



City Research Online

City, University of London Institutional Repository

Citation: Onorato, M. (2005). Essays on credit risk, risk adjusted performance and economic capital in financial institutions. (Unpublished Doctoral thesis, City University London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/8452/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk



Cass Business School
City of London

**Essays on Credit Risk, Risk Adjusted Performance
and Economic Capital in Financial Institutions**

**A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN FINANCE**

Mario Onorato

London 2005

Acknowledgements

A dissertation does not just appear out of nowhere, and although it is supposed to be a contribution by one person for a PhD, there are still a lot of people who have helped me out over the years. I have been fortunate enough to have had the support of so many people and without it this would not have been possible. While most people did not help directly on the project, every one of them contributed in some way towards helping me to get where I am today, even things like just being a friend and going out and having fun. For everyone listed here I am eternally grateful for their help.

The research that has gone into this thesis has been thoroughly enjoyable. That enjoyment is largely a result of the interaction that I have had with my supervisors and colleagues. I feel very privileged to have worked with my supervisors Alistair Milne and Giovanni Barone Adesi. To each of them I owe a great debt of gratitude for their patience, inspiration and friendship. The Department of Finance at Cass Business School has provided an excellent environment for my research. I spent many enjoyable hours with department members and fellow students chatting about my latest crazy ideas over a cup of coffee. Without this rich environment I doubt that many of my ideas would have come to fruition.

I would like to thank Father Francis of Certosa di Pesio and my friends there for their prayers, wisdom, and advice about all aspects of life.

A special thanks also to my family who have been extremely understanding and supportive of my studies. It is through their encouragement and care that I have made it through all the steps to reach this point in life, and I couldn't have done it without them. My family has always taken care of me and I love them all very much.

Most importantly of all, I would like to thank my wonderful wife Grazia, who has put up with my crazy PhD life style for the last three years. My research has resulting in many late nights, and some of my travelling has kept me away for weeks at a time. She has always been there supporting me and understood the importance of my work. She is the lovely part of my life and keeps me normal. I'd like to thank her who has encouraged me so much over the years.

Table of Contents

List of Tables.....	VIII
List of Figures.....	IX
INTRODUCTION.....	1
<u>CHAPTER ONE:</u>	
An Integrated Pricing Model for Defaultable Loans and Bonds.....	6
1. Introduction.....	7
2. Defaults and Transition Risk.....	8
2.1 Actuarial based method.....	10
2.2 The equity based method.....	11
3. Recovery by seniority of debt risk.....	12
4 The theoretical integrated pricing model.....	14
4.1The stochastic interest rate term structure model.....	16
4.2 The stochastic spread term structure model.....	16
4.2.1 Risk-neutral probability transition matrix	19
4.2.2 Multiple time horizon.....	24
4.2.3 One-period and cumulative transition probabilities.....	25
4.3 Spread term structure by rating class contingent to the seniority of debt.....	27
5. Applications.....	28
6. Conclusions.....	29

CHAPTER TWO:

A Framework for Risk Measures and Risk Adjusted Performance.....30

1. Introduction.....	31
2. Overview of the literature.....	34
3. Appealing Features of Risk Measures: Coherence and Stochastic Dominance.....	41
3.1 Stochastic Dominance.....	42
3.1.1 First Degree Stochastic Dominance (FDSD).....	43
3.1.2 Second Degree Stochastic Dominance (SDSD).....	43
3.1.3 Remarks on Stochastic Dominance.....	44
3.2 Coherence.....	46
4. Taxonomy of risk measures: an end user's perspective.....	47
4.1 Risk as Maximum Potential Loss.....	49
4.2 Risk as the magnitude of deviation from a target: two sided and one sided risk measures.....	51
4.2.1 LPM when $\alpha = 0$ and the Expected Value is the target value τ	54
4.2.2 LPM when $\alpha = 1$ and the Expected Value is the is the target value τ	55
4.2.3 LPM when $\alpha = 2$ and the Expected Value is the target value τ	56
4.2.4 LPM when $\alpha = 1,2$ and the Expected Value is $\tau = E(X)$	56
4.2.5 Conditional measures of shortfall risk.....	58
4.2.5.1 Conditional Tail Expectation, CTE.....	58
4.2.5.2 Tail Value at Risk.....	58
4.2.5.3 Mean Excess Loss (MEL) or Expected Shortfall (ESF).....	59
4.2.5.4 Remarks.....	60

