raw data we put them in rank order and then correlate the ranking numbers[66]. The obvious attraction of doing this is that the correlation measure remains invariant under distortion. Provided that the transformations f and g are strictly increasing functions, they do not affect the rankings. This transformation-independence is common in nonparametric techniques.

Another idea is known as Kendall's tau. This is defined as

$$P[(X - X')(Y - Y') > 0] - P[(X - X')(Y - Y') < 0]$$

---

[66] It is possible to boil this down to a simple expression that just uses the squared differences in ranking numbers. See e.g. Press et al, "Numerical Recipes in C++", CUP, which discusses Kendall's tau as well.

in which (X', Y') is an independent copy of (X, Y). In other words this compares pairs of points on a scatter-plot. If two points are in this orientation $\begin{bmatrix} & \bullet \\ \bullet & \end{bmatrix}$ then they contribute to the first term in the above expression and not to the second, whereas the converse is true if they are orientated $\begin{bmatrix} \bullet & \\ & \bullet \end{bmatrix}$. So the difference measures on average how aligned in a top-right-to-bottom-left sense they are. The range of tau is −1 to 1. Again it is clear that tau is transformation-independent because it does not alter relative position.

The Spearman's rank correlation is equal to the linear correlation when the data are uniformly distributed (essentially because the ranking operation has no effect on the distribution). Neither measure is equal to the linear correlation for bivariate Gaussian data though, as for Spearman we have

$$\rho_S = 12 \iint\limits_{-\infty < x, y < \infty} (\Phi(x) - \tfrac{1}{2})(\Phi(y) - \tfrac{1}{2})\phi_2(x, y; \rho)\, dx\, dy = \frac{6}{\pi}\arcsin\frac{\rho}{2} < \rho$$

(though the difference between ρ_S and ρ is never substantial) while Kendall's tau is

$$\tau = 4 \iint\limits_{0 < x, y < \infty} \phi_2(x, y; \rho)\, dx\, dy - 1 = \frac{2}{\pi}\arcsin\rho < \rho .$$

As an example we start with some clean Gaussian data[67] (Figure 11.1a) taken from a bivariate Normal distribution in which both margins are Normally distributed with mean zero and variance 1, and the correlation is 0.71. From 25 data points the estimation using the standard estimator is quite reasonable (0.77). Spearman's rank is ρ_S=0.70, which on transformation to an equivalent linear Gaussian correlation is $2\sin(\pi\rho_S/6)$=0.72. Kendall's tau comes out as 0.55, which on transformation to an equivalent linear Gaussian correlation is $\sin(\pi\tau/2)$=0.76.

So all three methods come out roughly the same on clean data. Let us now consider what happens when we contaminate the data by corrupting the fourth data point, moving it to (2.9,−2.9). The effect on the standard estimator is quite pronounced, taking the estimate down to 0.34. The Spearman's rank correlation, on the other hand, is altered by a smaller amount, 0.55 (linear Gaussian correlation of 0.57), and the Kendall's tau changes still less (to 0.47, or a linear Gaussian correlation of 0.67). Because the non-parametric methods compare the relative sizes of the data points rather than their absolute values, they are considerably more robust to contamination. See Figure 11.1b.

The non-parametric approaches are considerably easier to implement and we shall therefore use them in practice.

---

[67] Generated using a random number generator.

Fig. 11.1a. Gaussian data (true correlation 0.71) before corruption.

Estimators: Standard ρ=0.77; Spearman ρ_S=0.70, ρ=0.72; Kendall τ=0.55, ρ=0.76.

	X	Y	X rank	Y rank
1	2.484	1.826	1	3
2	-0.882	-1.183	24	24
3	-0.812	-0.515	22	22
4	0.202	-1.056	11	23
5	-0.458	-0.303	17	18
6	-0.682	-0.449	20	21
7	0.318	1.835	9	2
8	-1.915	-1.376	25	25
9	1.909	1.609	2	4
10	0.288	-0.441	10	20
11	-0.663	-0.341	19	19
12	1.412	1.424	5	6
13	1.728	0.646	4	8
14	-0.092	0.223	16	10
15	0.704	-0.141	7	16
16	1.158	2.070	6	1
17	0.068	1.075	12	7
18	-0.628	-0.149	18	17
19	0.435	-0.139	8	15
20	0.035	0.044	13	12
21	-0.834	0.040	23	13
22	-0.730	0.045	21	11
23	-0.046	-0.114	14	14
24	-0.061	0.428	15	9
25	1.745	1.499	3	5
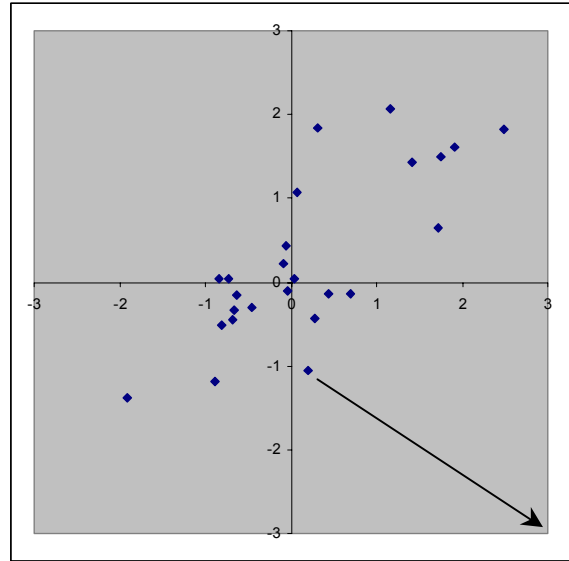
Source: CSFB

Fig. 11.1b. Gaussian data (true correlation 0.71) after corruption.

Estimators: Standard ρ=0.34; Spearman ρ_S=0.55, ρ=0.57; Kendall τ=0.47, ρ=0.67.

	X	Y	X rank	Y rank
1	2.484	1.826	2	3
2	-0.882	-1.183	24	23
3	-0.812	-0.515	22	22
4	2.900	-2.900	1	25
5	-0.458	-0.303	17	18
6	-0.682	-0.449	20	21
7	0.318	1.835	10	2
8	-1.915	-1.376	25	24
9	1.909	1.609	3	4
10	0.288	-0.441	11	20
11	-0.663	-0.341	19	19
12	1.412	1.424	6	6
13	1.728	0.646	5	8
14	-0.092	0.223	16	10
15	0.704	-0.141	8	16
16	1.158	2.070	7	1
17	0.068	1.075	12	7
18	-0.628	-0.149	18	17
19	0.435	-0.139	9	15
20	0.035	0.044	13	12
21	-0.834	0.040	23	13
22	-0.730	0.045	21	11
23	-0.046	-0.114	14	14
24	-0.061	0.428	15	9
25	1.745	1.499	4	5
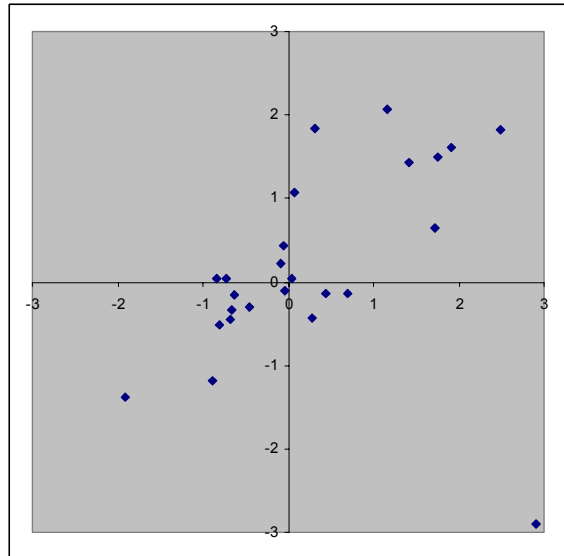
Source: CSFB

Time series and the differencing problem

Now that we have investigated methods of correlating time series together we need to understand correlation of time series. Consider a pair of time series $X(t)$ and $Y(t)$. By saying that they are correlated, what we actually are considering is the correlations between their returns, i.e.

$$\text{Corr}\big(X(t+\Delta t)-X(t), Y(t+\Delta t)-Y(t)\big)$$

or, similarly, the correlation between their log-returns, i.e.

$$\text{Corr}\big(\ln X(t+\Delta t)-\ln X(t), \ln Y(t+\Delta t)-\ln Y(t)\big).$$

The first thing to decide is what the differencing interval Δt should be. While it is common to talk about daily returns, we find the use of these is not always the best approach. First, if any data points are missing then a pair of daily returns is corrupted. Hence, a small amount of contamination greatly reduces the number of returns that are available. This would be also true of longer returns (such as weekly or monthly) but the effect is not so bad because some work-arounds are possible, e.g. if we are computing weekly returns and Tuesday 15th March's data is not available then we can always use Wednesday 16th instead, or the average of Monday 14th and Wednesday 16th. The error in doing this is likely to be fairly small compared with the true return. By similar reasoning, small amounts of contamination in each data point should produce smaller relative errors in the differences if the difference is taken over a longer time interval. Also by using a longer differencing interval, one substantially reduces the problem that prices quoted in different markets are not synchronous if they are on different time zones. Incidentally another advantage of using weekly (or some multiple of weekly) returns is that one does not have to worry about weekends: is Friday-Monday statistically different from Monday-Tuesday?

There is also another objection to the use of daily returns for our purposes. We are trying to evaluate the loss distribution at three-month horizon, not one day. It is not necessarily true that the one-day correlations are the same as the three-month ones (even if noise in the data could be discounted). An example of this is the autoregressive model of time series (we assume that the time step is 1), which is commonly used to model financial time series:

$$X(t)=aX(t-1)+e(t)$$

This models a time series that slowly de-correlates over time and reverts to $X=0$ (though some other mean level could be used instead), provided that $|a|<1$. The process $e(t)$ is uncorrelated, i.e. each $e(t)$ is independent of previous values. Let Y follow the same process,

$$Y(t)=bY(t-1)+f(t)$$

and let the increments be correlated according to

$$\text{Corr}(e(t),f(t))=\rho.$$

A little work shows that the k-period returns are correlated as follows:

$$\text{Corr}\big(X(t+k)-X(t), Y(t+k)-Y(t)\big)=\rho\,\frac{((ab)^k-1)/(ab-1)}{\sqrt{(a^{2k}-1)/(a^2-1)}\sqrt{(b^{2k}-1)/(b^2-1)}}$$

If $a=b$ then this is just ρ, regardless of k, but in general there is k-dependence. In fact the k-period correlation is smaller than ρ when $k>1$, and declines to $\rho\frac{\sqrt{(a^2-1)(b^2-1)}}{ab-1}<\rho$ as $k\to\infty$. This is an example of declining correlation.
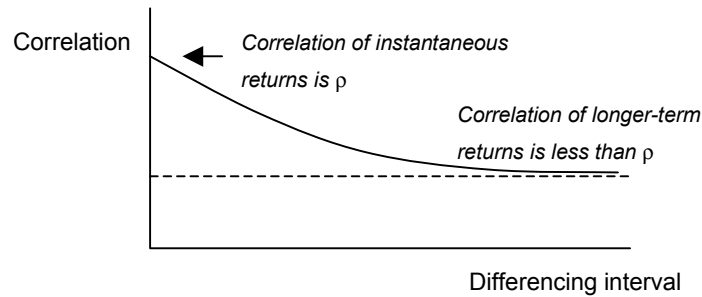
But if we return to the subject of contamination, the opposite effect can be observed. Let $X(t)$ and $Y(t)$ be Brownian motions. We do not observe them but instead see $\tilde{X}(t)$ and $\tilde{Y}(t)$ which are assumed to arise from the addition of uncorrelated observation noise:

$$X(t)=X(t-1)+e(t) \qquad Y(t)=Y(t-1)+f(t)$$
$$\tilde{X}(t)=X(t)+\tilde{e}(t) \qquad \tilde{Y}(t)=Y(t)+\tilde{f}(t)$$

Here $e(t)$, $f(t)$ play a significantly different roles from $\tilde{e}(t),\tilde{f}(t)$: the former represent the effect of new information on the market (and ignoring them makes the model pointless), whereas the latter are observation noise. We can only measure the correlations between the returns of $\tilde{X}(t)$ and $\tilde{Y}(t)$. It is easy to show that

$$\mathrm{Corr}\left(\tilde{X}(t+k)-\tilde{X}(t),\tilde{Y}(t+k)-\tilde{Y}(t)\right)=\frac{\rho k\sigma_e\sigma_f}{\sqrt{k\sigma_e^2+\sigma_{\tilde{e}}^2}\sqrt{k\sigma_f^2+\sigma_{\tilde{f}}^2}}<\rho$$

which starts less than ρ but approaches it as $k\to\infty$.
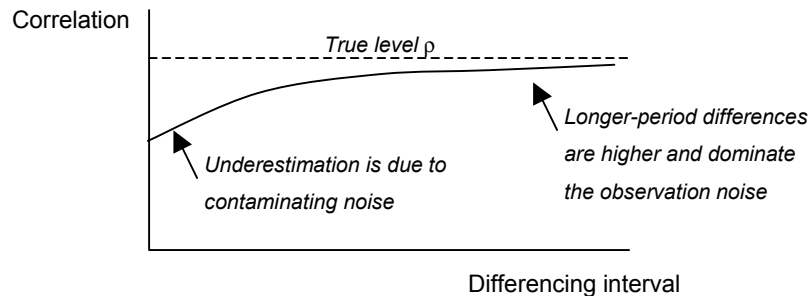
In consequence, if we want to know correlations at 3-month horizon we should use a differencing interval close to 3 months, rather than 1 day. On the other hand, this approach is not ideal either. If one has say 6 months or 1 year of data, then there is considerable uncertainty in estimating correlations between 3-month returns, simply because of the lack of data. For example, if 1 year's worth of data are used, there are

about 250 independent one-day returns. Now consider weekly returns. By differencing Jan1-Jan8, Jan2-Jan9, Jan3-Jan10 and so on one has again about 250 data points but they are no longer independent: the Jan1-Jan8 return is strongly correlated with the Jan2-Jan9 one, and so on. So the effective number of returns is considerably lower than 250: it is higher than 50, as the Jan1-Jan8 return is independent of the Jan8-Jan15 one, and so on, giving 50 weeks' worth, but the others (Jan2-Jan9, Jan9-Jan16 etc.) do not provide much more information. With three-month returns the problem is considerably worse.

Our approach is to use 1-monthly returns taken on weekly averaged data. That is, we average[68] the variable in question (equity, spread, or the like) each week, and then take the 4-week returns. This is schematically shown below. The idea is to have a differencing interval as close to 3 months as possible, while still having enough independent data points to work with.

Fig. 11.2. Each dot represents a weekly observation, and each arrow a 4-week return.
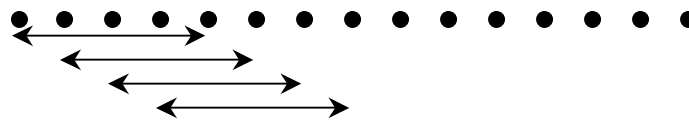
Issues in credit-equity correlation

There has been a certain amount of discussion about correlations in the credit and equity markets and we think it is worth discussing these now.

In the Merton framework the debt and equity are both modeled as contingent claims on some underlying firm value that follows a geometric Brownian motion. In the standard option-pricing terminology we therefore have

$$\text{Firm value: } dS_t = r_t S_t dt + \sigma S_t dW_t$$

$$\text{Equity: } dE_t = r_t E_t dt + \frac{\partial E}{\partial S} \sigma S_t dW_t$$

$$\text{Debt: } dD_t = r_t D_t dt + \frac{\partial D}{\partial S} \sigma S_t dW_t$$

This model unrealistically assumes amongst other things that the volatility is constant. Accordingly, all three variables move together and their instantaneous correlations are 100%. Now suppose we have two issuers, S_{1t} and S_{2t}, and let their returns be correlated through their Brownian motions by $dW_{1t} dW_{2t} = \rho\, dt$. Then the correlation between S_{1t} and S_{2t} is just ρ, and the same for E_{1t} and E_{2t}, and D_{1t} and D_{2t}.

---

[68] A robust method of averaging is to remove the highest and lowest and take the mean of the middle three. This is similar to the way LIBOR is calculated. Or one could take the median. These are more robust to outliers than a straight arithmetic mean.

This is just for the instantaneous returns, though. In reality when we consider weekly, monthly or three-monthly returns, the position becomes more difficult because we cannot assume that anything is Normally distributed any more. (The log of the firm value is, but that is not observable.)

It is here that the work of the first part of this chapter comes into play. The relationships between debt price, debt spread and equity are all nonlinear, and we cannot be certain what they are. (Of course CUSP™ goes a long way in assisting with this, because in simple models the Black-Scholes formulae can be used to transform from one to the other, but it is convenient if we can separate the correlation issues from the equity-credit model.) By using non-parametric estimators we can in principle end up with robust correlation estimates that do not depend on whether we use equity data, debt spread data or, inasmuch as it is observable, firm value data[69].

In practice, however, there are several flies in the ointment:

- There may be market regimes in which spread correlations are higher than equity correlations on account of supply and demand, liquidity, and the like. (Incidentally one would expect the use of weekly or monthly returns to be less susceptible to this problem than the use of daily returns.) So for debt-only portfolios one ought to use spread correlations, and for equity-only portfolios one ought to use equity correlations. However, one consistent set of correlations is necessary if the portfolio contains both equity and debt.

- Debt and equity have the same position (long) in the firm value, they take opposite positions in volatility (debt is short vol, equity is long). If the volatility is negatively correlated with the underlying (there is good evidence to support this in the equity market, to the extent that equity implied vol is negatively correlated with equity price), this causes the debt and equity correlations between two issuers to be different.

- Default-rate-implied asset correlations[70]. By these we mean the asset-asset correlations that are used in the risk management of loan portfolios where the emphasis is on default risk rather than spread risk. The higher the correlation, the higher the default rate volatility, so the question is what sort of figure corresponds to observed patterns of default in historical data. The Basel II Capital Accord uses 20%. This is lower than market-implied values, which is reasonable given that regulatory capital is supposed to be a lower bound for a portfolio's risk rather than an accurate estimate.

- Default-implied asset correlations. When pricing a credit basket or CDO one needs to know the joint default probabilities and although the Gaussian copula is the standard pricing model, the correlations that are really needed for the pricing are correlations between default *times*. However, there is a similarity between firm-value-based models and default-time models and it is not unreasonable (though not correct either) to use firm value correlations as a substitute for default-time correlations.

---

[69] The firm value is not directly observed in the market but it can be imputed from debt and equity data if one assumes the assumption that the arbitrage relationship Debt + Equity = Firm Value always holds. However, this becomes complicated when there are multiple claims on the firm.
[70] This does not mean the same as default event correlations, which are quite a different thing (see Chapter).

The table below summarises the "typical" correlations used in various situations.

Type of correlation	Instrument	Purpose	Imputed asset-asset correlation
Historical	Loans	Credit risk mgt, e.g., Basel II Capital Accord	Roughly 20%
Market-implied	CDS	Pricing CDOs/baskets	20-40% (higher in senior tranches)
Market-implied	Equities	Pricing baskets	40-50%
Historical	Equities	Risk management	40-50%

Source: Credit Suisse First Boston

Before moving on, we need to make some further comment on debt and equity moving together (or not). It is clearly unreasonable to say that debt and equity must always move together, and there is a small amount of published work on this[71]. In this chapter we are only testing the use of debt data to calibrate the factor correlation model. We shall consider the use of equity data, as an alternative, later on.

PR+2 Methodology – Part I

Bearing in mind all of the above discussions, we now explain the methodology used for estimating the PR+2 model. The basic idea is to use two layers of correlation: issuers to their individual sectors, and the sectors to each other. It is important to understand that a "sector" here does not mean a weighted average of issuers, as it does in index products; rather, it means a risk factor whose presence is inferred from the tendency of issuers in the same industrial environment to move up and down together. We use the notation

$$\Delta Z_i = c_i V_{S(i)} + \sqrt{1 - c_i^2} U_i$$

for linking the standardized log-return ΔZ_i of the ith issuer's firm (asset) value to its corresponding sector risk factor $V_{S(i)}$ ($S(i)$ is the sector in which the ith asset resides) and an issuer-specific part U_i. All of ΔZ, V, U are Normal with mean 0 and variance 1. Hence the asset-asset correlation for two issuers i,j in the same sector is $c_i c_j$. Between sectors we shall for the moment use the notation ρ. Then the correlation between issuers in different sectors is $c_i c_j \rho_{S(i)S(j)}$. From the above discussion on rough correlation numbers it appears that we want c_i to be around 0.6-0.7 (remember that this has to be squared to get the asset-asset correlation).

By asset, we mean here the firm (asset) value. The debt price, which is the observed data, is a transformation of it:

$$D_i(t) = f_i^D(Z_i(t))$$

where g^D is obtainable from the Black-Scholes formulae.