4.3	Utility theory based risk measures.....	61
5.	Distortion risk measures: definition, examples and properties.....	63
5.1	The Quantile Distorted Risk Measure.....	65
5.2	The TVaRp(X) Distorted Risk Measure.....	65
5.3	The Wang transform risk measure.....	66
5.4	Concave distortion risk measures.....	67
5.5	The Beta distortion risk measure.....	68
5.6	The relationship between stochastic dominance and ordered distortion risk measures.....	68
6.	Risk adjusted performance measure.....	69
6.1	Sharpe Ratio.....	69
6.1.1	Information Ratio.....	72
6.2	Treynor Ratio.....	72
6.3	Jensen alfa J_α	73
6.4	The Return on Value at Risk (ROVaR).....	74
6.5	Generalised Sharpe Ratio (GSR).....	76
6.6	RAP measures based on downside risk.....	77
6.6.1	Sortino Ratio.....	77
6.6.2	Upside-Potential Ratio.....	78
6.6.3	The Generalised Sortino Ratio: the Kappa and Omega RAP.....	79
6.6.4	Remarks.....	80
7.	Portfolio Management and Risk Adjusted Performance.....	81

7.1 Portfolio management when using quadratic utility theory and Mean-Variance principle.....	81
7.2 Portfolio management when using RAP downside risk measure.....	84
7.3 A new RAPM based on Distorted Risk Measures: Portfolio and Capital Management Perspectives.....	85
8. Conclusions.....	87
Appendix 1.....	90
Appendix 2.....	92
Appendix 3.....	94
Appendix 4.....	96

CHAPTER THREE:

**Default probability, capital structure and the cost of economic
capital.....**

1. Introduction.....	100
2. Literature Review.....	101
3. Economic capital and performance measurement.....	109
3.1 Alternative definitions of A conceptual framework for attributing capital	110
3.2 A conceptual framework for attributing capital.....	113
3.2.1 The Merton model of default.....	113
3.2.2 Economic capital and value measurement in financial institutions: the bondholders perspective.....	114

3.2.3	Economic capital and value measurement in financial institutions: the shareholders perspective.....	116
4.	Capital structure and cost of equity with constant default probability.....	117
4.1	Modelling assumptions and notations.....	118
4.2	The determination of promised debt repayments.....	121
4.3	Capital structure and cost of equity under unlimited liability.....	123
4.4	Capital structure and cost of equity (risk capital) under limited liability.....	128
4.4.1	Limited shareholders liability and deposits are insured.....	128
4.4.2	Limited shareholders liability and deposits are uninsured.....	131
4.5	A comparative analysis of leverage and cost of risk capital with fixed default probability, limited and unlimited liability.....	133
5.	From risk capital to economic capital: introducing franchise value and taxes.....	134
5.1	Cost of Risk Capital with Franchise Value.....	134
5.2	Tax and franchise value impact on cost of risk capital when deposits are uninsured.....	136
5.3	Tax and franchise value impact on cost of risk capital when deposits are insured.....	137
6.	Capital structure and cost of target risk capital under limited and unlimited liability, insured deposits, franchise value and taxes.....	138
6.1	The determination of promised debt repayments under unlimited liability and target credit rating.....	138
6.2	Capital Structure under unlimited liability and target credit rating.....	139
6.3	Cost of target risk capital under unlimited liability, limited liability and insured deposits.....	140

6.4 Cost of economic capital under limited liability, insured deposits, franchise value and taxes.....	143
7. Conclusions and further research.....	144
Appendix 1.....	147
References.....	148