---

[71] "A robust test of Merton's structural model for credit risk", R. Jarrow et al., www.defaultrisk.com . This chapter uses a crude version of the "sign" approach that we are using. See also our own group's report "Debt or Equity: The Chicken or the Egg?" released at the end of 2003.

The next stage represents a departure from "normal" practice and is designed to get around two problems:

- Returns are not Normally distributed and may be contaminated;

- Equity and debt are nonlinear transformations of the underlying asset and it is convenient to have a method that works on both.

We use the ideas behind the non-parametric estimation techniques described earlier and, in the spirit of Kendall's tau, consider only the *signs* of the changes in $D(t)$, not the absolute values. In other words, an increase is represented as +1 and a decrease as -1. This has an important consequence: because we are assuming that debt and firm value always move in line, we must always have

$$\operatorname{sgn}(D_i(t) - D_i(t - \Delta t)) = \operatorname{sgn}(Z_i(t) - Z_i(t - \Delta t))$$

and we do not need to know about g^D and g^E.

We are now ready to embark on the parameter estimation, and concentrate first on one particular sector (so we estimate the c's). Suppose for the moment that only one return is given for each issuer in that sector, and we call its sign γ_{it}, where i is the issuer and t is the time point at which the return is taken. So if the equity or debt of the ith issuer increased in value, γ_{it}=+1, and if it decreased in value, γ_{it}=-1, and if it stayed the same to within some small tolerance level, γ_{it}=0 (see note[72]). Following the conditional independence approach we condition on V (the sectorial risk factor) and then integrate out. Conditionally on V, the distribution of the ith asset return Z_i is $N(c_i V, 1 - c_i^2)$. Then

the probability of Z_i being positive is $\Phi\left(\dfrac{c_i V}{\sqrt{1 - c_i^2}}\right)$ and the probability of it being negative

is $\Phi\left(\dfrac{-c_i V}{\sqrt{1 - c_i^2}}\right)$. So the probability of observing the data point γ_{it} is $\Phi\left(\dfrac{\gamma_{it} c_i V}{\sqrt{1 - c_i^2}}\right)$.

Multiplying up (as the issuers are conditionally independent) and integrating V out gives

$$p(\gamma_{1t}, \ldots, \gamma_{nt} \mid c_1, \ldots, c_n) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{\gamma_{it} c_i V}{\sqrt{1 - c_i^2}}\right) \phi(v)\, dv$$

which is the probability of observing the data set given the (as yet unknown) parameters c. Now we have to deal with the fact that we have more than one time point for each issuer. If the returns were taken over non-overlapping time intervals (e.g. Jan1-Jan8, Jan8-Jan15, etc) they would be independent and we would have

$$p(\gamma_{1\ldots n, t(1)\ldots t(N)} \mid c_1, \ldots, c_n) = \prod_{t=t(1)}^{t(N)} \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{\gamma_{it} c_i V}{\sqrt{1 - c_i^2}}\right) e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} \right]$$

---

[72] If we use spreads, then obviously the spread changes in opposite direction to the value, so all the signs are reversed. This does not affect anything.

But (see Fig. 11.2) we are using overlapping intervals, so independence no longer holds. To take this into account we raise to the power of an "overlap-factor" which for 4-week returns taken at 1-week steps is $\kappa = \frac{1}{4}$. This gives:

$$p(\gamma_{1\ldots n,t(1)\ldots t(N)} \mid c_1,\ldots,c_n) \approx \left\{ \prod_{t=t(1)}^{t(N)} \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{\gamma_{it} c_i v}{\sqrt{1-c_i^2}} \right) e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} \right] \right\}^{\kappa}$$

which can be readily computed (the integration is done by Gaussian quadrature). We can now maximize this expression with respect to the c's to get the maximum likelihood estimate, i.e. the set of c's from which the data are most likely to have come. (In fact if we choose to do just this then we do not need to bother with κ.)

However this approach is not as good as it could be and tends to produce estimates that wander about too much over time. To counteract this we introduce the final ingredient into the estimation process: the concept of a *prior*.

The approach described above is known as maximum likelihood (ML), or classical statistical inference. A philosophical objection to it is that it maximizes the probability of observing the data given the parameters; but what is more natural is surely to consider instead the probability distribution of the parameters given the data. This idea requires swapping the conditioning around and is known as *Bayesian inference*.

Suppose that we wish to find the distribution of some parameter θ given data x (this is the standard notation used in the literature). Bayesian inference asserts that unknown parameters have distributions. First we have the prior distribution which is the distribution put on θ before any data are observed. This is denoted $p(\theta)$ and its selection is largely a matter of subjective judgment. Suppose now that data are observed and we know the likelihood function $p(x|\theta)$, i.e. the probability of observing the data given the parameters. Then the posterior distribution of θ given x is obtained by Bayes' theorem:

$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)}$$

What is commonly done is to maximize the posterior probability $p(\theta|x)$. This is known as the MAP (maximum a posteriori) estimate. To summarise:

- ML : maximize $p(x|\theta)$ w.r.t. θ;
- MAP : maximize $p(\theta|x)$ w.r.t. θ.

In finding the MAP estimate we can ignore the $p(x)$ in the denominator (as it doesn't depend on θ).

To relate this to our problem, the comparison of the "standard" notation with the notation we have been using above is:

Prior	$p(\theta)$	$p(c)$
Likelihood fn	$p(x|\theta)$	$p(\gamma_{1\ldots n,t(1)\ldots t(N)} \mid c_1,\ldots,c_n)$
Posterior	$p(\theta|x)$	$p(c_1,\ldots,c_n \mid \gamma_{1\ldots n,t(1)\ldots t(N)})$

Turning now to the formulation of the prior we wish to choose a distribution that represents our view that the correlation parameter *c* is around 0.6 with some degree of uncertainty. We should choose a distribution that vanishes for *c*>1 as such a correlation is impossible. A reasonable idea is to use the expression[73]

$$p(c) \propto \exp\left(-\frac{1}{2}\left(\frac{\text{arctanh}(c) - \text{arctanh}(\mu)}{\sigma}\right)^2\right)$$

(the normalization coefficient need not be computed). The graph of this function is shown in Fig. 11.3 below (μ=0.6 and σ=0.5). It is apparent that this is the right sort of shape. In the absence of any data at all(!!) a MAP estimate using this prior would give a correlation estimate of 0.6.

Fig. 11.3. Example prior (μ=0.6, σ=0.5)
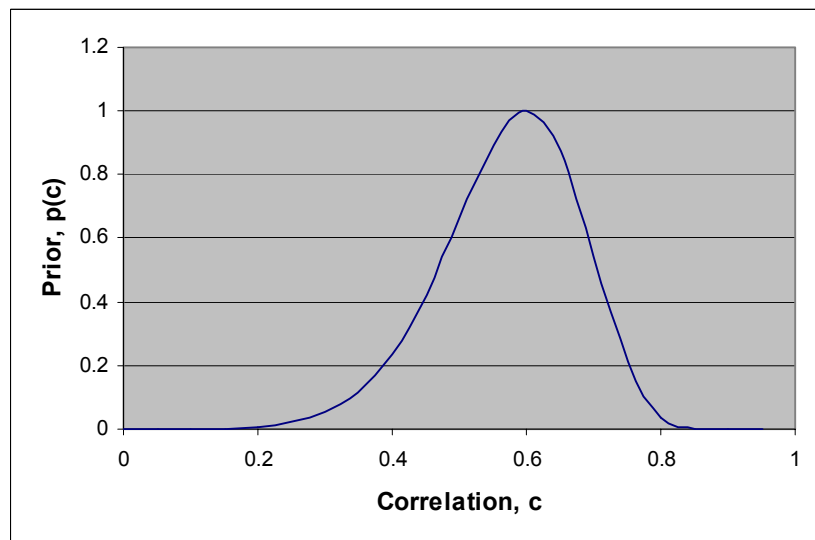
Source: CSFB

The lower the value of σ, the narrower the distribution and the greater the penalty placed on correlations differing from the "modal" value of 0.6.

The maximization is done using standard numerical procedures, e.g. the Excel Solver or other commercial packages e.g. NAG. This allows us to calibrate our database of issuers sector by sector.

The next question is how to solve for the sector-to-sector correlations. There are various ways of trying to do this, which include:

- Aggregating spread data within each sector to give an "industry sector spread", and then correlating those together. This is the traditional approach.

- Forming a likelihood function for the entire set of issuer data and using this to estimate the sector-sector correlations.

We shall discuss these later. Meanwhile we now show how well our new issuer-sector method works in practice.

---

[73] This *loosely* says that the inverse hyperbolic tangent of the correlation coefficient is assumed Normally distributed. The shape of the prior is the important thing, not the mathematical precision of its derivation.

Back-testing of issuer-sector correlation

We have backtested the methodology on credit spread data from a wide range of high-grade issuers over the last year and now present an overview of the backtesting procedure and results.

For each month we took the historical spread data available on that date (i.e. the previous 6 or 12 months) and performed three analyzes:

1. Calculation of issuer-sector correlation by the existing PR+1.5 method (use daily returns over 6 months and aggregate issuers within a sector; then correlate the issuers with that aggregate to produce an estimate of the correlation[74])—in the graphs this is called *Old method*;

2. As Method 1 but using monthly returns over a 12-month period, rather than daily returns—in the graphs this is called *Monthly returns;*

3. Using the PR+2 method as described in this chapter (Bayesian with signed monthly returns over a 12-month period)—in the graphs this is called *New method.*

Analysis 2 is in a sense mid-way between the current and new methods: like the new method it uses monthly returns, but like the current method it uses the raw spread returns rather than the signed returns coupled with a prior. Incidentally it is possible to run portfolios in PR+1.5 using the PR+2 correlation estimator and vice versa, because the analytical engines use the same correlation model.

In general, we find that the new correlation method produces considerably more stable estimates, which vary only a small amount from month to month. It also appears that the use of monthly returns gives more stable results than daily ones, and this is consistent with our earlier suggestions.

To give an illustration, we show typical results for several well-known issuers.

---

[74] In the old method a lower threshold of 20% was applied too. This is not necessary in the new method.

ABS US (Retail)
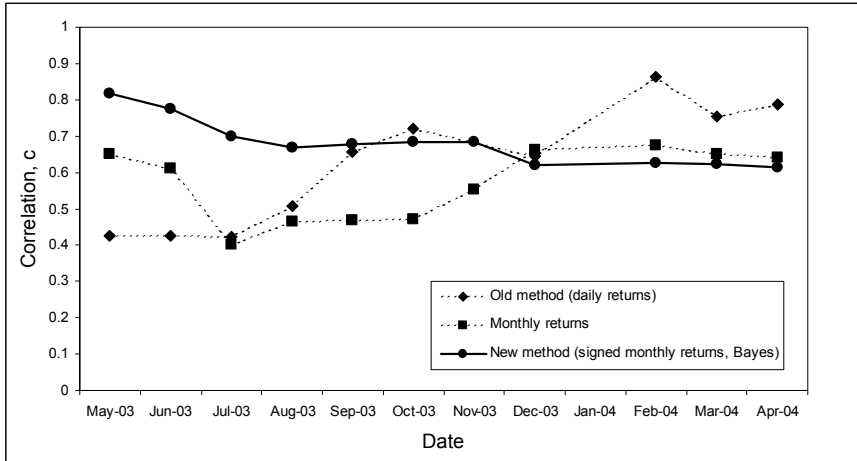
AEP US (Electric Utilities)
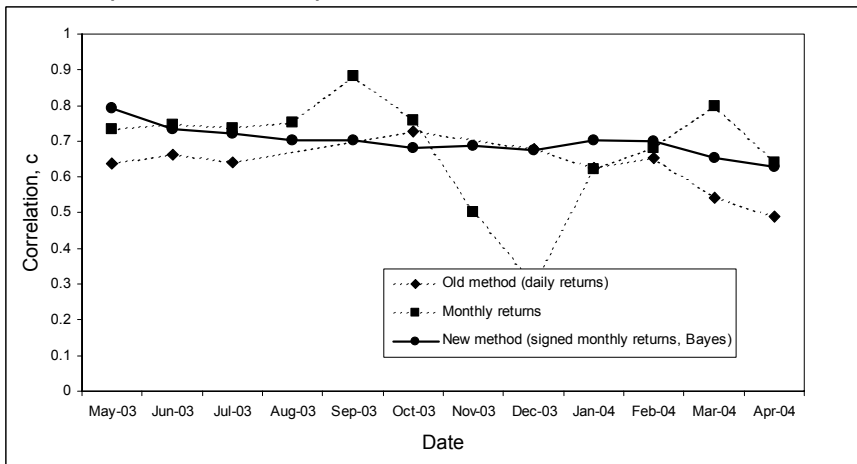
WYE US (Healthcare)

Source: CSFB

CREDIT SUISSE | FIRST
SUISSE | BOSTON
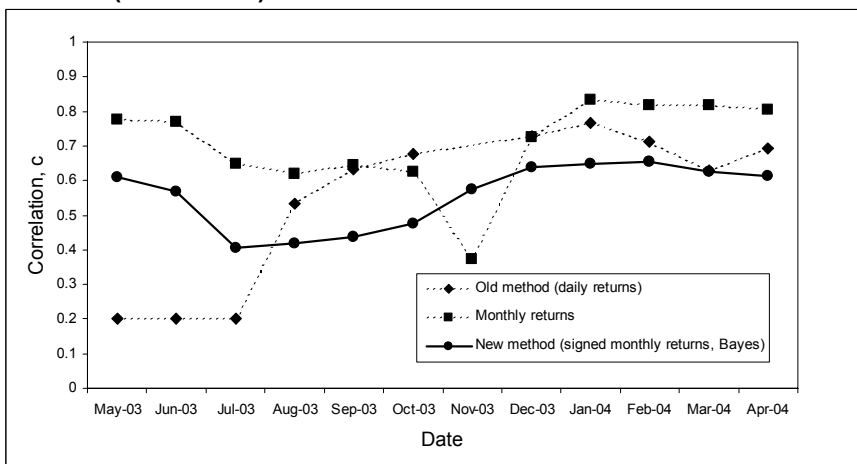
ALL US (Insurance)
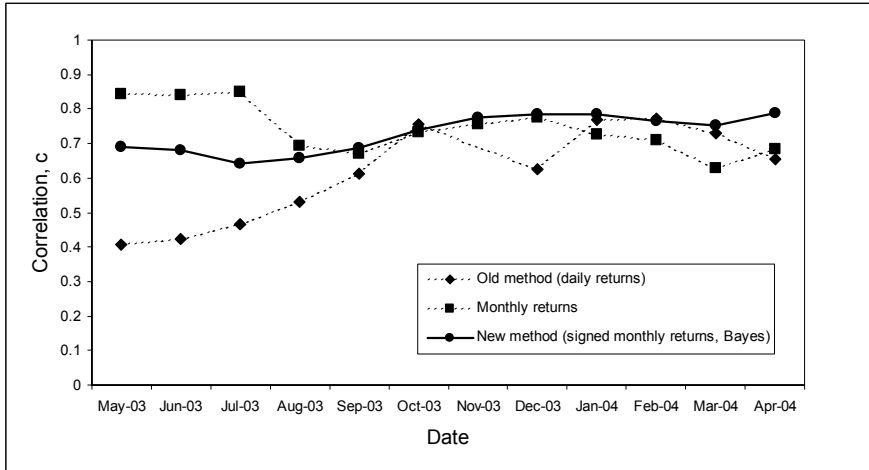
IBM US (Technology)
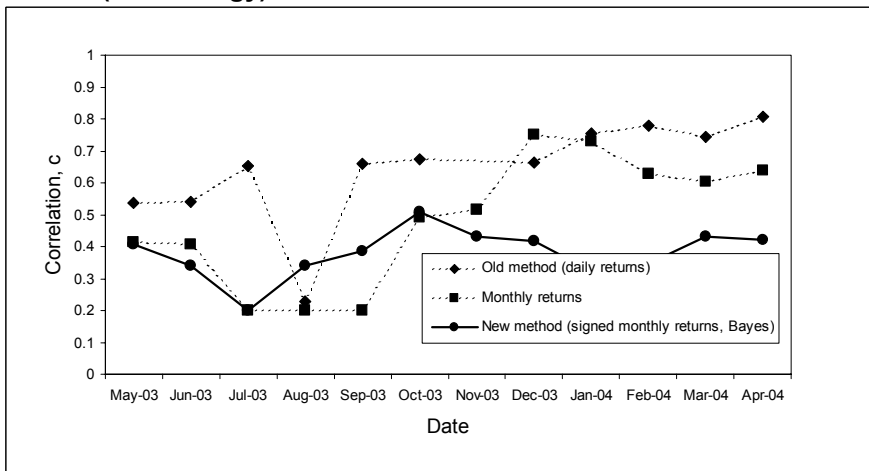
APA US (Energy)
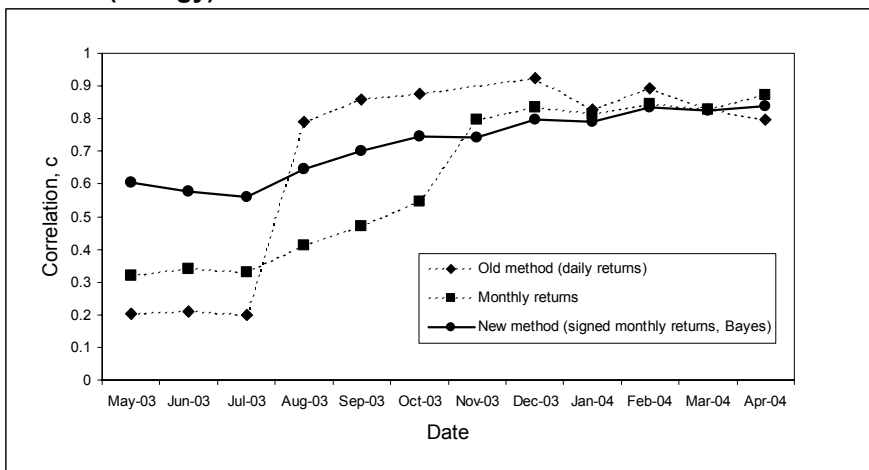
Source: CSFB

NWS US (Media)
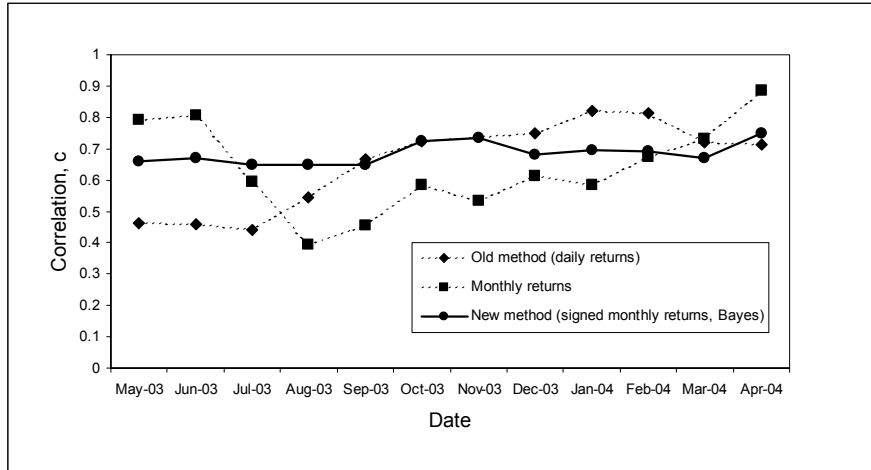
MER US (Finance)
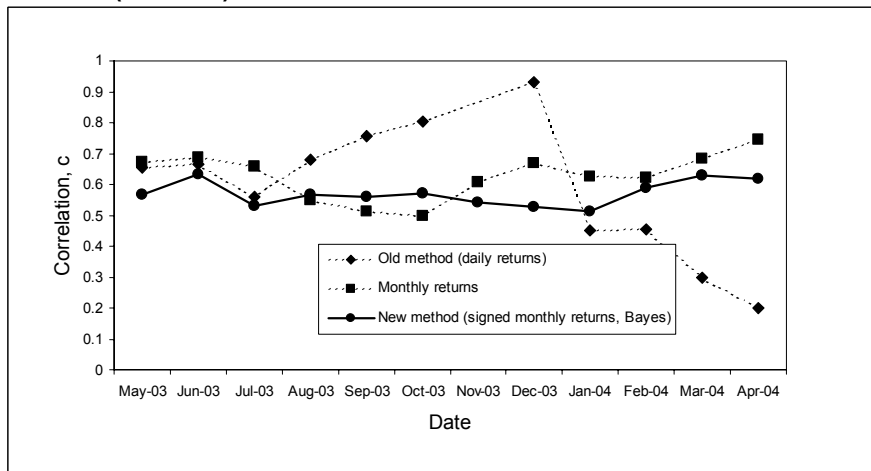
FTE FP (Telecoms)
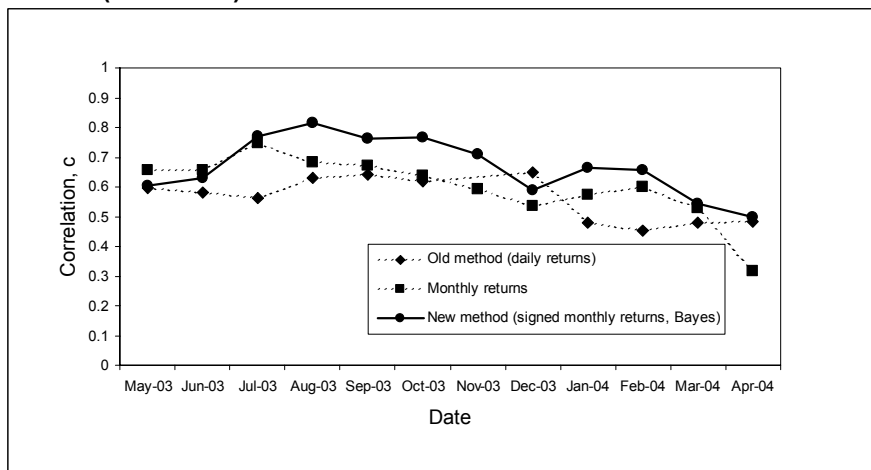
Source: CSFB

DTE GR (Telecoms)
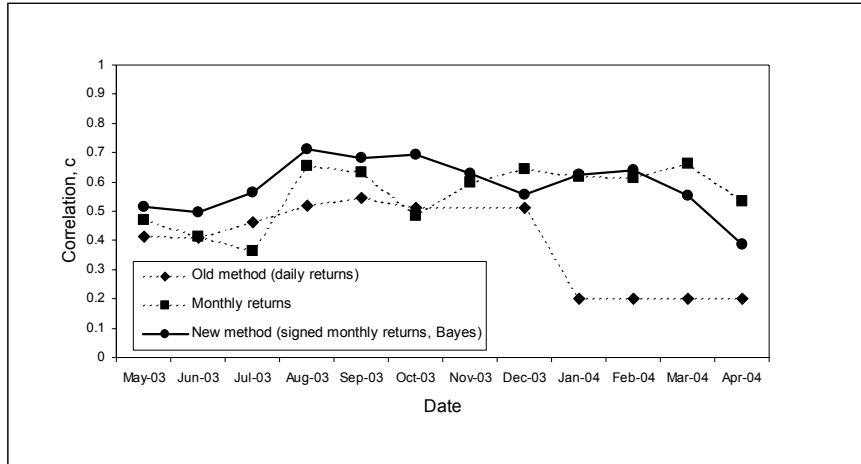
VNO US (Reits)
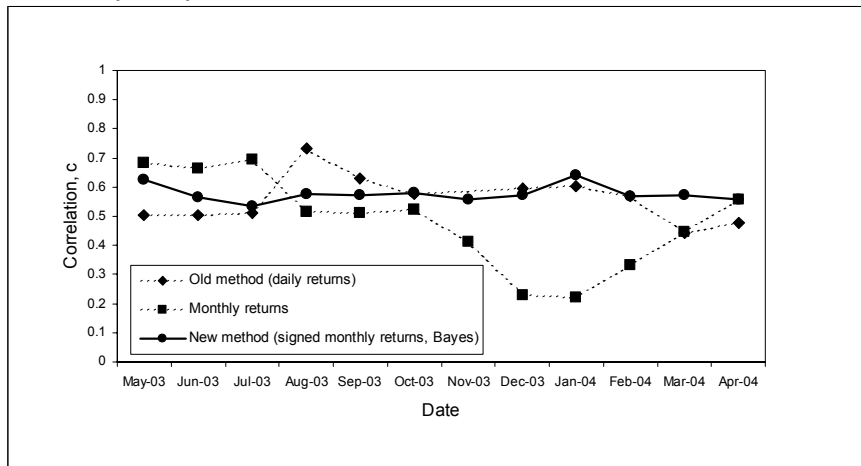
SRE US (Gas pipelines)
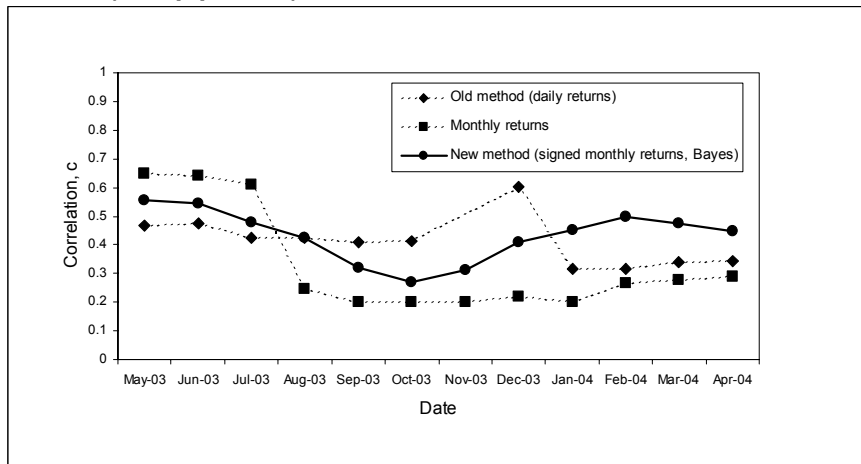
Source: CSFB

PR+2 Methodology – Part II

The starting-point for the estimation of issuer-sector correlations was the joint likelihood function

$$p(\gamma_{1t}, \ldots, \gamma_{nt} \mid c_1, \ldots, c_n) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{\gamma_{it} c_i v}{\sqrt{1 - c_i^2}} \right) \phi(v) dv$$

where γ_{it} is the sign of the return of the i'th issuer at time point t, and $c_1, \ldots, c_n$ are the issuer-sector correlation parameters (correlation between firm value return of the issuer with the relevant sectorial risk factor). When combining data from many time points one can just multiply them, but if the time intervals are overlapping an adjustment needs to be made as the returns are no longer independent. This gives

$$p(\gamma_{1\ldots n, t(1) \ldots t(N)} \mid c_1, \ldots, c_n) \approx \left\{ \prod_{t=t(1)}^{t(N)} \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{\gamma_{it} c_i v}{\sqrt{1 - c_i^2}} \right) \phi(v) dv \right] \right\}^{\kappa}$$

where for 4-week returns taken at 1-week steps we have $\kappa = \frac{1}{4}$. The form of the prior is

$$p(c_1, \ldots, c_n) = p(c_1) p(c_2) \cdots p(c_n)$$

with

$$p(c) \propto \exp\left(-\frac{1}{2} \left(\frac{\operatorname{arctanh}(c) - \operatorname{arctanh}(\mu)}{\sigma} \right)^2 \right)$$

which expresses our prior opinion that the correlation parameters are independently distributed and that each is "not far from μ", with σ controlling how tightly bound to μ they are. We then maximize the a posteriori probability with respect to the c's, sector by sector.

Having obtained the issuer-sector parameters, we now turn to the sector-sector modeling. The first point that we want to emphasize is that the PR+2 calculator allows an arbitrary correlation matrix for sectors to be specified. Therefore one possibility is to aggregate spread data, sector by sector, and correlate the time series of "sector-average spread". However, this is not consistent with our issuer-sector approach, where the sector is simply a risk factor. One could try and estimate a time series for each factor and then correlate the factors together. However, we are investigating an approach that is completely consistent with the methodology we used for the issuer-sector analysis: we try and estimate the correlations directly from the signed issuer returns. To do this we shall assume a simplified correlation structure for the sectors. This substantially reduces the number of parameters to be estimated, as we shall now show.

The picture that we have in mind is this:

Fig. 11.4. "Graphical" model of correlation.
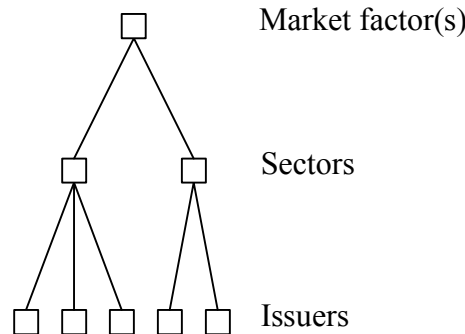
Lines indicate direct dependence. Each issuer relates to (only) one sector and the sector factors are themselves correlated. As before the issuer return is written

$$\Delta Z_i = c_i V_{S(i)}^{SCTR} + \sqrt{1 - c_i^2} U_i^{ISSR} .$$

If the sectors are correlated through a single risk factor V^{MKT}, each having a residual component U_S^{SCTR}, then the sector risk factors are decomposed as

$$V_S^{SCTR} = b_S V^{MKT} + \sqrt{1 - b_S^2} U_S^{SCTR}$$

so that the correlation between two different sector factors S_1, S_2 is $\rho_{S_1 S_2} = b_{S_1} b_{S_2}$. Thus the correlation between issuers in the same sector is $c_{i_1} c_{i_2}$, and in different sectors is $c_{i_1} c_{i_2} \rho_{S_1 S_2} = c_{i_1} c_{i_2} b_{S_1} b_{S_2}$. Note that this representation is parsimonious: for example if there are 20 sectors then the number of correlation parameters is only 20 for the one-factor approach, as compared with 20x19/2=380 for a full correlation matrix. It is also very simple to ensure positive-definiteness: we just need to ensure that all the *b*'s lie between –1 and 1. In fact, our preconception about them is a bit more restrictive than that: loosely, a range of 0.5-0.8 seems reasonable, so that sector-sector correlations are about 30%-60%. This means that we think that issuers in different sectors are about half as correlated as issuers in the same sector, on average[75].

If we adopt this one-factor correlation model of sectors, we can write the issuer return in the following form:

$$\Delta Z_i = \underbrace{c_i b_{S(i)} V^{MKT}}_{\text{market}} + \underbrace{c_i \sqrt{1 - b_{S(i)}^2} U_{S(i)}^{SCTR}}_{\text{sector specific}} + \underbrace{\sqrt{1 - c_i^2} U_i^{ISSR}}_{\text{issuer specific}} .$$

---

[75] This is partly based on experience with KMV PortfolioManager™ and with CDO pricing.

To analyze the data at sector level we write down the likelihood function, i.e. probability of observing the data (=signed returns) given the parameters. To do this we proceed as before. First we condition on the market factor V^{MKT}, and call its value w. Now for each sector, let the value of the sector-specific factor be u_S. Then within one particular sector, the issuer return has distribution

$$N\left(c_i\left(b_S w + \sqrt{1-b_S^2}\, u_S \right), 1-c_i^2 \right)$$

and the probability of its return being positive or negative is

$$\Phi\left(\frac{\pm c_i(b_S w + \sqrt{1-b_S^2}\, u_S)}{\sqrt{1-c_i^2}} \right).$$

By conditional independence the probability of observing a given set of signs for all the issuers in one particular sector is

$$\int_{-\infty}^{\infty} \prod_{i=1}^{n_S} \Phi\left(\frac{\gamma_{i,S,t} c_i\left(b_S w + \sqrt{1-b_S^2}\, u_S \right)}{\sqrt{1-c_i^2}} \right) \phi(u_S)\, du_S$$

where $\gamma_{i,S,t}$ denotes the signed return of the ith issuer in sector S (the issuers in that sector are labelled i=1,...n_S). Now the sectors are independent conditionally on the market factor, so we can now combine them. The probability of observing the set of signs across all issuers and sectors (but still at one time point) is

$$\int_{-\infty}^{\infty} \prod_S \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n_S} \Phi\left(\frac{\gamma_{i,S,t} c_i\left(b_S w + \sqrt{1-b_S^2}\, u_S \right)}{\sqrt{1-c_i^2}} \right) e^{-v^2/2}\, \frac{du_S}{\sqrt{2\pi}} \right] dw.$$

We now combine results from all time points as before and insert the "overlap-factor" to get

$$p(\text{data} \mid b_1,\ldots,b_m) \approx$$

$$\left\{ \prod_{t=t(1)}^{t(N)} \left[\int_{-\infty}^{\infty} \prod_S \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n_S} \Phi\left(\frac{\gamma_{i,S,t} c_i\left(b_S w + \sqrt{1-b_S^2}\, u_S \right)}{\sqrt{1-c_i^2}} \right) \phi(u_S)\, du_S \right] \phi(w)\, dw \right] \right\}^{\kappa}$$

It is important that we keep the estimates of the c's that we derived before, and as a result the above expression is only to be viewed as a function of the sector correlation parameters (the b's). Although the computation looks awkward it is not particularly difficult or time-consuming: the sector integrals (inner square bracket) are computed one by one for each sector and time point, and then the outer integration is done[76].

---

[76] Consequently the computation is essentially one integral inside another, not a multidimensional integral: the workload does not increase exponentially with the number of sectors.

We now wish to find the *b*'s (there are *m* of them say, where *m* is the number of sectors) and, importantly, we have already found the *c*'s. We use the same functional form of prior as before:

$$p(b) \propto \exp\left(-\frac{1}{2}\left(\frac{\operatorname{arctanh}(b) - \operatorname{arctanh}(\mu)}{\sigma}\right)^2\right)$$

$$p(b_1, \ldots, b_m) = p(b_1)p(b_2)\cdots p(b_m)$$

though of course we can allow the mean and dispersion parameters to be different (we use 0.7 for the mean for each sector as we think the sectors are quite strongly correlated). It is then just a matter of multiplying the likelihood function by the prior and minimizing the resulting function, as before.

Back-testing

We have seen how the issuer-sector correlation estimates are reasonably stable from month to month (results were shown for twelve different issuers taken from a variety of sectors and of differing size). Here we show the results for the sector-sector correlations. Remember that what is being shown here is the sector-market correlation, i.e. the correlation of the sector risk-factor to the presumed "market" risk-factor that is responsible for cross-sector correlation. To obtain the sector-sector correlation for two different sectors one simply multiplies their sector-market correlation parameters.

Our universe of issuers is partitioned into nineteen sectors, though a couple of them contain very few issuers. Here we show some of the results from the new estimation method. Currently we estimate a full covariance matrix for the sectors, so the methods are not directly comparable. What we can do, however, is to estimating a "market time series" (combining all issuers) and "sector time series" (combining all issuers within a particular sector) and correlating the two. This allows us to give a sector-market correlation that is consistent with the current methodology. We refer to this as the "old method" on the graphs, and the Bayesian approach described in this chapter as the "new method".

As can be seen from the graphs, the new method (shown as the solid line) exhibits rather less monthly fluctuation than the current one (shown as the dotted line), while giving fairly similar "average" values. This cannot simply be ascribed to the Bayesian prior forcing all the correlation parameters towards 0.7—though this would occur if the dispersion parameter σ were set to a very low value. For, looking at Insurance, one sees that the correlations are at a pretty static 90%.

Conclusions

In the first part of this chapter, using some ideas from the theory of robust statistics we developed and back-tested a new method of estimating correlations from historical spread data. The method consists in separately estimating issuer-sector and sector-sector correlations, where here a "sector" means a risk factor that causes issuers in the same industry sector to have a significant common component of variation. It appears that this method produces more stable results than the use of "raw" daily returns, which we have suggested is inadequate for a variety of theoretical and practical reasons.

In the second part we have completed the description of our new method for modeling correlations in PortfolioRisk+. The emphasis has been on robust estimation, which is why we have used non-parametric methods and also used priors.

We have hinted that, at least in principle, equity returns could also be used in the calibration. It would therefore be interesting to discover whether the use of equity returns would produce substantially different results.

Banks
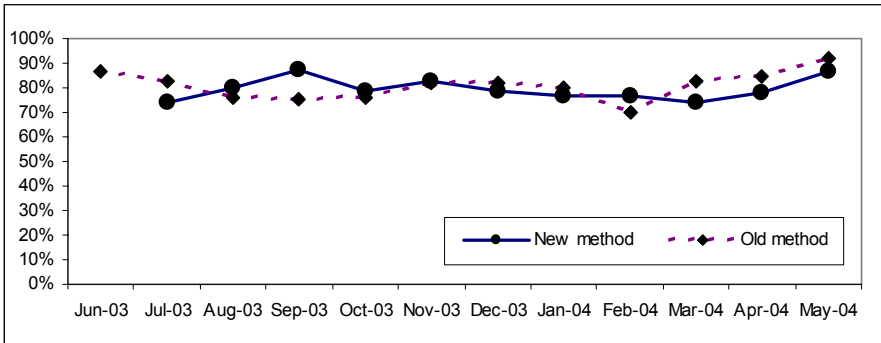
Financials
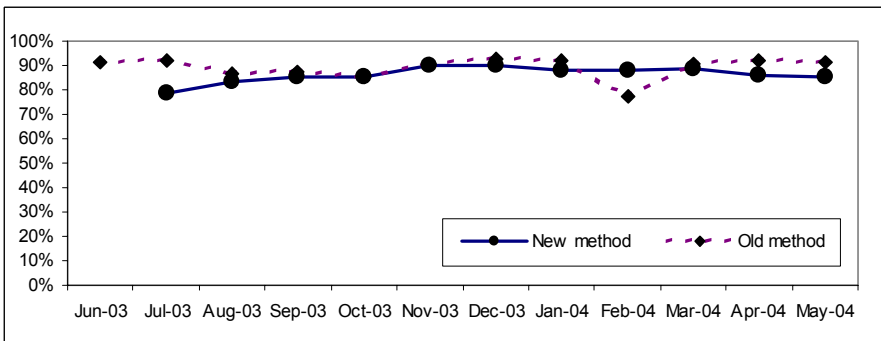
Insurance
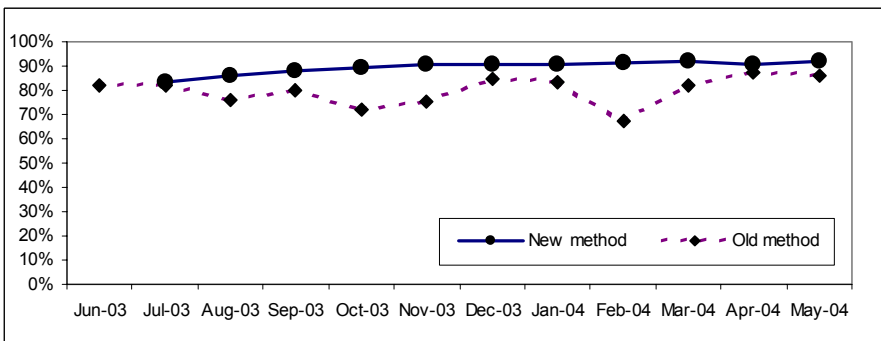
Media
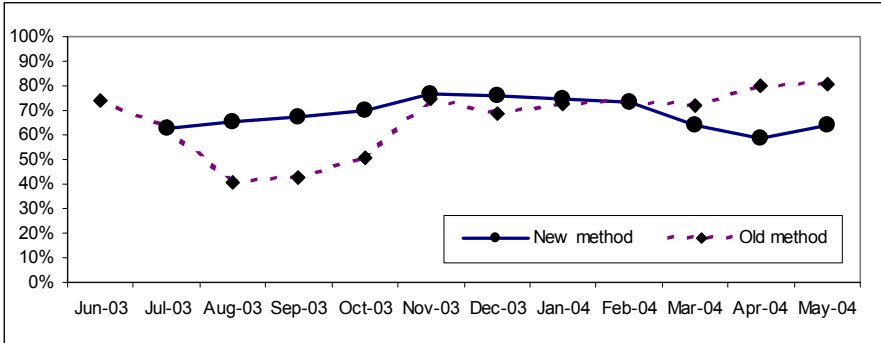
Technology
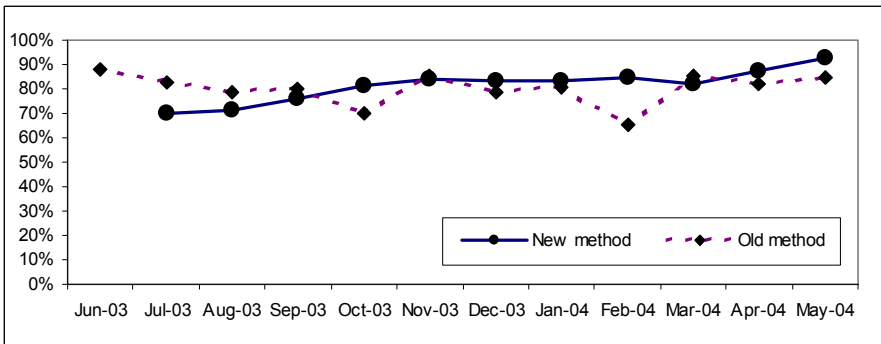
Telecoms
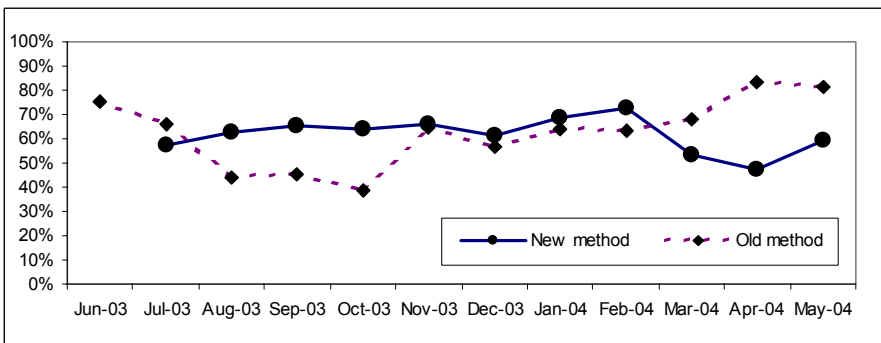
Appendix – Example of Bayes' theorem

Suppose that we suspect that a coin is biased. We believe that the probability θ of its showing heads is 0.4, though we are not certain that that is so. Suppose that we choose a Beta distribution for θ, with mean 0.4 and standard deviation 0.15. This means that (roughly)

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}, \text{ with a=3.87 and b=5.8}$$

(a, b having been chosen to match the mean and variance: the mean is a/(a+b) and the variance is ab/(a+b)^2(a+b+1)).