List of Tables

CHAPTER ONE: An Integrated Pricing Model for Defaultable Loans and Bonds

Table 1: Recovery Rates by Seniority and Original Bond Rating, 1971-2001.....13

CHAPTER TWO: A Framework for Risk Measures and Risk Adjusted Performance

Table 1: Example of Degrees of the Lower Partial Moment. 57

Table 2: Sharpe Ratio comparison: single investment vs portfolio71

Table 3: Sharpe Ratio Jensen Alfa and Treynor ratio comparison..... 74

Table 4: Portfolio A and B loss distribution..... 86

Table 5: Comparison between downsides and distorted risk measure..... 86

Table 7: Sharpe Ratio of Investment A and B.....96

Table 8: SR and GSR of Investment A.....97

Table 9: SR and GSR of Investment B98

List of Figures

CHAPTER ONE: An Integrated Pricing Model for Defaultable Loans and Bonds

Figure 1: pdf (Beta) subordinated non-investment grade.....20

CHAPTER TWO: A Framework for Risk Measures and Risk Adjusted Performance

Figure 1: The expected value integral representation.....42

Figure 2: Quantile Risk Measure.....50

Figure 3: Shortfall-probabilities under a target return τ for two portfolios A and B with their corresponding return density functions.....54

Figure 4: Relation between the target shortfall $LPM(1;\tau)$ and the shortfall-conditioned mean $\mu_{x \leq \tau}$ for returns $x \leq \tau$ two distributions F and G.....55

Figure 5: Expected Shortfall, Quantile and Tail VaR.....60

Figure 6: The relationship between the *expected* and the *distorted* value of a random variable.....64

Figure 7: The Quantile Distorted Risk Measure.....65

Figure 8: The Tail VaR Distorted Risk Measure66

Figure 9: SR vs GSR Return Probability Distribution Functions of Investment A and B.....96

Figure 10: SR vs GSR Cumulative Distribution Functions of Investment A and B97

CHAPTER THREE: Default probability, capital structure and the cost of economic capital

Figure 1: Debt, equity and leverage behaviour when changing volatility and fixed negative correlation (-0.5) is assumed..... 129

Figure 2: Debt, equity and leverage behaviour when changing correlation and fixed volatility (0.15) is assumed.....125

Figure 3: Debt, equity and leverage behaviour when changing both volatility and correlation..... 125

Figure 4: Cost of equity (risk capital) and correlation relationship.....126

Figure 5: The put value as a function of both volatility and correlation with fixed high PD (0.5%).....129

Figure 6: The put value as a function of both volatility and correlation with fixed low PD (0.01%)..... 129

Figure 7: Cost of equity as a function of correlation and PD.....140

Introduction

This dissertation consists of three autonomous essays, discussing the following topics:

1. the pricing of defaultable bonds, loans and plain vanilla credit derivatives,
2. the use of risk-adjusted performance measurement, for optimal portfolio management in the banking, asset management and insurance industries
3. return on economic capital as a measure of value created by the holding of bank assets and the operation of bank business units..

Even if, at a first glance, these three essays appear to be independent developed, it is possible to show how they belong to a single general framework: the analysis of the measurement and management of financial (market and credit) risk. This dissertation has the aim to shed some light on three aspects of this process: the first two being the valuation of the risky debt and the analysis of the risk adjusted performance of a portfolio (such as credit assets) exhibiting non-normal returns; the third the analysis of bank portfolio choice taking account of the conflicting interests of bond-holders and share-holders.

The link between the first two essays can be explained by considering that to build a risk adjusted performance measure one needs to estimate the portfolio value distribution across time. To this aim the first chapter provide a tool to evaluate a portfolio of credit asset in different point in time by simultaneously taking in consideration the interaction of market (represented by the interest rate risk) and credit risk (the risk drivers being the transition risk and the recovery rate by the seniority of debt risk) . From the difference between the portfolio value in two point in time it is possible to estimate the portfolio loss (or eventually) gain distribution. Through the analysis of the probability density function (pdf) of the loss distribution it is possible to estimate the risk measures (downside risk measures, in the banking and asset management industry, “distorted” risk measures, in the insurance industry) presented in the first part of the second chapter¹.

The analysis of the most important features a risk measure should possess is completed through the incorporation of the recent new developments in actuarial science, and applied to the insurance industry. These are discussed in the second part of the second essay. In fact, actuaries are the first claiming that risk measures should go beyond the coherence principles first introduced by Artzner and al. (1997). Nowadays the coherence framework for risk measures represents the best practice, both from an academic and professional perspective, in the asset management and banking industry. Actuaries (Wang 1997) claim that, to be successful in pro-active risk management one needs to carefully consider all the information

¹ In this work only the ex-ante perspective is considered. It is out of the scope of the dissertation to analyse the ex-post performance (performance evaluation)

incorporated in the whole loss distribution, not just the tail; this is because stochastic dominance matters. In this dissertation a risk adjusted performance that can be used in the banking and asset management industry on top of the newly built coherent and stochastic dominance compliant “distorted” risk measures is introduced and discussed with the aim of providing superior information for pro-active portfolio management and optimal economic capital analysis.