Now we observe that in 50 coin tosses, 24 show up heads. By the Binomial Theorem the probability of observing this is

$$p(x \mid \theta) = \frac{50!}{24!26!} \theta^{24}(1-\theta)^{26}$$

So the posterior probability distribution of θ is

$$p(\theta \mid x) \propto \theta^{24+2.87}(1-\theta)^{26+4.8} .$$

The use of a Beta distribution as a prior was a neat choice because the posterior distribution is another Beta distribution. Its mean is (24+3.87)/(24+3.87+26+4.8)=0.467. The MAP estimate is the mode of the distribution, which is (a-1)/(a+b)=0.45. Whichever of these we choose to take as our view of biasedness, our conclusion is that the coin is less biased than we previously thought it to be.
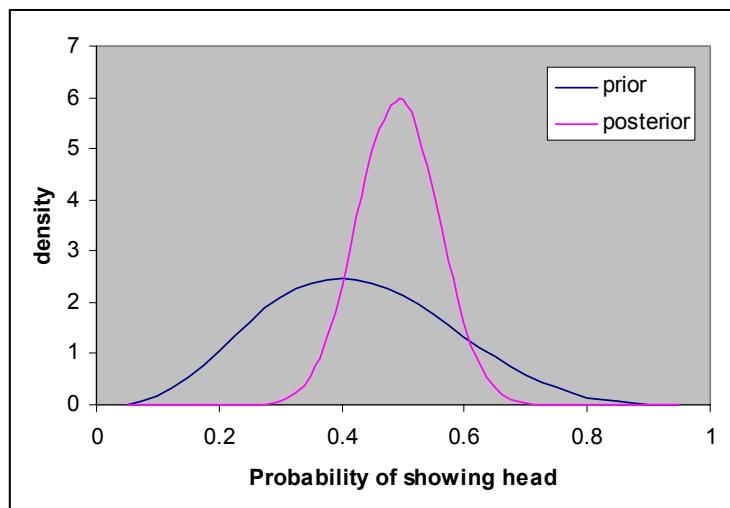
Source: CSFB

THIS PAGE IS INTENTIONALLY LEFT BLANK

12

Volatility, correlations, and the CAPM

We are now going to sew together some of the fundamental ideas that we have introduced in the previous chapters. These are the mean-variance theory of risk, risk contributions, and correlation. We show how all these are used in portfolio management and how they relate to the way the market prices risk.

Portfolio volatility

All investors trade risk in some form and to some extent. For example, an investor in credit-risky bonds has the risk either that the bond will increase or decrease in value (mark-to-market viewpoint) or that the bond defaults (buy-and-hold viewpoint). The Merton model argues that the risk of a bond (whether default or mark-to-market risk) arises from the volatility of the firm's assets. In return for taking this risk, the bondholder is rewarded with a coupon that not only covers his expected loss but also rewards his risk-taking. This latter reward is known as the risk premium: no risk, no premium.

Once the investor has assembled a portfolio, he can look at everything at a portfolio level and take the view that he is being rewarded for taking an aggregated risk. There then arises the question of whether enough reward is being given for the assumed risk. As we have previously discussed, there are several ways of measuring risk[77], but the simplest one to start with is the mean-variance framework.

Here the risk of a portfolio is measured by its standard deviation at some time horizon; if the horizon is short then the term *volatility* is effectively synonymous with standard deviation. It is convenient to keep the same word as we can then enquire how a firm's asset volatility feeds through to bond price or spread volatility, and thence to portfolio volatility. Portfolio volatility depends on two things:

- individual asset volatilities, and
- correlation between assets.

Increasing either of these causes the portfolio volatility to increase. For a picture, see Fig. 12.1 (next page).

Estimating market correlations is therefore of fundamental importance in assessing in a portfolio context what is worth buying and what is not. This is why we have spent a lot of effort on building robust correlation estimators.

---

[77] "Risk measures: How long is a risky piece of string?", Chapter 9.

**CREDIT
SUISSE** | FIRST
BOSTON

Fig. 12.1. Effect of asset volatility and correlation on portfolio volatility (NB the graphs are freehand sketches only)

(a) Moderate asset volatility, moderate correlation ⇒ moderate portfolio volatility.
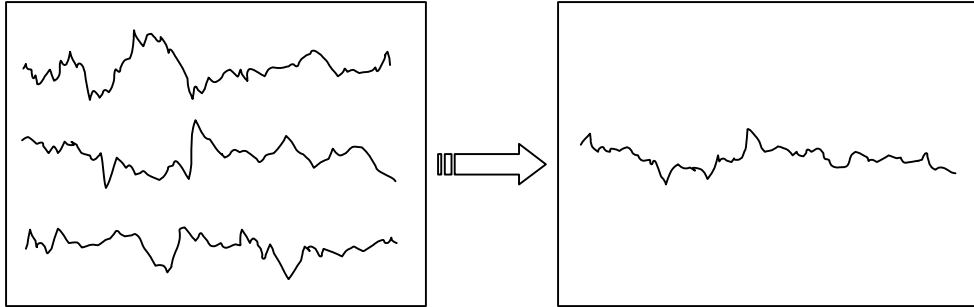
(b) Higher asset volatility ⇒ higher portfolio volatility.
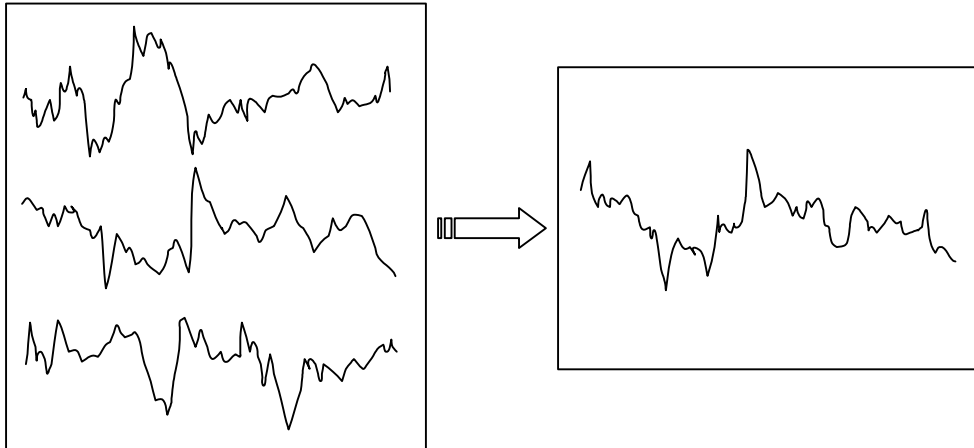
(c) Higher correlation ⇒ higher portfolio volatility.
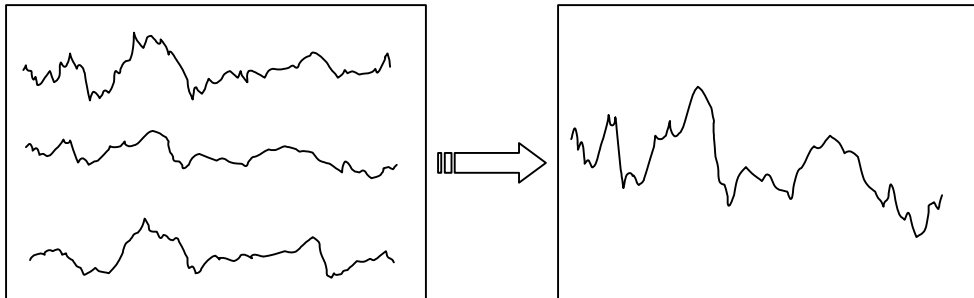
Risk contributions

We argued in Chapter 2 that the best approach to modeling correlation is *not* to write down a huge matrix of pairwise correlations, but instead to seek factors that govern the dependence. However, if all we want to do is construct a formula for portfolio variance then a covariance matrix is the quickest route.

We consider a portfolio $Y = \sum_i a_i X_i$ in which X_i is the return of the i'th asset and a_i is the i'th asset allocation. The portfolio variance is therefore

$$\sigma_Y{}^2 = \sum_{i,j} a_i a_j \sigma_i \sigma_j \rho_{ij}$$

where σ_i is the i'th standard deviation and ρ_{ij} is the pairwise linear correlation coefficient between asset i and asset j. This gives the portfolio standard deviation as

$$\sigma_Y = \left(\sum_{i,j} a_i a_j \sigma_i \sigma_j \rho_{ij} \right)^{1/2}.$$

Following our discussion of risk measures, we wish to find the sensitivity, or delta, of the portfolio risk (σ_Y) to asset allocation. This is

$$\frac{\partial \sigma_Y}{\partial a_i} = \frac{1}{\sigma_Y} \sum_j a_j \sigma_i \sigma_j \rho_{ij} = \frac{1}{\sigma_Y} \text{Cov}[X_i, Y]$$

so that the portfolio risk is most sensitive to assets that have a high covariance with the market portfolio. A high covariance results either from high volatility or high correlation to the rest of the portfolio.

Capital Asset Pricing Model (CAPM[78])

How does the market price risk? Let Y be the "market portfolio" and declare that the market prices in an optimal way by equilibrium: low-returning risky assets are sold and high-returning less-risky assets are bought until equilibrium is achieved.

We saw in our previous discussions on risk measures[79] that a portfolio is optimal in the sense of Sharpe ratio (return ÷ risk maximized) when the excess returns are proportional to the risk contributions. That is,

$$r_i = k \frac{\partial \sigma_Y}{\partial a_i} \quad \text{for each } i$$

with r_i denoting the expected excess return of the i'th asset. Putting everything together we arrive at the famous Capital Asset Pricing Model:

$$r_i = k \frac{\partial \sigma_Y}{\partial a_i} = \frac{k}{\sigma_Y} \text{Cov}[X_i, Y] = "\beta_i"$$

---

[78] Introduced by Markowitz (1952) and Sharpe (1964).
[79] "Risk measures: How long is a risky piece of string?", Chapter 9.

("beta" being the universally used CAPM terminology). So, loosely, one is rewarded for volatility that is correlated with the market. Moreover the reward is linear in the risk (Fig. 12.2, left picture).

This argument does depend on market efficiency and if taken too far it has some unsavoury implications for portfolio managers: you may as well hold the market portfolio, and you get no reward for diversifying your portfolio (because investors can do that themselves). But, if we argue that inefficiencies exist in the market, owing to, for example, players holding different amounts of market information, players having trading constraints, or players having different risk measures, then the picture on the right is a truer representation of reality and now portfolio selection becomes a valuable exercise:

Fig. 12.2. Risk vs. return in the CAPM world and in the real world

CAPM (Efficient market)   Real world (inefficient)
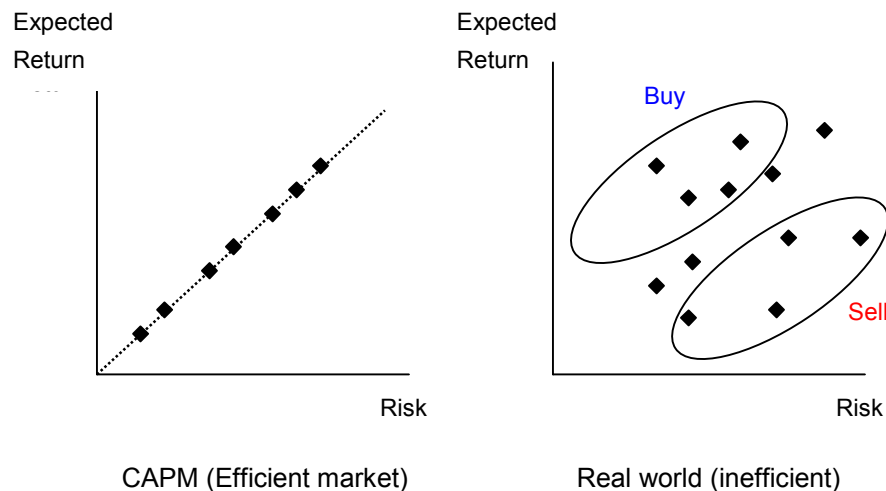
At this point we have shown how to make trade recommendations based on the assumption that we start with a portfolio similar to the "market portfolio" (in practice, a suitable index or benchmark). For example, if ABC Inc has a high return to risk, then we can recommend buying it. We can do this without going into greater depth than looking at first-order risk contributions to the market portfolio, which we assume to be similar to those in our portfolio.

However, as we have said before[80], this simple analysis does not tell us how much to buy: as we increase the weight in ABC its risk contribution increases in a nonlinear way (as we shall discuss next, this arises from portfolio risk being a convex function of asset allocation) and if we buy too much then we will be over-exposed to it. Of course, in the case where our portfolio is very different from the market portfolio, the recommendation may even be the wrong way round: for example it would not be correct to buy ABC Inc if our portfolio happened to be greatly overweight in that asset already. Hence to do a proper portfolio optimization we need to know the precise portfolio composition and to consider second derivatives of risk w.r.t. asset allocation. We briefly consider this latter point next.

---

[80] "Portfolio optimization: The importance of convexity", Chapter 10.

Convexity

One differentiation more shows that

$$\frac{\partial^2 \sigma_Y}{\partial a_i \partial a_j} = \frac{1}{\sigma_Y}\left(\text{Cov}[X_i, X_j] - \frac{\text{Cov}[X_i, Y]\,\text{Cov}[X_j, Y]}{\sigma_Y{}^2} \right)$$

To prove the all-important convexity, which is necessary for a unique optimal portfolio, we need to show that the matrix of second partial derivatives is positive definite. This means that, if we write $H_{ij} = \dfrac{\partial^2 \sigma_Y}{\partial a_i \partial a_j}$, then we are to show that

$$\sum_{i,j} u_i H_{ij} u_j \geq 0$$

for any vector u. This can be written out in full as

$$\text{(Prove:)} \qquad \sum_{i,j}\left(\text{Cov}[u_i X_i, u_j X_j] - \frac{\text{Cov}[u_i X_i, Y]\,\text{Cov}[u_j X_j, Y]}{\sigma_Y{}^2} \right) \geq 0$$

which after further manipulation gives

$$\text{(Prove:)} \qquad \text{Var}\left[\sum_i u_i X_i\right] \geq \frac{1}{\sigma_Y{}^2}\left(\text{Cov}\left[Y, \sum_i u_i X_i\right]\right)^2$$

A useful trick that we shall use in later work is to consider the function

$$Q(t) = \text{Var}\left[\sum_i (a_i - t u_i) X_i\right] = \text{Var}[Y] - 2t\,\text{Cov}\left[Y, \sum_i u_i X_i\right] + t^2\,\text{Var}\left[\sum_i u_i X_i\right]$$

Now $Q(t)$ must be positive for all values of t, because it is the variance of something. By "completing the square" we find that

$$Q(t) = \left(\sigma_Y - \frac{t}{\sigma_Y}\text{Cov}\left[Y, \sum_i u_i X_i\right]\right)^2 + t^2\left(\text{Var}\left[\sum_i u_i X_i\right] - \frac{1}{\sigma_Y{}^2}\text{Cov}\left[Y, \sum_i u_i X_i\right]^2 \right)$$

$$\min Q(t) = t^2\left(\text{Var}\left[\sum_i u_i X_i\right] - \frac{1}{\sigma_Y{}^2}\text{Cov}\left[Y, \sum_i u_i X_i\right]^2 \right) \geq 0$$

which is what we wanted to show.

It is worth noting that we have been a little sloppy in our terminology. Positive definiteness means that $\sum_{i,j} u_i H_{ij} u_j > 0$, not ≥ 0. In fact, no risk measure can satisfy the strict inequality, because it is a corollary of the homogeneity property of a risk measure R that

$$\sum_{i,j} a_i a_j \frac{\partial^2 R}{\partial a_i \partial a_j} = 0\,.$$

(We could have foreseen this in the construction of $Q(t)$, which vanishes at t=1 if the u's are equal to the a's.) All this says is that if we scale up all the exposures by the same relative amount then the portfolio risk scales up in a linear way too: but then it can't be strictly convex. In practice this is not an issue because such a scaling is forbidden in any sensible portfolio optimization: the total value must remain the same.

Conclusions

We have shown how portfolios work in a mean-variance framework and established that the "delta" of an asset, i.e. the sensitivity of standard deviation to asset allocation, is the covariance between the asset and the portfolio. We have also related this important result to the CAPM. However, mean-variance is not ideal for bond portfolios, in which the downside risk greatly exceeds the upside. There is a great deal of interest in "downside" risk measures in the risk management community and it is also apparent that the market puts a premium on downside risk: for example, in short-dated out-of-the-money put options. So in the next chapter we shall start looking at portfolio optimization in the VaR and CVaR (expected shortfall) frameworks, and showing how these compare with the mean-variance framework.

<div style="text-align: right">

13

</div>

Contributions to VaR and CVaR

In this chapter we continue with the subject of risk contributions. The risk contribution of an asset in a portfolio is fundamental to portfolio optimization, and it is defined by sensitivity of portfolio risk to fractional change in asset allocation. The classical mean-variance theory was developed by Markowitz in the 1950s and we have considered it in some detail in the last few chapters. It is a theory that is well-established and well-founded but, as we shall see here, it is not the best for dealing with very asymmetrical or fat-tailed distributions. Consequently Value-at-Risk (VaR) has become popular. However, as we shall discuss in depth here, VaR contribution is quite difficult to compute and certain portfolio models do not permit a sensible definition of it. For substantially the same reasons it is also very difficult to estimate VaR contribution from a Monte Carlo simulation. However, there are some more fundamental problems with VaR associated with its lack of convexity, which causes difficulties with portfolio optimization, and the only satisfactory way to resolve this is to move away from VaR and to use expected shortfall or "CVaR" instead.

We round off with an example that shows how VaR, CVaR and standard deviation contributions can be very different. VaR and CVaR contributions emphasize severe, rare risks (tail risks) whereas standard deviation contributions emphasize small, frequent risks (middle-of-distribution risks).

Recap: from standard deviation to VaR

We have spent some time in the last few chapters talking about standard deviation and its influence on financial theory: risk management, portfolio management, and the Capital Asset Pricing Model. When looking at the distribution of returns of complex financial instruments such as derivatives, some disadvantages become apparent. Take first Figure 13.1, which gives two attempts at modeling the distribution of P&L of a portfolio that holds one or more credit-risky bonds (long positions only).

Figure 13.1. Approximations to the statistical distribution of returns of a credit-risky bond (or portfolio). Dotted: Normal N(0,1); Solid: Gamma(2,1) shifted to make its mean zero. Both distributions have mean 0 and variance 1.
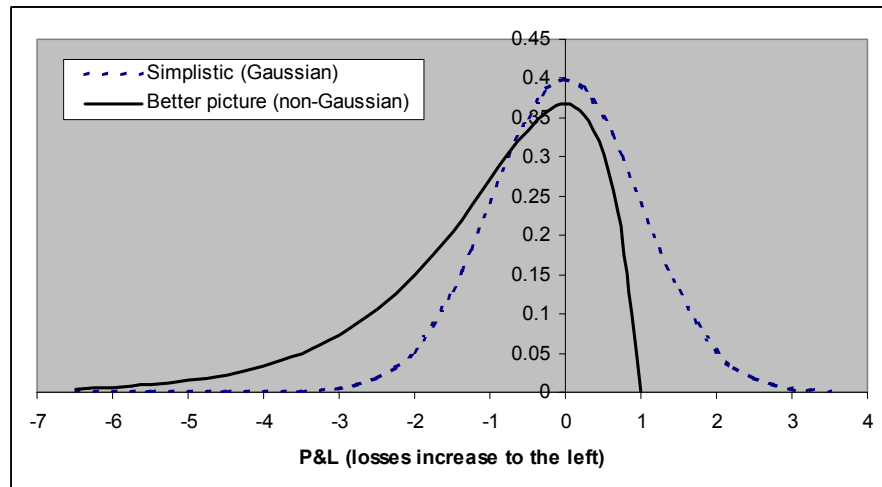
Early models of risk management, which persisted in times gone by when computers were slow and sophistication was low, assumed that asset returns were Normal (Figure 13.1, dotted line). On the other hand the true distribution may well be closer to that shown by the solid line. For bond portfolios in which the upside is small and the downside is big this is a considerably more realistic picture. The point about Figure 13.1 is that both the distributions have mean 0 and variance 1. So mean-variance does not tell them apart.

Now there is nothing intrinsically *wrong* with using standard deviation as a risk measure even when the distributions involved are skewed and/or fat-tailed. If one is worried mainly about common, not-too-severe events, and not about the tails, then standard deviation may well be appropriate. However, portfolio managers and risk managers often do care about downside risk. For one thing, regulatory authorities require them to care. In today's world, no investment bank would be permitted to state the market risk of its derivatives portfolio on the basis of Normal distribution assumptions[81]. Similarly, no investment bank would be permitted to state the credit risk of its loan book on the basis of Normal distribution assumptions. For investment-grade bond portfolios, the risk of default is perceived as quite low and the spread risk is more important. In that case, the Mertonian interpretation of a bond portfolio as a portfolio of options on the underlying firms is the most pertinent, and it makes sense to view the portfolio as a portfolio of derivatives.

In fact there is direct evidence that the market cares about downside risk too. For a start the equity option market shows "skew" (or smile) indicating that out-of-the-money put options are more expensive than a Normally distributed world would suggest.

---

[81] More precisely, if they do use a very unsophisticated methodology, the risk figures are increased by a "multiplier" chosen by the regulatory authority. A better methodology that has been satisfactorily back-tested incurs a lower multiplier.

Figure 13.2. Typical equity option skew ("smile").

Implied vol

Strike
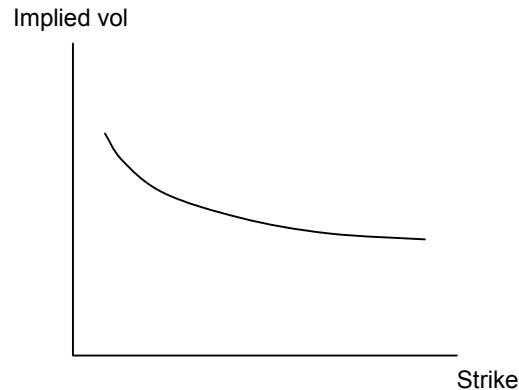
So option dealers implicitly care about non-Normal distributions. There are two possible explanations for this:

- The underlyings genuinely do have bigger downside, so put options have a greater expected payout, and hence a greater price.

- Supply and demand and risk aversion increase the desire of buyers, who are typically long-equity funds, to obtain downside protection. This causes the price to increase, and buyers will pay it even if on a purely actuarial basis it seems too high: in other words there is a risk premium.

In fact both explanations are generally accepted as true.

Now that we have decided that downside risk is important, we want to know how to quantify it. The Value at Risk, at lower[82] tail probability p, is given by y such that

$$P(Y < y) = p$$

In other words, it is the smallest level of loss that will be exceeded no more than $100p\%$ of the time, at some specified time horizon.

For example, with the distributions of Figure 13.1, the VaR at 98% confidence (p=0.02) is −2.05 for the Gaussian distribution and −4.8 for the non-Gaussian one, so VaR does recognize the difference in downside risk.

Having established what VaR is, we now wish to find its sensitivity to asset allocation because, as we have said previously, this is necessary in portfolio optimization.

---

[82] We are using the convention that loss is negative. In many articles on risk management loss is taken to be positive and then the VaR is in the right-hand tail of the distribution.

VaR contribution

Fundamentals

We assume that the portfolio is given as a weighted sum of assets, as $Y = \sum_{j=1}^{n} a_j X_j$.

We are going to prove the following result which states the sensitivity of the VaR of Y to the asset allocations a_j :

$$\frac{\partial}{\partial a_i} VaR[Y] = E[X_i \mid Y = VaR] \tag{1}$$

This means, in words, that the sensitivity of the VaR to the i'th allocation is equal to the expected loss from the i'th asset conditionally on the portfolio loss being equal to the VaR. In other words, we imagine all possible events that might occur, then forget about all of them except the ones in which we lose exactly whatever we have computed the VaR to be, and then find the average loss from the asset in question.

There are various ways of proving (1), but before we go into those in depth we shall make some observations. First, the sum of the VaR contributions (recall: risk contribution=allocation x sensitivity) is the VaR, by 1-homogeneity[83]:

$$\sum_{i=1}^{n} a_i \frac{\partial}{\partial a_i} VaR[Y] = E\left[\sum_{i=1}^{n} a_i X_i \mid Y = VaR\right] = E[Y \mid Y = VaR] = VaR$$

which is quite transparent. Next, let us recall the standard deviation result:

$$\frac{\partial}{\partial a_i} \sqrt{V[Y]} = \frac{1}{\sqrt{V[Y]}} Cov[X_i, Y]$$

For standard deviation, the risk contribution is given as the covariance between the asset and the portfolio: the higher the correlation between the two, the higher the contribution. Now take VaR: if the asset is likely to have much depreciated in value[84] when the portfolio has much depreciated in value, then the VaR contribution is large—and this is another way of expressing that the asset and portfolio are correlated. So in that sense, VaR and standard deviation contribution are conceptually similar. The difference is that VaR contribution does not look at all scenarios in the same way that standard deviation does: it targets only the situations in which the portfolio loses one precise amount of money[85]. This has some important ramifications.

Discrete models and the problems they cause

The first concerns any situation in which the loss distribution is discrete, by which we mean that the portfolio loss cannot take any possible value, but instead only a set of discrete values. An example is default-only models of credit risk, in which a bond either defaults or does not default. A more sophisticated example is if a bond can upgrade or

---

[83] "Portfolio optimization: The importance of convexity", Chapter 10.
[84] e.g. for a bond: downgraded, defaulted, spread greatly increased
[85] In a sense, the standard deviation contribution is a weighted average of the VaR contribution taken over all tail probabilities between 0 and 1: so the standard deviation looks at the whole loss distribution, and the VaR at just one slice through it.

downgrade to one of a finite number of rating states. Another situation in which this occurs is when the model is specifically designed to work on a "grid"[86]: for simplicity and ease of computation, the loss arising from the portfolio is assumed to lie on a grid of equispaced points.

To see what the consequence of discreteness is, consider the following picture of the loss distribution:

Figure 13.3a. VaR for discrete loss distribution.
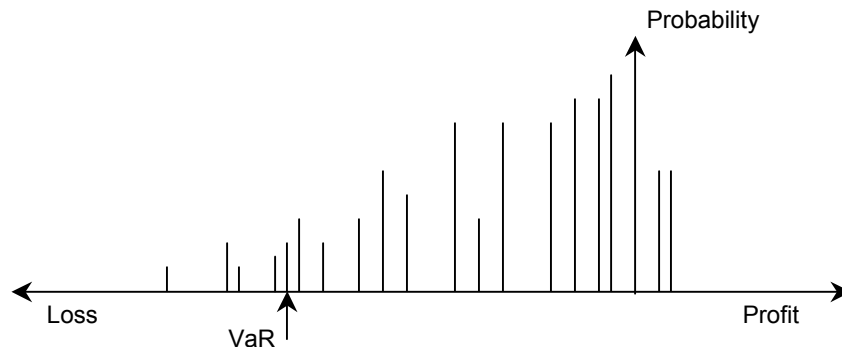
We define a "portfolio event" as a specific combination of individual asset events. For simplicity, we start by assuming that each possible portfolio event gives rise to a different portfolio loss. Then each "spike" in the figure represents one combination of events. When one of the asset allocations is altered slightly, the probability of a specific portfolio event does not change; however, the portfolio loss does: the height of each spike is unchanged, but the spikes move a little left or right. For small changes in asset allocation, the spike do not move past each other, i.e. their corresponding losses remain in the same order. Hence the VaR, which is determined by tail probability, corresponds to the same event as it did before the asset allocations were perturbed. So for small perturbations the VaR varies linearly in the asset allocations, and its sensitivity is just given by the values of the assets in the corresponding event. This is what (1) says (without the expectation symbol), because conditioning on $Y=$VaR tells us exactly what each asset did.

However, there is a fairly obvious problem. Suppose that, on the particular event corresponding to losing the VaR, one particular asset—call it asset #1—gives little loss, but that it has a large loss in the events that give rise to a similar portfolio loss. Suppose further that another asset, #2 say, is quite similar (e.g., a bond of similar rating). Then one can easily have a situation like that in Figure 13.3b.

---

[86] CreditRisk+ is an example. However, many portfolio models can be implemented in "grid" form, which has some computational advantages. Grid-based methods are very commonly used in CDO pricing as well. However, with those it is not very important to find the sensitivity w.r.t. asset allocation. We shall talk about these issues later.

Figure 13.3b. Inconsistent sensitivity of VaR for discrete loss distribution. At present event B determines the VaR

event A: asset#1 = -10, asset#2 = -5, ...

event B: asset#1 = -1, asset#2 = -8, ...

event C: asset#1 = -9, asset#2 = -3, ...

Loss

Profit

VaR
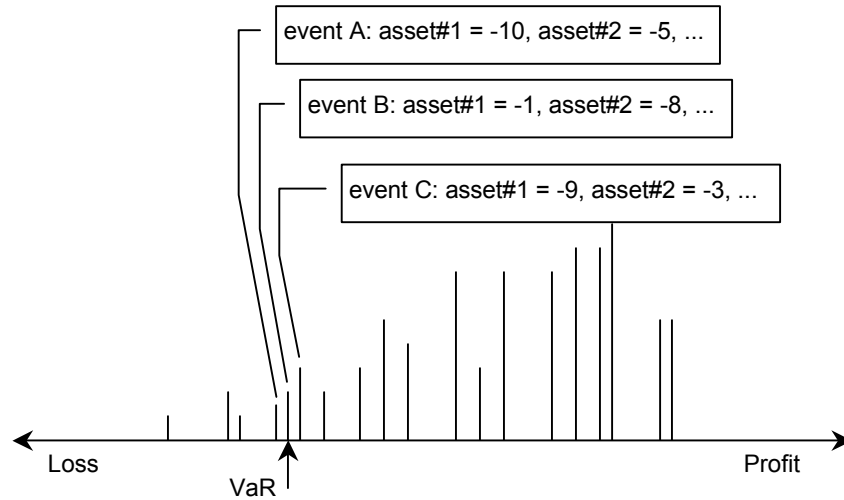
The VaR sensitivity of asset#1 is –1, and of asset#2 is –8. Given that they are similar assets, this looks very odd indeed. Further, a small change in the confidence level will alter the picture drastically: decreasing the tail probability will make the sensitivities –10 and –5, while increasing it makes them –9 and –3.

This makes VaR sensitivity extremely sensitive to tail probability, which is clearly undesirable: you do not want the picture to change by altering the confidence level from 98% to 98.1%. The consequence of this problem is that when larger changes are made to the asset allocations, and the "spikes" (Figure 13.3b) become re-ordered (Figure 13.3c), the VaR sensitivities suddenly jump (Figure 13.3d). This makes portfolio optimization very difficult indeed.

Figure 13.3c. Now event C determines the VaR, and the sensitivities have jumped

event A: asset#1 = -10, asset#2 = -5, ...

event B: asset#1 = -1, asset#2 = -8, ...

event C: asset#1 = -9, asset#2 = -3, ...

Loss

Profit

VaR
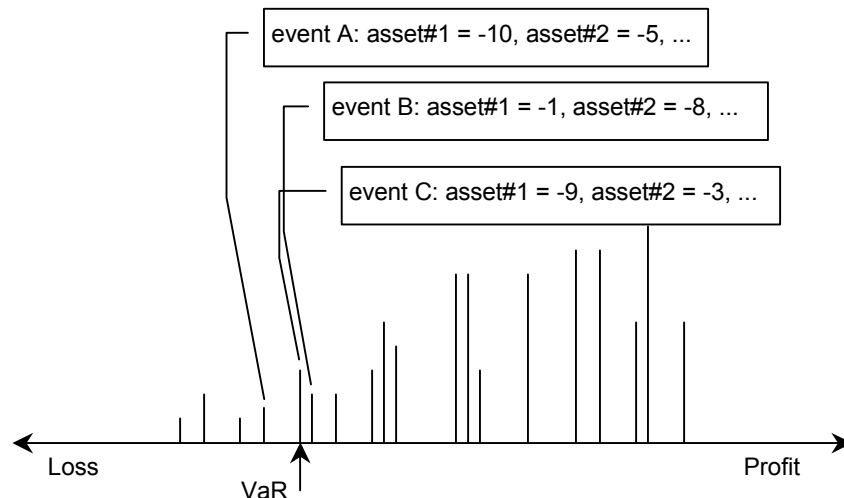
Figure 13.3d. Variation of VaR with asset allocation: an optimization nightmare

allocation in *j*th asset
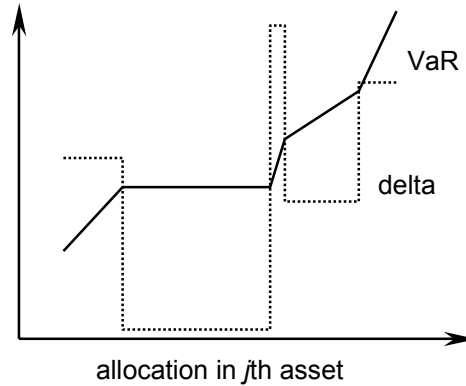
At present we have assumed that each event gives rise to a different level of portfolio loss. In practice more than one event may give rise to the same portfolio loss, e.g. in a portfolio of five assets, in one event the asset values might be (-1, -2, -9, -3, +1), and in another (-5, -3, -2, -2, -2): the portfolio losses are −14 in each case. For grid-based methods this sort of situation will almost certainly arise because the losses must be integer

Asset #	Possible gains/losses
1	+1, -2, -8
2	+2, -4, -10
3	+3, -2, -17
4	+2, -9
5	+2, -10

Source: Credit Suisse First Boston

multiples of the grid spacing. In that case, one might imagine that the formula for the VaR contribution would be given by averaging the contributions arising from all the events that gave rise to that particular loss (the VaR): indeed this is so, and that puts the expectation symbol into Eq. (1), which rounds off the "proof" nicely. However, it does little to alter the problem that we have identified above. Take for example the following situation, in which there are five assets and their possible gains and losses are given in the table above.

If the VaR is −19, then the combinations are

(+1,-4,+3,-9,-10), (+1,-10,-2,2,-10), (-2,-4,-17,+2,+2), (-8,+2,-17,+2,+2).

If it is −18, they are

(+1,-10,-2,-9,+2), (+1,+2,-2,-9,-10).

If it is −17, they are

(-8,-4,+3,+2,-10), (-2,-10,+3,+2,-10).

So for VaR=−17 the risk contribution of asset#4 is +2 (as both the combinations have that as the loss), at VaR=−18 it is −9, and at VaR=−19 it is somewhere in between +2 and −9, depending on the relative probabilities of the various combinations. The other assets behave, or rather misbehave, in the same way.

This points to the conclusion that trying to do portfolio analysis and optimization using grid-based models is not a very good idea because the question of finding a "sensible" VaR contribution is not well-posed. At a first glance, before one has appreciated this problem, it seems very reasonable to use them because moving the distribution onto a discrete grid hardly alters the distribution of losses. The catch is that the sensitivities to asset allocations can become totally messed up. Mathematically, it is possible for two functions to be very close, but for their gradients to be very different, as for example below, and this is the essence of the problem.

Functions are close; gradients are completely different
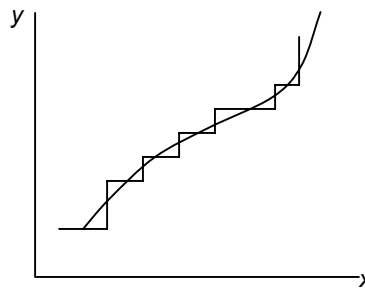
Monte Carlo simulation has all the same problems

The difficulties of defining and finding VaR contribution are not confined to discrete models. They also occur with *any* model, if Monte Carlo simulation is used. This is because the output of a Monte Carlo simulation is necessarily a (discrete) list of numbers. In finding the expected value of a particular asset given that the portfolio value (or loss in value) is equal to the VaR, we are faced with the problem that only in one simulation is the portfolio value equal to the VaR—so the expectation is being calculated from a single Monte Carlo simulation! Not surprisingly this gives useless results. Increasing the number of simulations has no effect.

The only remedy is to try and work with smooth, continuous distributions wherever possible. This means that if analytical methods are used, they should be ones that give continuous distribution approximations. If on the other hand Monte Carlo simulation is used, the distribution should be smoothed out. The method of *kernel estimation* has been suggested for this purpose[87] as a method of stabilizing Monte Carlo results, so we take it next.

Kernel estimation

The basic idea behind kernel estimation is that we use not just one simulation (i.e. that on which the portfolio value equaled the VaR) to estimate the conditional expected loss, but also those "close by" (i.e. simulations on which the portfolio value was close to the VaR).

---

[87] C. Gourieroux, J.-P. Laurent & O. Scaillet, "Sensitivity Analysis of Values-at-Risk", J. Emprirical Finance, 7:225-245, 2000.

The kernel estimate of the distribution is given by

$$\hat{f}(y) = \frac{1}{Nh} \sum_{k=1}^{N} \psi\left(\frac{y - y_k}{h}\right) \quad \text{(density)}$$

and

$$\hat{P}(Y < y) = \frac{1}{N} \sum_{k=1}^{N} \Psi\left(\frac{y - y_k}{h}\right) \quad \text{(tail probability)}$$

Here (y_k) denote the simulations; ψ is some probability density function chosen by the user (a Normal distribution, $\psi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ is the most widely used), Ψ is its associated cumulative probability function (so for a Normal distribution, $\Psi(x) = \Phi(x)$), and h is a width parameter. The effect is that each simulation becomes "fuzzy", and the distribution is smoothed out more as h is increased.

Fig. 13.5. Increasing levels of smoothing: h=0; h>0 small; h>0 larger.

To find the VaR sensitivities we just differentiate the tail probability with respect to the asset allocations:

$$\frac{\partial}{\partial a_i} \hat{P}(Y < y) = \frac{1}{Nh} \sum_{k=1}^{N} \left(\frac{\partial y}{\partial a_i} - x_{ik}\right) \psi\left(\frac{y - y_k}{h}\right);$$

here x_k is the value of the i'th asset in the k'th simulation. As the tail probability is to be constant, with y varying, the left-hand side is to be zero, and we arrive at

$$\frac{\partial y}{\partial a_i} = \frac{\dfrac{1}{Nh} \sum_{k=1}^{N} x_{ik} \psi\left(\dfrac{y - y_k}{h}\right)}{\dfrac{1}{Nh} \sum_{k=1}^{N} \psi\left(\dfrac{y - y_k}{h}\right)}$$

which is very similar to Eq. (1) in spirit: the expression is a weighted average of values of the i'th asset (X_i) using weights that kill off any simulations that did not give a portfolio loss close to the VaR ($\psi\left(\dfrac{y - y_k}{h}\right)$ is negligible unless $y \approx y_k$). We can think of the above expression as being $E[X_i \mid Y \approx VaR]$. And taking the limit $h \to 0$ does give us Eq. (1).

But does it work well in practice? Kernel estimation is certainly an improvement over using the raw simulation data, but it is not ideal, and to obtain stable estimates requires a very large number of simulations. Direct estimation from a model is a preferable approach, if the model is susceptible to analytical treatment. Although this is more difficult to implement, it gives excellent results, and we discuss it next.

Conditional independence models

Suppose that, conditionally on some risk factor (*V* is the notation we have been using in previous chapters, so we continue with it), the values of assets are independent. We have argued[88] that a fair approximation to the loss distribution can be obtained through reasoning that conditionally on *V* the loss distribution is roughly Normal (by the Central Limit Theorem). This leads to the following approximation to the tail probability:

$$P(Y < y) \approx E\left[\Phi\left(\frac{y - \mu_{Y|V}}{\sigma_{Y|V}} \right) \right] \;\; \text{or} \;\; P(Y < y) \approx \sum_k h_k \Phi\left(\frac{y - \mu_k}{\sigma_k} \right)$$

(The first expression is the general one and uses the conditional mean $\mu_{Y|V}$ and standard deviation $\sigma_{Y|V}$; the second one assumes that the distribution of *V* is discrete, i.e. there is a discrete set of risk-factor states. Note that it is quite OK for the distribution of the *risk factor* to be discrete; the important thing is that the distribution of the *portfolio loss* be continuous, which is an entirely separate matter.) As usual Φ denotes the cumulative Normal probability function, and ϕ (used later) is the density.