Another important motivation for the second chapter is that through the use of distorted risk adjusted performance measurement the risk or asset manager can efficiently select an optimal portfolio whose financial asset show non-normality and fat tail (like the credit assets). It is therefore claimed that even if the underlying distribution shows this type of shape it is possible to provide, from a theoretical perspective, superior information because avoiding to include in the set of feasible portfolios also the stochastically dominated ones.

Summarising, to optimise the risk-return-value profile of the portfolio at the hand risk managers need:

- to evaluate the financial assets across time
- to estimate the related pdf loss distribution
- to adopt risk adjusted performance measure (RAPM) that goes beyond coherence thus controlling for non-normality and fat tails of the loss distribution at the hand.

The last aspect of the financial risk management process, analysed in the third chapter, analyses the determination of the optimal bank target credit rating which allows to maximise the target cost of economic capital by considering, simultaneously, the response to two conflicting objectives: the shareholder and the bondholder perspective. The first two chapters show some “tools” that risk managers have to manage to extrapolate win-win strategies for the conflicting objectives of both shareholders and bondholders. In fact, bondholders main concern is that the bank may default thus loosing (fully or in part) the provided financial resources. Similar concern characterizes the regulators whose main objective is to ensure the stability of the financial market and the (ongoing) solvency of the financial institutions. It is therefore obvious that regulators in all financial industries focus their attention on the downside risk and therefore they pay attention to the tail risk incorporated in the pdf loss distribution. Regulators, consequently, to cope with the problem of estimating capital requirement to assure the ongoing solvency of the financial institutions suggest to use downside measures of risk which emphasizes the capability of financial institutions to remain solvent even in the presence of critical (stress) financial conditions.

On the other hand shareholders and top management are interested in the maximisation of the value of the financial institution (or, stated alternatively, the bank portfolio total return). This means that the shareholders consider a crucial indicator of the well being status of the financial institutions the (expected) portfolio values of the bank portfolio in the future. They are ultimately interested in measuring how the financial institutions are capable to generate superior risk adjusted returns and therefore their focus it is not the tail of the loss distribution but the mean of the portfolio value distribution across time. They know that the bank must remain solvent across time, but shareholders consider this a constraint and not an objective. They want to be successful by achieving the optimal risk adjusted return on the capital of the bank.

The third area in which this dissertation seeks to make a contribution is in the analysis of the relationship between the risk capital (the capital that a bank has to have to remain solvent) and the economic capital (the capital a bank needs to have to reward the shareholders for the risk being taken). This distinction is crucial to identify how much capital a bank must have to avoid default and, at the same time, to be competitive in the financial market. In the third chapter is shown how to select investment alternatives and measure the contribution in risk adjusted value terms. It is claimed that, to this aim, a bank has to make use of a specific risk adjusted performance, namely the cost of economic capital, that is able simultaneously to provide information to:

1. reach the target credit rating (default probability) the bank is willing to achieve to be competitive in the financial market thus defining the amount of risk capital to have
2. define how much leverage (debt to economic capital ratio) a bank has to use to maximise the shareholder value thus providing a superior risk premium to the shareholders.

This explains how the third chapter is linked with the first and the second one. The third chapter consider an equilibrium framework and therefore the results provided should be interpreted accordingly. In another joint research with A. Milne “When does RAROC measure shareholder value? Theoretical diagnosis and practical prescription” this framework is extended to deal with non-normality and fat tail.

Having now summarised the motivations for this thesis, we can briefly review some specific contributions of the different chapters.

The main contribution of the first chapter is to define a general framework to price risky debt through an integrated reduced pricing model. In the integrated pricing model the pricing of any risky security is shown to reflect the return on a risk-free asset plus a risk margin. The

risk margin, in theory, must compensate the investor for all the risks assumed which are represented by:

- transition and recovery (by seniority of debt) risk
- liquidity risk
- credit exposure risk
- default correlation risk
- collateral risk
- concentration risk
- market and credit correlation risk.