It is now a simple matter to find the VaR sensitivity, because all we have to do is differentiate with respect to asset allocation:

$$\frac{\partial}{\partial a_i} P(Y < y) = E^V\left[\left(\frac{1}{\sigma_{Y|V}} \frac{\partial y}{\partial a_i} - \frac{1}{\sigma_{Y|V}} \frac{\partial \mu}{\partial a_i} - \frac{y - \mu_{Y|V}}{\sigma_{Y|V}^2} \frac{\partial \sigma_{Y|V}}{\partial a_i} \right) \phi\left(\frac{y - \mu_{Y|V}}{\sigma_{Y|V}} \right) \right]$$

We wish the tail probability to be fixed and the VaR (i.e. *y*) to vary, so the left-hand side is to be zero; rearranging gives

$$\frac{\partial y}{\partial a_i} = \frac{E^V\left[\left(\frac{1}{\sigma_{Y|V}} \frac{\partial \mu_{Y|V}}{\partial a_i} + \frac{y - \mu_{Y|V}}{\sigma_{Y|V}^2} \frac{\partial \sigma_{Y|V}}{\partial a_i} \right) \phi\left(\frac{y - \mu_{Y|V}}{\sigma_{Y|V}} \right) \right]}{E^V\left[\frac{1}{\sigma_{Y|V}} \phi\left(\frac{y - \mu_{Y|V}}{\sigma_{Y|V}} \right) \right]}.$$

Computation of this is easy once the sensitivities of the conditional mean and variance to asset allocation are found; and as this is easy enough, the method is very workable.

General proof of Eq. (1)

We round off the discussion of the theory of VaR sensitivity with a general proof of the equation (1). This uses the *characteristic function*, which we introduced a while back[89]. Although the derivation is technical, it is of interest because a vast amount of theory can be developed using exactly the same machinery: first and second derivatives of expected shortfall and saddle-point analytics. The reader is referred to the Appendix.

---

[88] "Getting the full picture", Chapter 8.
[89] "Characteristically elegant", Chapter 6.

Second derivative of VaR

We pointed out in the chapter on optimization[90] that it is important for the portfolio risk to be a convex function of the asset allocations. VaR does not have this property, and that is nothing to do with the model, or how the VaR is calculated: it is simply a deficiency of VaR itself. Although we could go through the mathematics to show where things go wrong, the exercise is messy and not very useful and the reader is asked to take it on trust. When we say that VaR is not convex we do not mean that it goes wrong in all situations: for simple portfolio models it may well prove to be convex, and in certain types of portfolio approximation it always is. However, CVaR, as we shall see in a moment, is much better behaved and it makes sense to drop VaR in favor of it.

From VaR to CVaR

Recall that (if we adopt the convention that loss is negative) the expected shortfall or CVaR is defined by[91]

$$CVaR = E[Y \mid Y < VaR].$$

First derivative of CVaR

In a remarkable similarity with Eq. (1), CVaR has a sensitivity given by

$$\frac{\partial}{\partial a_i} CVaR[Y] = E[X_i \mid Y < VaR] \tag{2}$$

In other words, the CVaR sensitivity is the expected loss from the asset in question given that the portfolio loss is *worse than* the VaR. We prove Eq. (2) in the Appendix.

Again it is easy to show that the CVaR contributions add to give the CVaR:

$$\sum_{i=1}^{n} a_i \frac{\partial}{\partial a_i} CVaR[Y] = E\left[\sum_{i=1}^{n} a_i X_i \mid Y < VaR\right] = E[Y \mid Y < VaR] = CVaR.$$

The difference between Eqs. (1,2) is important. We have spent some time pointing out all the problems that are encountered in trying to estimate a conditional expectation based on an exact level of portfolio loss. The CVaR sensitivity requires the expectation to be conditional on a *range* of losses instead, and that is considerably easier. Obviously estimation from a Monte Carlo simulation is still quite "noisy", as for example if the tail probability is 1% and 10,000 simulations have been run then the expectation is being estimated from only 100 simulations—and the problem is worse for a lower tail probability. On the other hand it is possible to derive the sensitivity of the kernel estimate of the CVaR, and use that instead: that has an additional smoothing effect.

However, as with the VaR, the most satisfactory approach is to obtain the CVaR and its derivatives using conditional independence models. To pick up where we left off with the VaR, assume that there is some risk-factor V on which the assets are conditionally independent, and assume that Y is Normal conditionally on V. The CVaR is given by

$$E[Y \mid Y < y] = \frac{1}{P} E^V\left[\mu_{Y|V} \Phi(z_{Y|V}) - \sigma_{Y|V} \phi(z_{Y|V})\right]$$

---

[90] "Portfolio optimization: The importance of convexity", Chapter 10.
[91] Sometime a minus sign is inserted in front, but this makes no real difference. The only thing to watch for is the convexity later on.

with $z_{Y|V} = \dfrac{y - \mu_{Y|V}}{\sigma_{Y|V}}$; y denotes the VaR and P is the lower tail probability. By differentiating we obtain the neat result for the sensitivity:[92]

$$\frac{\partial}{\partial a_i} E[Y \mid Y < y] = \frac{1}{P} E^V \left[\frac{\partial \mu_{Y|V}}{\partial a_i} \Phi\left(z_{Y|V}\right) - \frac{\partial \sigma_{Y|V}}{\partial a_i} \phi\left(z_{Y|V}\right) \right].$$

Second derivative of CVaR

Another nice result, which again looks like Eq. (1,2), is the formula for the second derivative (or Hessian as it is often called) of CVaR:

$$\frac{\partial^2}{\partial a_i \partial a_j} CVaR[Y] = -Cov[X_i, X_j \mid Y = VaR] \times \frac{f(y)}{P} \tag{3}$$

In other words, the second derivative is the conditional covariance of the two assets given that the portfolio loss is *equal to* the VaR, multiplied by a factor which we won't bother about ($-1 \times$ density of the loss distribution at the VaR $\div$ tail probability). As it is known that the covariance matrix of any two random variables must be positive definite (see Chapter 12), convexity has been proven[93]. We prove Eq. (3) in the Appendix to this chapter.

Notice that if we want to calculate this second derivative, all the problems that we had with the VaR sensitivity come back to haunt us, because the conditioning is on an exact level of loss (Y=VaR in Eq. (3) cf. Y<VaR in Eq. (2)). So for discrete models the Hessian has the undesirable property of jumping as the asset allocations are changed, and as the tail probability (used to define the CVaR) is changed. Also it is difficult to estimate from Monte Carlo simulations.

On the other hand it is fine in the conditional independence framework. Differentiating the expression for first derivative:

$$\frac{\partial^2}{\partial a_i \partial a_j} E[Y \mid Y < y] = \frac{-1}{P} E^V \left[\left(\frac{\partial^2 \sigma_{Y|V}}{\partial a_i \partial a_j} + \sigma_{Y|V} \frac{\partial z_{Y|V}}{\partial a_i} \frac{\partial z_{Y|V}}{\partial a_j} \right) \phi\left(z_{Y|V}\right) \right].$$

The resulting expression for the Hessian is necessarily convex[94]. This concludes our examination of CVaR.

---

[92] There is a bit more to this than meets the eye. It might look as though we have missed out some terms from here, arising from the fact that $z_{Y|V}$ depends on the asset allocations. Actually, those terms all cancel.

[93] In fact, because of the minus sign in (3), the risk measure is concave. But this is only because we have defined risk as being negative here, which is a trivial issue. All the properties needed for optimization are obeyed.

[94] The first part of the expression is the Hessian of the standard deviation risk measure, and the second is a "perfect square".

Example

We are going to demonstrate in this section that standard deviation, VaR and CVaR give different pictures of the split-down of risk in a portfolio.

For simplicity we have taken a simple model of a portfolio of defaultable bonds or loans. Much of what we are going to demonstrate applies to other asset classes as well, but this example is more easily explained.

Table 13.1 (next) shows the portfolio, in which loans are assumed either not to default, or to default without recovery (so the loss is either 0 or 1, times the exposure). The loans are correlated according to the Gaussian copula model that we have seen before. This means that a Normally distributed risk factor is assumed and the conditional default probability of each loan is $\Phi\left(\dfrac{\Phi^{-1}(p_i) - \beta_i V}{\sqrt{1 - \beta_i{}^2}}\right)$ where p_i is the expected default frequency

(EDF) and β_i is the correlation parameter of the ith loan. Each loan is made to a different issuer. As can be seen from the table, different assets have different exposures, correlations, and EDFs. Also assets 5, 8, 14 have big exposures and low EDFs: this is done deliberately. In the context of a real portfolio this is quite reasonable as it is normal to have bigger positions in (or, lend more to) higher-grade credits.

Table 13.1. Test portfolio of loans and results of portfolio calculations. Note particularly #5, #8, #14, which have much bigger exposures and lower default probabilities; this makes them "tail risks". Note also that #36-#50 are less correlated with the rest of the portfolio than #21-#35.

	Allocation (M$)	EDF	Correl
1	1	2.00%	0.5
2	2	1.00%	0.5
3	1.3	2.00%	0.5
4	1	10.00%	0.5
5	25	0.10%	0.5
6	2	1.00%	0.5
7	3	1.00%	0.5
8	12	0.50%	0.5
9	2	0.70%	0.5
10	2	0.70%	0.5
11	3.4	0.70%	0.5
12	1.7	3.00%	0.5
13	2.1	1.00%	0.5
14	15	0.20%	0.5
15	2	0.70%	0.5
16	2	0.70%	0.5
17	3.4	0.70%	0.5
18	1.7	3.00%	0.5
19	2.1	1.00%	0.5
20	2	0.70%	0.5
21	2	0.70%	0.7
22	3.4	0.70%	0.7
23	1.7	3.00%	0.7
24	2.1	1.00%	0.7
25	2	0.70%	0.7
26	2	0.70%	0.7
27	3.4	0.70%	0.7
28	1.7	3.00%	0.7
29	2.1	1.00%	0.7
30	2	0.70%	0.7
31	2	0.70%	0.7
32	3.4	0.70%	0.7
33	1.7	3.00%	0.7
34	2.1	1.00%	0.7
35	2	0.70%	0.7
36	2	0.70%	0.3
37	3.4	0.70%	0.3
38	1.7	3.00%	0.3
39	2.1	1.00%	0.3
40	2	0.70%	0.3
41	2	0.70%	0.3
42	3.4	0.70%	0.3
43	1.7	3.00%	0.3
44	2.1	1.00%	0.3
45	2	0.70%	0.3
46	2	0.70%	0.3
47	3.4	0.70%	0.3
48	1.7	3.00%	0.3
49	2.1	1.00%	0.3
50	1.5	0.80%	0.3

Figure 13.6. Skewed loss distribution from test portfolio
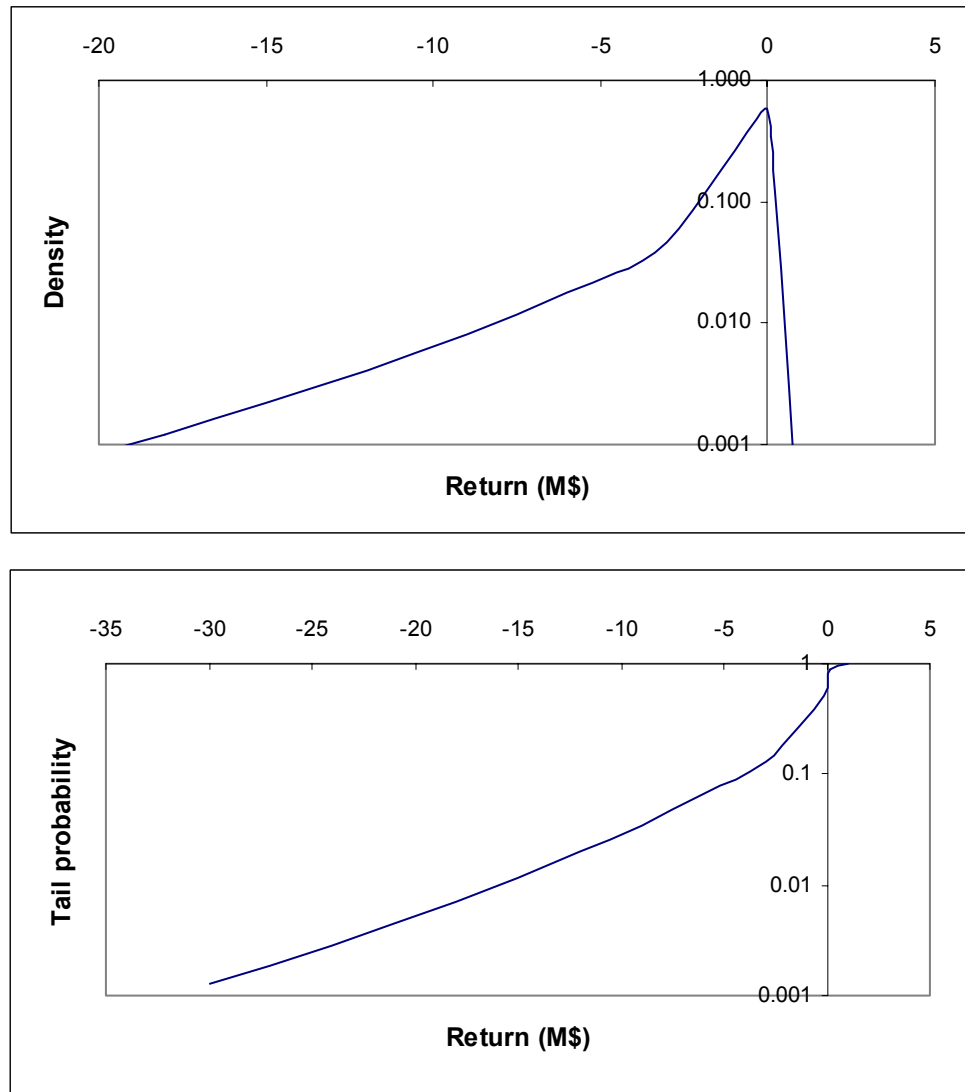
Figure 13.6 shows the loss distribution. From the top picture (probability density as a function of loss: note logarithmic scale on vertical axis) it is clear that the downside is much bigger than the upside. The lower picture shows the probability mass and this allows you to work out the VaR by eye (though not the CVaR). For example at 99.5% confidence, the tail probability is 0.005 and the VaR is about –$20.25M. The CVaR is incidentally –$27.7M (it must be a greater loss than the VaR).

Figure 13.7 (next page) shows the risk contributions using three different risk measures: standard deviation, VaR(99.5%), CVaR (99.5%). There are two obvious points to be made.

First, assets 21-35 have higher contributions, by all measures, than assets 36-50. As the only difference between these two groups is correlation, this must be the explanation. We have already decided on theoretical grounds that the higher the correlation of an asset the higher its risk contribution, so this comes as no surprise.

Secondly, VaR and CVaR identify assets 5, 8, 14 as tail risks and they receive much more "attention" than when standard deviation is used. Because these issuers are very unlikely to default they do not contribute a huge amount to the variance. In fact, asset #4 does not have the highest risk contribution as measured by standard deviation, but very clearly does by VaR and CVaR.

The consequence of this for a portfolio manager is important. For an optimal portfolio, risk contribution must be in constant ratio to excess return, across the portfolio. As VaR and CVaR penalize tail risks more than standard deviation does, it must be the case that in the VaR- or CVaR-optimal portfolios the tail-risky assets will have lower allocations than they will in the standard-deviation-optimal portfolio.

Figure 13.7. Risk contributions by standard deviation, VaR and CVaR for test portfolio. (The signs of the VaR and CVaR contributions have been reversed to make them positive.) Less correlated assets have a lower risk contribution (compare #21-#35 with #36-#50), by all three measures. Big exposures to unlikely risks are heavily penalized by VaR and CVaR, compared with standard deviation.
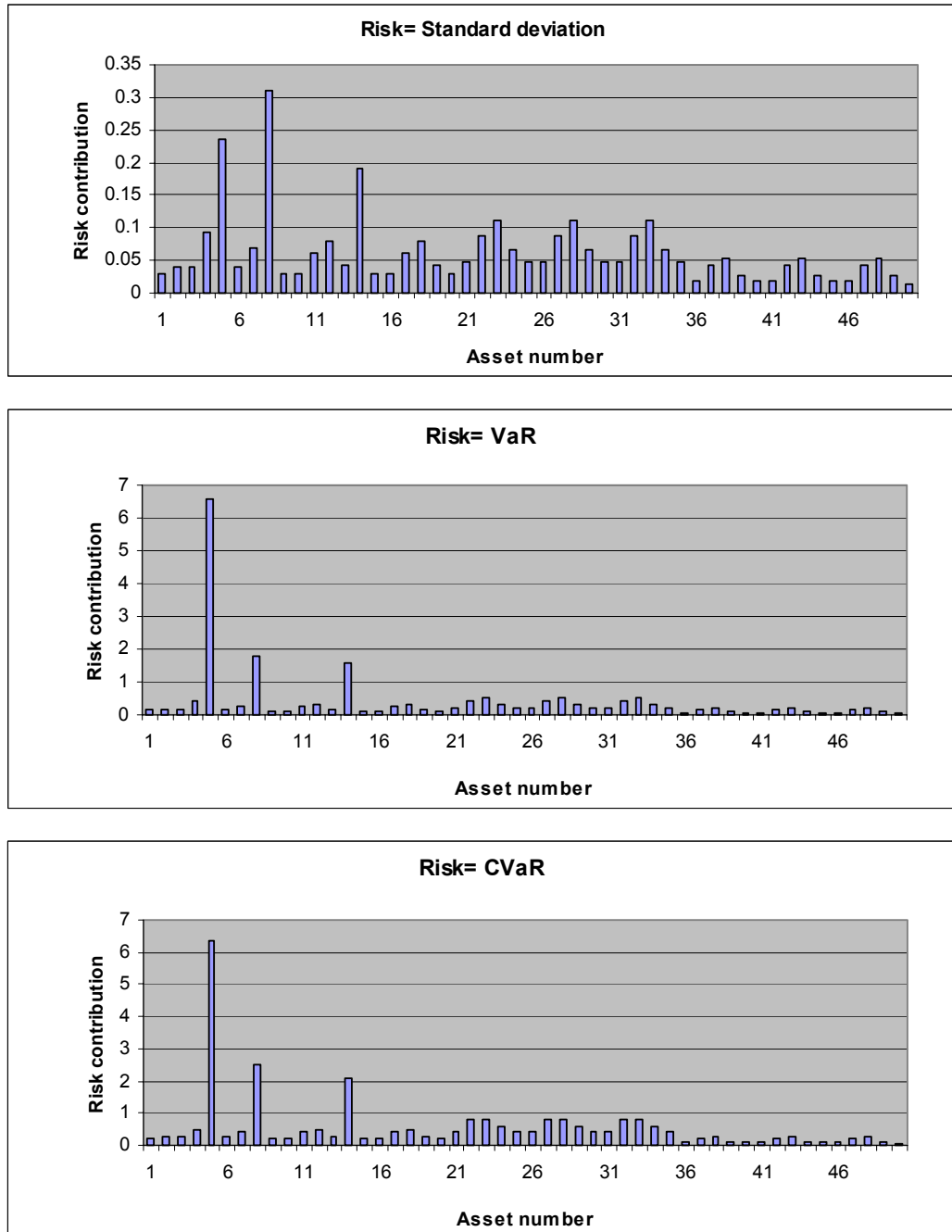
Conclusions

We have covered quite a lot of ground in this chapter and the conclusions can be summarized as follows.

VaR/CVaR vs. standard deviation

- First, VaR and CVaR both give a more complete picture of potential loss than standard deviation does. This is particularly relevant for distributions that have appreciable downside risk.

CVaR as an improvement on VaR

- VaR is very difficult to estimate satisfactorily from Monte Carlo simulation of a model.

- CVaR is theoretically sounder than VaR because of convexity. This does not mean that VaR will always fail to be convex—often it is—but it is not possible for CVaR to exhibit non-convexity. Importantly, the *estimated* CVaR will always be convex too, provided it is estimated "in a sensible way".

- Using CVaR does not completely remove all the difficulties associated with estimating VaR; rather, CVaR is "one order of differentiability smoother than" VaR. This means that -

    o estimating the first derivative of CVaR is about as awkward as estimating the VaR;

    o estimating the second derivative of CVaR is about as awkward as estimating the first derivative of VaR; and so on.

- VaR and CVaR can be cleanly estimated from analytical conditional independence models, and this is why we regard these as the best way to do portfolio analysis. These techniques have the additional advantage of not introducing simulation error.

VaR and CVaR vs. standard deviation

- VaR and CVaR penalize severe but unlikely risks much more than standard deviation does. This is not intuitively surprising, but a mathematical explanation is quite difficult (though an elegant one is given in the reference cited below[95]). Consequently if VaR or CVaR is used as a risk measure then these tail risks will seem less attractive, from a risk/return perspective, than if standard deviation is used.

So beware: altering your risk measure causes your optimal portfolio to vary.

---

[95] R. Martin et al: "VaR: Who contributes and how much?", RISK, August 2001.

Appendix

General proof of Eq. (1): VaR sensitivity

Recall that the characteristic function is defined by

$$C(\omega) = E[e^{i\omega Y}], \quad i = \sqrt{-1},$$

and that the density of Y is recovered from the characteristic function by the inverse Fourier integral

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega x} d\omega.$$

The tail probability can be obtained by integrating with respect to x (from $-\infty$ to y):

$$P(Y < y) = \frac{-1}{2\pi i} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} \frac{1}{\omega} d\omega.$$

(This needs careful interpretation because the integral runs through a singularity at $\omega = 0$, but we shall not go into the details[96].) Differentiating w.r.t. the j'th asset allocation gives

$$\frac{\partial P}{\partial a_j} = \frac{-1}{2\pi i} \int_{-\infty}^{\infty} \frac{\partial C(\omega)}{\partial a_j} e^{-i\omega y} \frac{1}{\omega} d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial y}{\partial a_j} C(\omega) e^{-i\omega y} d\omega.$$

Now we want the tail probability to remain fixed, while the VaR (i.e. y) varies, so the left-hand side must be zero. Also we can write out

$$\frac{\partial C(\omega)}{\partial a_j} = \frac{\partial}{\partial a_j} E[e^{i\omega Y}] = \frac{\partial}{\partial a_j} E\left[e^{i\omega \sum a_j X_j}\right] = E\left[\frac{\partial}{\partial a_j} e^{i\omega \sum a_j X_j}\right] = i\omega E[X_j e^{i\omega Y}]$$

and combine to get

$$\frac{\partial y}{\partial a_j} = \frac{1}{2\pi} \int_{-\infty}^{\infty} E[X_j e^{i\omega Y}] e^{-i\omega y} d\omega \bigg/ \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} d\omega$$

or

$$\frac{\partial y}{\partial a_j} = E\left[X_j \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(Y-y)} d\omega\right] \bigg/ \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} d\omega$$

The integral in the numerator is the Fourier representation of the delta function (spike) at $Y=y$, and we already know the denominator (as the density function at y) so the result is

$$\frac{\partial y}{\partial a_j} = E\left[X_j \delta(Y-y)\right] \big/ f(y) = E[X_j \mid Y = y], \quad \text{Q.E.D.}$$

A large amount of theory rests on this method of proof, including as we shall see later the results for expected shortfall, and all the saddle-point methodology. We have not talked about the saddle-point method yet, as it is very advanced, but it can be viewed as a method of handling the Fourier integrals used in the above derivation.

---

[96] Essentially one perturbs ω at the origin to give it a *positive* imaginary part.

General proof of Eq. (2): CVaR sensitivity

To work out the CVaR we first need to calculate, for the distribution of *Y*, the expected payout of a put option "struck" at the VaR.

VaR (=*y*)　　　　　　*Y*
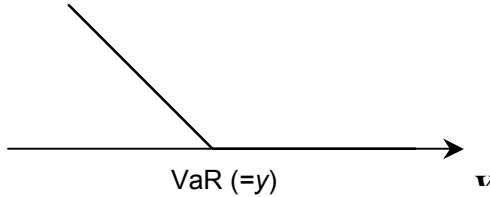
This can be obtained by integrating the tail probability function

$$P = P(Y < y) = E[1(Y < y)] = \frac{-1}{2\pi i} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} \frac{1}{\omega} d\omega \qquad (*)$$

with respect to *y*:

$$Put = E[\max(y - Y, 0)] = \frac{-1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} \frac{1}{\omega^2} d\omega.$$

The CVaR is obtained from the equation

$$CVaR = VaR - \frac{Put}{P} = y + \frac{1}{2\pi P} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} \frac{1}{\omega^2} d\omega$$

where *P* is the lower tail probability. Differentiating with respect to asset allocation gives

$$\frac{\partial}{\partial a_j} CVaR = \frac{\partial y}{\partial a_j} + \frac{1}{2\pi P} \int_{-\infty}^{\infty} \frac{\partial C(\omega)}{\partial a_j} e^{-i\omega y} \frac{1}{\omega^2} d\omega + \frac{\partial y}{\partial a_j} \frac{1}{2\pi i P} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega y} \frac{1}{\omega} d\omega$$

and now the first and third terms cancel, because of (*). Using the result for $\partial C(\omega)/\partial a_j$ that we obtained before, we arrive at

$$\frac{\partial}{\partial a_j} CVaR = \frac{E[X_j 1(Y < y)]}{P(Y < y)} = E[X_j \mid Y < y], \quad \text{Q.E.D.}$$

General proof of Eq. (3): Second derivative of CVaR

We have obtained that

$$\frac{\partial}{\partial a_j} CVaR = \frac{1}{2\pi P} \int_{-\infty}^{\infty} \frac{\partial C(\omega)}{\partial a_j} e^{-i\omega y} \frac{1}{\omega^2} d\omega$$

and differentiate again:

$$\frac{\partial^2}{\partial a_j \partial a_k} CVaR = \frac{1}{2\pi P} \int_{-\infty}^{\infty} \frac{\partial^2 C(\omega)}{\partial a_j \partial a_k} e^{-i\omega y} \frac{1}{\omega^2} d\omega - \frac{\partial y}{\partial a_k} \frac{1}{2\pi P} \int_{-\infty}^{\infty} \frac{\partial C(\omega)}{\partial a_j} e^{-i\omega y} \frac{i}{\omega} d\omega$$

Previous working shows that the second term tidies up to give $\dfrac{\partial y}{\partial a_k}\dfrac{\partial y}{\partial a_j}$, and we already know the first derivatives of VaR. Also

$$\frac{\partial^2 C(\omega)}{\partial a_j \partial a_k} = \frac{\partial^2}{\partial a_j \partial a_k} E[e^{i\omega Y}] = E\left[\frac{\partial^2}{\partial a_j \partial a_k} e^{i\omega \sum a_j X_j}\right] = -\omega^2 E[X_j X_k e^{i\omega Y}]$$

Thus

$$\frac{\partial^2}{\partial a_j \partial a_k} CVaR = -\frac{E[X_j X_k \delta(Y = y)]}{P} + E[X_j \mid Y = y]E[X_k \mid Y = y]\frac{f(y)}{P}$$

$$= -Cov[X_j, X_k \mid Y = y] \times \frac{f(y)}{P}$$

as required.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Appendix

Single-name Credit risk modeling

Overview of models

There are three main quantitative approaches to single-name credit risk modeling:

- **Reduced form models** concentrate on the spread between non-risky and risky rates. They deal with debt directly (a great strength) and say nothing about equity (a weakness). Default arrival is described by a Poisson process and is totally unpredictable. The key ingredients of these models are interest rates, hazard rates and recovery value.

  Reduced form models were studied by Artzner & Delbaen (1995), Jarrow & Turnbull (1995) and Duffie & Singleton (1999).

- **Structural models** concentrate on the evolution of a firm's assets. Default happens when assets fall below a threshold (either discrete or continuous). In general, default is predictable and overnight default risk is close to zero. The firm is "dissolved" at maturity. Equity is priced as a call option on the assets of the company, while debt is priced as the present value of the terminal debt minus a put option on the assets. Clearly, the sum of equity and debt equals the value of the assets. These category of models are used in practice for a variety of purposes (KMV, CUSP, Credit Grades, etc.)

  Structural models have been particularly successful recently, for two main reasons: firstly, they are able to unify equity and debt in a single model; secondly, they naturally handle the non-linear payoff of corporate securities and, in particular, the asymmetrical nature of credit risk caused by the possibility of default.

  The original option based structural model was developed by Merton (1974) on the basis of Black & Scholes's work (1973). Black, Fischer & Cox (1976) expanded the model by introducing a barrier triggering default before maturity. Advanced structural models featuring jump-diffusion characteristics, where a firm can default unexpectedly due to sudden drops in assets value, were introduced by Zhou (1997) and subsequently developed by Kijima & Suzuki (2001) and Hilberink & Rogers (2002).

- **Hybrid models** combine the reduced form and structural approaches. They model default as a completely unexpected event characterized by an instantaneous default probability like the reduced form models, but they express this probability as a function of the share price, somehow bringing the capital structure of the company into the picture. Hybrid models are particularly useful to model securities that combine equity and debt features, such as convertible bonds. They have important implications for purely equity products and elegantly explain volatility skews.

We now review the different categories of models in greater detail, providing some highlights about the underlying theory. We start with reduced form models, move on to hybrid models, and finally thoroughly examine structural models.

Reduced form models

Reduced form models assume that default occurs without warning at an exogenous default rate or intensity, and is therefore unpredictable. These models do not try to explain the reasons why default occurs, and the intensity model is inferred from market prices.

Assumptions

The dynamics of the intensity are specified under the risk-free probability Q and all expectations are calculated under this probability measure (the notation used is $E^Q[\cdot]$).

Taking as given the random default time τ, we define the default process N by

$$N_t = 1_{\{\tau \le t\}} = \begin{cases} 1 & \text{if} \quad \tau \le t \\ 0 & \text{if} \quad \text{not} \end{cases}$$

The default process described is increasing, namely it has an upward trend: the conditional probability at time t that the firm defaults by time $s \ge t$ is not less than N_t itself. We can decompose N_t in two separate processes: a martingale (whose value at time t equals the conditional expectation of future values) and a so-called *compensator* A^τ (which captures the upward trend in N). The compensator describes the cumulative, conditional likelihood of default, and it is parameterized through a non-negative process λ by setting

$$A_t^\tau = A_{\min(t,\tau)}, \quad \text{with} \quad A_t = \int_0^t \lambda_s ds \,.$$

The variable λ_t describes the conditional default rate or *intensity*. Under the assumption that the intensity λ is constant, the default process is a homogenous Poisson process with intensity λ, stopped at its first jump. Thus τ is exponentially distributed with parameter λ and the pricing (or risk-neutral) probability of default is given by

$$q(T) = 1 - e^{-\lambda T} \,.$$

Given the default probability, it is possible to calculate the intensity as

$$\lambda = \frac{d(T)}{1 - q(T)},$$

where d is the density of q. In the view of this representation, in the statistical literature λ is often called hazard rate.

On the other hand, if $\lambda = \lambda(t)$ is a deterministic function of time, then N is an inhomogeneous Poisson process with intensity λ, stopped at the first jump. The default probability is given by

$$q(T) = 1 - e^{-\int_0^T \lambda(u)du}.$$

Finally, if $\lambda = (\lambda_t)$ is a stochastic process such that conditional on the realization of the intensity, N is an inhomogeneous Poisson process stopped at its first jump, then N is a Cox process, or doubly-stochastic Poisson process. The conditional default probability given the intensity path up to time T is

$$q(T) = 1 - e^{-\int_0^T \lambda_u du},$$

and the expected default probability can be expressed as

$$q(T) = 1 - E^Q\left[e^{-\int_0^T \lambda_u du}\right].$$

Valuation

Reduced form models have the desirable characteristic of producing tractable valuation formulas. If we consider a zero coupon bond paying 1 at maturity T if there is no default, or R at time T if default occurs, the bond price is

$$B_0^T = e^{-rT} E^Q\left[R\, 1_{\{\tau \le T\}} + 1_{\{\tau > T\}}\right] = e^{-rT} - e^{-rT}(1 - R)q(T).$$

The variable R is the *recovery value* of the bond, namely the fraction of the principal amount that the investor manages to get back in case of default, and $q(T)$ is the market-implied default probability. It is interesting to point out that the value of such a bond is the value of an otherwise equivalent risk-free bond, minus the present value of the expected loss due to default. Even more interesting is the case where the intensity λ is constant and recovery value is zero, in which case the valuation formula becomes

$$B_0^T = e^{-rT}(1 - q(T)) = e^{-(r + \lambda)T},$$

where we can calculate the price of the defaultable zero coupon bond the same way we would valuate a risk-free bond, only making sure to increase the discount rate by λ.

It is worth pointing out once again that default is modeled as being always unexpected. This characteristic is one of the major strengths of reduced form models. We will see that for basic structural models the short-term default probability is close to zero, which is in contrast with the empirical evidence. On the contrary, in the case of reduced form models default can happen at any point in time, and this naturally explains the greater than zero observed short-term credit spreads.

Hybrid models

Hybrid Models build upon the results of the Reduced Form models by defining specific functions for the hazard rate λ. In these models the hazard rate, namely the instantaneous probability of default, is typically expressed as a function of the share price. This allows us to tie the likelihood of default into information reflecting the current fundamental value of the firm. To this extent Hybrid Models move away from the Reduced Form models' approach – which do not try to explain the reasons of default at all – and take a step in the direction of Structural Models, which we will examine in the next section.

Assumptions

In order to define the hazard rate function, it is convenient to express the equity value of a firm in terms of units of a so-called "money market account" $B(t)$. This can be done simply by dividing the stock price $S(t)$ by

$$B(t) = e^{\int_0^t r(u)du},$$

($r(u)$ is the continuously compounded interest rate earned on the account), obtaining

$$s(t) = \frac{S(t)}{B(t)}.$$

Under this assumption the discounted price is modeled as a martingale with

$$ds(t) = \sigma s(t)dW(t)$$

where $W(t)$ is a standard one-dimensional Brownian motion.

We can therefore express the hazard rate as a function of this relative stock price. Madan and Unal (1998) make the assumption that

$$\lambda(t) = \phi(s(t))$$

$$= \frac{c}{\left[\ln\left(\frac{s(t)}{\delta}\right)\right]^2}.$$

If δ lies below the current value of $s(t)$, then λ is a decreasing function of the equity value, with the hazard rate tending to infinity as $s(t)$ approaches δ. This simply means that the instantaneous risk of default increases as the equity value goes down. On the other hand, if δ is above the current equity value, then a positive relation is possible. This may be appropriate when the value of the equity increases as a consequence of increased assets volatility, therefore creating greater risk for the bond-holders (see the Structural Models section for greater details about viewing contingent claims on a company's assets as options). The exact value of δ has to be determined empirically.

Valuation

We have seen before how it is possible to derive the risk-neutral probability of default when the hazard rate is a stochastic function of time $\lambda(t)$. Obviously, the risk-neutral probability of *no* default is simply the complement of $q(t,T)$, namely

$$F(t,T) = E^Q\left[e^{-\int_t^T \lambda(u)dt}\right].$$

The price of the defaultable bond, as a function of $F(t,T)$, is therefore given by

$$B_t^T = e^{-r(T-t)}E^Q[F(t,T) + R(1 - F(t,T))], \qquad (A.1)$$

where R is again the recovery value.

Given the assumptions we have made we are in a Markov setting, and the pricing probability of default can be expressed as

$$F(t,T) = \psi(s,t),$$

where s is the discounted equity price as previously defined. It is easy to observe that

$$e^{-\int_0^t \lambda(u)du}\psi(s)$$

is a risk neutral martingale and hence by Itô's lemma the function ψ has to satisfy the partial differential equation

$$\frac{1}{2}\sigma^2 \frac{\partial^2 \psi(s,t)}{\partial s^2} + \frac{\partial \psi(s,t)}{\partial t} = \phi(s)\psi(s,t), \qquad \text{subject to } \psi(s,T) = 1.$$

The solution is given by

$$\psi(s,T) = G_a\left(\frac{2}{d^2(s,T)}\right)$$

$$d = \frac{\ln(s/d)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2}$$

$$a = \frac{2c}{\sigma^2}$$

and $G_a(y)$ satisfies the ordinary differential equation

$$y^2 G'' + \left(\frac{3y}{2} - 1\right)G' - aG = 0, \qquad \text{subject to } G(0) = 1, \ G'(0) = -1.$$

Once the function $\psi(s,t)$, and therefore $F(t,T)$ has been determined, plugging it into Equation A.1 gives the price of the defaultable bond.

Structural models

In 1974, Robert Merton introduced a simple model that links equity to debt. This model relies upon the sharing rule of the firm's assets between two main classes of claimholders, the shareholders and the bondholders. Models of this type have since become known as structural models, as these describe the capital structure of the firm.

Assumptions

Merton assumed an underlying process that represents the total value of the firm's assets A. The equity and debt claims are priced off this process. The main hypotheses of Merton's framework are:

1. The value of the firm follows a standard geometric Brownian motion (GBM), thus providing the same mathematical tools available in the Black-Scholes world. The process is described in mathematical terms as follows:

$$\frac{dA_t}{A_t} = \mu dt + \sigma dW_t, \qquad A_0 > 0,$$

   where μ is a drift parameter, σ is a volatility parameter and W is a standard Brownian motion. Applying Itô's lemma the above equation can be re-written in finite terms:

$$A_t = A_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

2. The total value of the firm is financed by equity E and one representative zero-coupon bond B_T, maturing at T with face value F, so we get

$$A_T = B_T + E_T.$$

3. The priority rule cannot be violated: shareholders obtain a positive payoff only in the case that debt-holders are completely reimbursed.

4. There are neither cash flow payouts nor issues of any type of security during the life of the debt or bankruptcy costs. It is also assumed that the firm can't be liquidated before maturity. This implies that default can only happen at maturity.

Valuation of the equity and debt claims

In the event the face value payment is not met, i.e., the firm defaults, the bondholders receive the entire value of the firm and the equity-holders of the firm receive nothing. Hence the value at maturity of the bond B_T is given by

$$B_T = \min(A_T, F).$$

Figure A.1 represents the payoff diagram for B_T.

Figure A.1
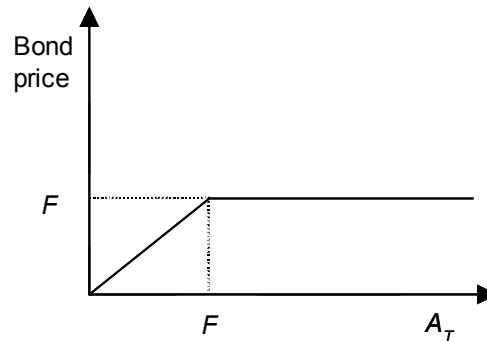
Source: CSFB

We can also write the above equation as

$$B_T = A_T - \max(A_T - F, 0)$$

from which we can identify E as

$$E_T = \max(A_T - F, 0).$$

Figure A.2 represents the payoff diagram for E_T.

Figure A.2
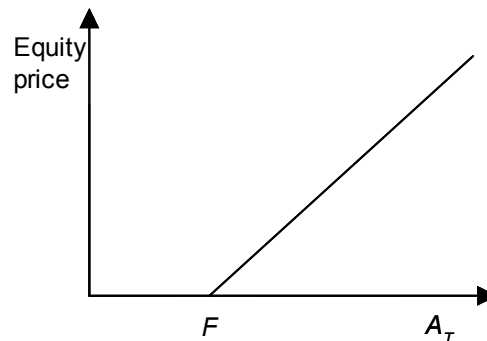
Source: CSFB

The value of the equity is then equal to that of a European call option c_0 on the assets of the company, where the maturity of the option corresponds to the maturity of the firm and the principal payment on the debt F corresponds to the strike of the option (see Figure A.2 above).

In the Black-Scholes economic setting the value of the equity is

$$E_0 = c_0 = A_0 \Phi(d_1) - e^{-rT} F \Phi(d_2),$$

where

$$d_1 = \frac{(r + \sigma^2/2)T - \ln(F/A_0)}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T},$$

$\Phi(\cdot)$ is the standard normal cumulative distribution function and r is the risk-free rate.

We can also write B_T as

$$B_T = F - \max(F - A_T, 0).$$

Therefore, the value of the bond at time 0 is

$$B_0 = Fe^{-rT} - p_0$$

where p_0 is the value of a European put struck at F maturing at T. The value of the put is

$$p_0 = -A_0\Phi(-d_1) + e^{-rT}F\Phi(-d_2).$$

So the debt-holder is a holder of a risk-free loan and short a put option on the firm's assets. It is easy to verify, using put-call parity, that the market value identity holds

$$A_0 = E_0 + B_0.$$

These equations clearly show how, although both equity and debt values depend on the leverage of the firm (the F/A_0 ratio embedded in the d_1 and d_2 variables), the assets value does not.

Default Probability
We have seen before that the assets value at time t is given by

$$A_t = A_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

We can use this result to calculate the default probability of the bond, namely the probability that the assets value falls below the face value of debt F at maturity. Since W_T is normally distributed with a zero mean and variance T, the default probability $p(T)$ is

$$p(T) = P[A_T < F] = P[\sigma W_T < \ln(F/A_0) - mT] = \Phi\left(\frac{\ln(F/A_0) - mT}{\sigma\sqrt{T}}\right),$$

where, once again, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $m = (\mu - \sigma^2/2)$.

Credit Spread
Finally, let us derive a formula for the credit spread. The credit spread is the difference between the yield of a defaultable zero-coupon bond and the yield of an otherwise equivalent default-free bond. It is a measure of the compensation the investors require for taking on the risk associated to the potential default losses. The relationship between the yield and the price at time t of a bond maturing at time T is

$$B(t,T) = e^{-y(t,T)(T-t)}.$$

The spread $S(t,T)$ at time t is therefore

$$S(t,T) = -\frac{1}{T-t} \ln\left(\frac{B_t^T}{\overline{B}_t^T}\right), \qquad T > t,$$

where $\overline{B}_t^T$ is the price of the default-free bond maturing at time T. The term structure of credit spreads is the schedule $S(t,T)$ against T, keeping t fixed. Under the Black-Scholes economic setting the price of the riskless bond is

$$\overline{B}_t^T = Fe^{-r(T-t)},$$

therefore the term structure is given by

$$S(0,T) = -\frac{1}{T} \ln\left(\Phi(d_2) + \frac{A_0}{F} e^{rT} \Phi(-d_1)\right), \qquad T > 0,$$

which is a function of maturity T, assets volatility σ, the leverage ratio F/A_0 and the risk-free rate r.

Advanced structural models

One problem of the traditional structural Merton model is that default can only happen at maturity. Even if the value of the firm gets very close to zero (before maturity), there is no mechanism to trigger default. This does not reflect the economic reality, as usually there are covenants in place that grant to credit-holders the right to reorganise the company if its value goes below a certain level.

To get around this undesirable behavior, a **knock-out threshold** or **barrier** can be introduced, at the level of the firm's debt. If the value of the assets drops below the barrier at any point in time, the firm defaults. Thus, default can occur during the life of the firm, as opposed to the traditional structural models, in which default can only happen at maturity. With this "trick" we can generate the entire default curve, i.e., the default probabilities over different time horizons. This allows us to price any debt instrument of any maturity using the calculated default probabilities, and to make relative value decisions between several issues of the same issuer. In principle, any risky security can be priced using the default curve.

Default Probability
If we allow the bond to default before maturity, when the value of the assets falls below the threshold D (which we assume to be less than the current asset value A_0), then the time at which the bond defaults is a continuous random variable defined as

$$\tau = \inf\{t > 0 : A_t < D\}.$$

The default probability is then

$$p(T) = P[M_T < D] = P\left[\min_{s \leq T}(ms + \sigma W_s) < \log(D/A_0)\right],$$

where M is the minimum of the firm values $M_t = \min_{s \leq t} A_s$ and $m = (\mu - \sigma^2/2)$.

Since the distribution of the minimum of an arithmetic Brownian motion is inverse Gaussian, the default probability is given by

$$p(T) = \Phi\left(\frac{\ln(D/A_0) - mT}{\sigma\sqrt{T}}\right) + \left(\frac{D}{A_0}\right)^{\frac{2m}{\sigma^2}} \Phi\left(\frac{\ln(D/A_0) + mT}{\sigma\sqrt{T}}\right).$$

Another limitation of traditional structural models is that default can be anticipated. This is because the geometric Brownian motion assumed for the value of the firm is continuous at any point in time and can only "diffuse" towards the level of the firm's value that triggers default, therefore making it to some extent predictable. A continuous diffusion process as in Merton's original model has therefore (almost) zero probability of hitting the barrier over a small time interval. Consequently, structural models predict a null short-term credit spread (a defaultable bond is not riskier than a Government Bond if we know it will not default overnight), which is in contrast with empirical observation.

A second modification can be introduced to solve the short-term spread problem. The value of the firm still follows a standard geometric Brownian motion, but on top of it a **jump process** is added. This captures the fact that default could be triggered by a sudden, unexpected negative event. The jump process represents the possibility of the firm instantaneously defaulting due to the arrival of catastrophic information regarding, for example, litigation or fraud.

Figure A.3 represents two possible paths followed by the firm's assets where the above modifications have been introduced. When the value of the assets hits the barrier before maturity, the firm defaults. Also, a jump can occur, in which case the firm can potentially default instantly. This type of model is called a jump-diffusion model.

Figure A.3
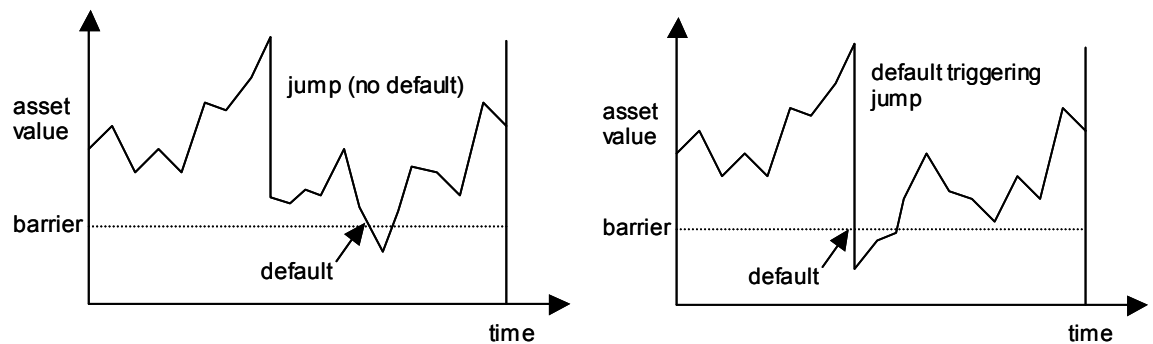
Source: CSFB

Finally, another advantage of advanced structural models over the traditional ones is their flexibility of calibration. Although the original Merton's model is over-determined (it's not possible to calibrate the model consistently to both the equity market and the credit market simultaneously), models characterized by jump-diffusion and continuous barrier are not, and can be tuned to fit equity and credit prices simultaneously.

The first jump-diffusion model was proposed by Zhou (1997). It extends Merton's original model by replacing the GBM with a general jump-diffusion process

$$dA / A = (\mu - \lambda\kappa)dt + \sigma dW + (\Pi - 1)dN, \qquad A(0) = A_0 \,,$$

where N is a Poisson process with intensity λ and Π is an i.i.d. log-normal variable such that $\ln \Pi \sim N(\mu_\pi, \sigma_\pi^2)$. This implies that

$$\kappa = E[\Pi - 1] = e^{\mu_\pi + \sigma_\pi^2 / 2} - 1.$$

This stochastic process has two components: one that characterizes the normal fluctuation in the firm value, caused by the continuous flow of economic information into the market, and another that describes the sudden changes, the jumps in the assets value, due to the arrival of unexpected important information. The jump-diffusion process seems therefore appropriate to model a firm's default risk: the market value of the assets move continuously most of the time, but default can be triggered at any time by a sudden, dramatic drop in the firm's enterprise value.

The key difficulty with this approach is the lack of tractability, which is due to the fact that barrier problems for jump-diffusion are very difficult to solve. Zhou used the brute force *Monte Carlo* method, which is extremely slow (especially if calibration is involved). This methodology involves simulating thousands of possible random paths that the assets of the firm can follow, computing the payoff to the investor in each single one, discount it back and finally averaging out the results. The simulation is run under the risk-neutral, market implied probability measure Q, employing the stochastic process above rewritten in the following form

$$dX / X = (r - \lambda\kappa)dt + \sigma dW + (\Pi - 1)dN \,,$$

where $X = A / F$. The equation can also be restated as

$$d \ln X = (r - \sigma^2 / 2 - \lambda\kappa)dt + \sigma dW + \ln(\Pi)dN \,.$$

CUSP: CSFB's structural model

CSFB's original Credit Underlying Securities Pricing model (CUSP) employs a structural approach to the modeling of credit risk.

CUSP is an extension to the original Contingent Claim model proposed by Merton. As shown before, Merton modeled equity as a call option on the firm's assets and debt as a risk-free loan plus a short put option on the firm's assets. He used a simplified model of the firm, with only two classes of securities: stocks and a zero-coupon bond maturing in one year (whereas CUSP uses a 10-year bond). At the end of the year, the company liquidates itself. If the value of the firm's assets at the time is above the face amount of the debt, the equity holders pay off the liability and walk away with the residual. On the other hand, if the asset value of the firm falls below the face amount of the debt, the equity holders declare bankruptcy and walk away with nothing, while the debt holders receive the recovery value of the firm. The fact that equity and corporate debt are both derivatives of a same underlying asset implies certain theoretical relationship between the two should hold (i.e. put-call parity).

The main limitation of Merton's model is that its implementations, such as CUSP, are necessarily issuer-based, rather than issue-based: they cannot differentiate risk/relative value over different issues of the same issuer, e.g., a 5-year bond vs. a 10-year bond. Consequently, CUSP models a 10-year liquid bond as a proxy for the "generic" company's credit.

Unlike other structural credit models, which focus upon default probabilities, CUSP's objective is to model the risk and reward associated with movements in credit spreads. This makes CUSP ideally suited as a tool for high-grade credit markets, where spread risk rather than default risk is the central issue. CUSP has affirmed itself as a market standard in credit structural modeling, alongside products such as KMV and CreditGrades.

CUSP's output measures are the Spread Widening Risk (SWR) and the Probability Weighted Return (PWR). SWR is CUSP's basic risk statistic, defined as the amount that the credit spread will widen given a one standard deviation decline in the equity price over a 3-month horizon. Again, SWR is calculated at the issuer level - i.e. one statistic for each issuer. PWR, on the other hand, is CUSP's basic return measure, defined as the expected bond return (for a hypothetical 10-year par bond). The PWR represents the debt 3-month total return averaged over the distribution of possible stock outcomes as given by the current equity volatility. The PWR incorporates both the change in debt price and carry over a 3-month horizon for a par 10-year bond.

A significant improvement that CUSP introduces over the original Merton's model is a "volatility smile" parameter. When the asset value of a highly leveraged firm goes below a certain level, CUSP automatically boosts the asset's volatility in order to obtain a higher SWR than would otherwise be calculated. Empirical testing shows that this extension to the original model fits better the market data.

At the present day there are more than 800 names modeled in CUSP that are made available via the web[97]. The delivery platform we have developed is highly accessible and interactive: the CUSP's universe can be explored in several different ways, and the data can be analyzed in both numerical and graphic form.

Typical uses of website include:

- listing issuers from a specific sector or portfolio and sort them by SWR, to identify the riskiest credits according to CUSP;

- sorting by SWR z-score to identify the most improved and least improved risks according to CUSP;

- sorting by PWR for CUSP's relative value ranking of the most attractive and least attractive issuers in a sector or portfolio;

- viewing historical charts of PWR, SWR, LIBOR Spread, Stock Price and Stock Volatility at an individual or aggregate level.

More details about CUSP, the way the model works and its possible uses can be found in the technical document "Introduction to CUSP" available on the website. Alternatively, you can contact directly the CUSP Team[98].

---

[97] The CUSP website is accessible to all CSFB's customers at www.cusp.com using the usual CSFB Research & Analytics ID and password.

[98] **US:** Jonathan Song, 1-212-325-1768, jonathan.song@csfb.com; Jason Rayman, 1-212-325-4908, jason.rayman@csfb.com; David Crick, 1-212-325-3070, david.crick@csfb.com. **Europe:** Recai Gunesdogdu, 44 207 883 7978, recai.gunesdogdu@csfb.com; Derek Hynes, 44 207 888 8107, derek.hynes@csfb.com. **Asia:** Yusuke Ueda, 81 3 5777 7174, yusuke.ueda@csfb.com

CUSP+: CSFB's advanced structural credit risk model

CUSP+ gets around most of the limitation that apply to the traditional structural models and to CUSP.

The two main changes introduced in CUSP+ are a knock-out threshold, which allows for the possibility of a default to happen before the debt matures, and a jump-diffusion process for the firm's value, which introduces the short-term component of credit risk – the firm can default suddenly and unpredictably.

In CUSP+ the knock-out threshold or barrier is equal to the face value of debt, while jumps are modeled such that the firm value drops to 40% of the face value of debt (recovery value), automatically triggering default. (Figure A.4). The main benefit of these two simplifications is that we developed a closed form solution to the problem, allowing us to avoid computational intensive (and rather imprecise) Monte Carlo simulations.

The "technical" innovations not only reflect more closely the real behavior of credit in the marketplace, they also allow for a more flexible calibration of CUSP+. The model can be calibrated to both the equity market and the credit market, depending on the investor's view. For example, if you believe the equity markets are correctly priced, you can calibrate the model to the equity price and two points on the option volatility surface, and imply the credit curve. Or you can calibrate to the credit market, and the model would generate the entire equity volatility surface. As a third alternative, if you believe the 6-month ATM volatility and 5-year CDS spread are both liquid, i.e., "correct", you might actually calibrate the model to these two prices plus the stock price. You can then find relative value on the credit curve off the 5-year point. These methodologies are illustrated in Figure A.5.

Hedge ratios can also be calculated from the model as well, as they represent the sensitivities or partial derivatives of the instrument with respect to the underlying parameters of the model.

Figure A.4

Figure A.5

---

equity market	calibration →	CUSP +	implies →	credit market *(spread curve)*
credit market	calibration →	CUSP +	implies →	equity market *(volatility surface)*
equity / credit market	calibration →	CUSP +	implies →	equity / credit market *(volatility surface, and spread curve)*
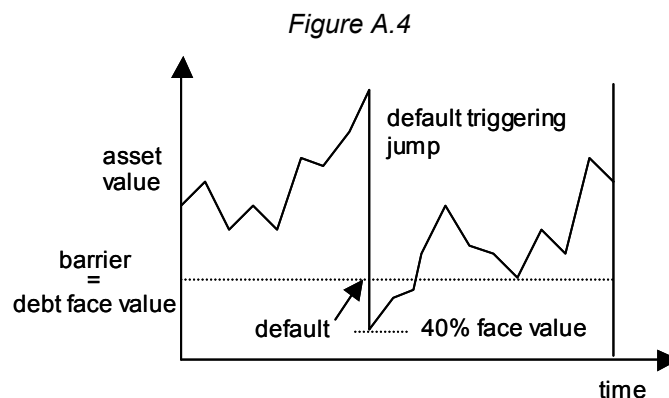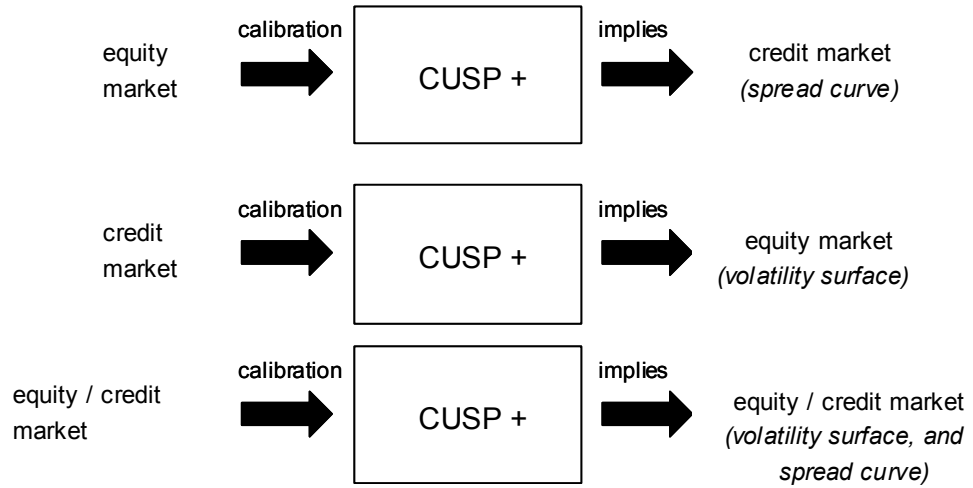
Source: CSFB

Analyst Certification

The analysts identified in this report each certify, with respect to the companies or securities that the individual analyzes, that (1) the views expressed in this report accurately reflect his or her personal views about all of the subject companies and securities and (2) no part of his or her compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this report.

CUSP® Certification and Definition

With respect to the analysis in this report based on the CUSP® methodology, CSFB certifies that (1) the views expressed in this report accurately reflect the CUSP® methodology and (2) no part of the firm's compensation was, is, or will be directly related to the specific views disclosed in this report.

CUSP® is an analytical model that relates to an issuer's capital structure, stock price, and the option-implied volatility of its shares to credit risk. CUSP® provides systematic monitoring of credit risk from forward-looking, market-based measures (one standard deviation spread widening, SWR) and relative-value tracking that incorporates both risk and return (probability weighted return, PWR). Details on the CUSP® methodology can be found on www.cusp.com or by contacting the CSFB Fixed Income Portfolio Strategy Group.

Important Disclosures

The analyst(s) involved in the preparation of this research report received compensation that is based upon various factors, including CSFB's total revenues, a portion of which are generated by CSFB's Investment Banking and Fixed Income Divisions.

CSFB policy generally prohibits analysts and members of their households from having a financial interest in the securities or derivatives of issuers that are the subject of this report.

CSFB policy generally prohibits analysts and members of their households from becoming an officer, director, or advisory board member of an issuer that is the subject of this report.

CSFB and/or its affiliates may have managed or co-managed a public or Rule 144A offering of securities for an issuer that is the subject of this report.

CSFB and/or its affiliates may trade as principal in the securities or derivatives of the issuers that are the subject of this report.

Recommendation Definitions

Buy: Indicates a recommended buy on our expectation that the issue will be a top performer in its sector.

Outperform: Indicates an above-average total return performer within its sector. Bonds in this category have stable or improving credit profiles and are undervalued, or they may be weaker credits that, we believe, are cheap relative to the sector and are expected to outperform on a total-return basis. These bonds may possess price risk in a volatile environment.

Market Perform: Indicates a bond that is expected to return average performance in its sector.

Underperform: Indicates a below-average total-return performer within its sector. Bonds in this category have weak or worsening credit trends, or they may be stable credits that, we believe, are overvalued or rich relative to the sector.

Sell: Indicates a recommended sell on the expectation that the issue will be among the poor performers in its sector.

Restricted: In certain circumstances, CSFB policy and/or applicable law and regulations preclude certain types of communications, including an investment recommendation, during the course of CSFB's engagement in an investment banking transaction and in certain other circumstances.

Risk Category Definitions

In addition to the recommendation, each issue may have a risk category indicating that it is an appropriate holding for an "average" high yield investor, designated as **Market**, or that it has a higher or lower risk profile, designated as **Speculative** and **Conservative**, respectively.

CSFB Credit Rating Definitions

CSFB assigns rating opinions to investment-grade and crossover issuers. Ratings are based on our assessment of a company's creditworthiness and are not recommendations to buy or sell a security. The ratings scale (AAA, AA, A, BBB, BB) is dependent on our assessment of an issuer's ability to meet its financial commitments in a timely manner. Within each category, creditworthiness is further detailed with a scale of High, Mid, or Low – with High being the strongest sub-category rating: **High AAA**, **Mid AAA**, **Low AAA** – obligor's capacity to meet its financial commitments is extremely strong; **High AA**, **Mid AA**, **Low AA** – obligor's capacity to meet its financial commitments is very strong; **High A**, **Mid A**, **Low A** – obligor's capacity to meet its financial commitments is strong; **High BBB**, **Mid BBB**, **Low BBB** – obligor's capacity to meet its financial commitments is adequate, but adverse economic/operating/financial circumstances are more likely to lead to a weakened capacity to meet its obligations. **High BB**, **Mid BB**, **Low BB** – obligations have speculative characteristics and are subject to substantial credit risk. CSFB's rating opinions do not necessarily correlate with those of the rating agencies.

Options Disclaimer

Structured securities, derivatives and options are complex instruments that are not suitable for every investor, may involve a high degree of risk, and may be appropriate investments only for sophisticated investors who are capable of understanding and assuming the risks involved. Supporting documentation for any claims, comparisons, recommendations, statistics or other technical data will be supplied upon request. Any trade information is preliminary and not intended as an official transaction confirmation. Use the following links to read the Options Clearing Corporation's disclosure document: http://www.cboe.com/LearnCenter/pdf/characteristicsandrisks.pdf

Because of the importance of tax considerations to many option transactions, the investor considering options should consult with his/her tax advisor as to how taxes affect the outcome of contemplated options transactions.

CREDIT SUISSE | FIRST BOSTON

AMERICAS	EUROPE	ASIA PACIFIC
1 212 325 2000	44 02 7888 8888	852 2101 6000