In practice it is virtually impossible simultaneously consider the joint stochastic evolution of these risk sources. In the integrated pricing model only the risk coming from interest rates, transition and recovery (by seniority of debt) variations are (jointly) modelled within a no arbitrage framework. In this chapter the most widespread approaches which may be used to jointly model interest rates and credit spreads are discussed, namely reduced form models and structural models (Merton 1974). In particular this chapter discusses the ones directly modelling credit spread components (transition and recovery risk). The model described in this chapter can be considered as a generalization of the Das & Tufano (DT, 1996) model, which is an extension of the Jarrow-Lando-Turnbull (JLT,1997) model, and uses credit ratings to characterize the transition risk. Unlike the JLT model, the DT model makes the recovery rate in the event of default stochastic, and provides a two-factor decomposition of credit spreads. In this essay this approach is generalised by considering different set-ups for different seniority classes of debt. Therefore, their main features are those of being mark to market (in contrast to the default mode paradigm) and that the spread term structure by rating class is adjusted for a spread term structure which is contingent on the seniority of debt within an arbitrage free framework.

In the second chapter, for the sake of portfolio optimization and sound risk management, it is shown that it is essential for a risk measure to properly reflect the risk differentials in alternative strategies or portfolios. The main contribution of this essay is to show how a risk measure, although being coherent, ignores useful information in a large part of the loss distribution, and consequently lacks incentive for mitigating losses below quantile (VaR) risk measure. Therefore a good risk measures for pro-active portfolio management must go beyond coherence. When trying to asses the main appealing features a RAPM should show, it is also important to recognize that do exist many (and conflicting) objectives, the financial environment is uncertain, dynamic and complex, and the underlying financial instruments

(their returns) may show asymmetry, fat tails and liquidity problems besides the -traditional- financial and operational risks.

In this set up, a further contribution is based on the intuition that the analyst has to clearly identify what are the answers the performance measure should correctly give and not worry about possible paradoxical cases that are out of the set of scenarios considered by the decision maker. This means that risk measures that have to answer the question of the right capital requirement a financial institution should have not necessarily must have the same features of a RAPM used for proactive risk management. Strictly speaking, in the first case there is no reason to know all the information incorporated in the whole loss distribution because for the aim of the regulators only the censored tail risk matters. But clearly from a business perspective the type of information the analyst have to ask to the portfolio value distribution is somewhat different and therefore the set of desirable properties change thus making the stochastic dominance principles matter.

In order to be conscious of the limitations and shortcomings of the risk measures currently in use in the best practice in several financial applications in this chapter the understanding of the theoretical relationships that exist between the stochastic dominance principles and coherence axioms will be discussed with the final aim to provide guidelines for the risk adjusted performance measure correct use in this well defined context.

The main contributions of the third chapter is to provide the combined first analysis of the impact of default probability, capital structure, limited liability, deposit insurance, franchise value and taxes on Risk Adjusted Return on Capital (RAROC or cost of economic capital) type risk measures. Previous work of Crouhy and al (1999) has addressed part of this problem but does not provide a complete analysis. The suggested model also adopts a rigorous equilibrium perspective by enforcing the pricing of banking assets using the Capital Asset Pricing Model (CAPM). This alters significantly some of the results reported by Crouhy and al. (1999), in particular showing that their results greatly overstate the impact of volatility on the required “hurdle” for return on economic capital. Finally, this framework also clarifies the relationship amongst different measures of capital, namely the market value of the equity, the risk capital and the economic capital.

Please note that the first chapter of this work is forthcoming in European Journal of Operational Research EJOR, Volume 163, Issue 16 May 2005, pages 65-82 (M. Onorato and E. Altman). The third chapter is jointly written with Alistair Milne.

Chapter 1

An Integrated Pricing Model for Defaultable Loans and Bonds¹

"Maximising expected return is fine ... if you are risk-neutral. Nobody is.

Maximising expected utility is fine ... if you understand it.

But anyway... will I lose money?"

Zagst (2002)

¹ We thank Giovanni Barone Adesi, Winfried Hallerbach, Anthony Saunders, Jaap Spronk, Rangarajan Sundaram, Constantine Thanassoulas, three anonymous referees, and seminar participants at Italian Banking Association, Milan, for helpful discussion of this study. Earlier versions of this chapter were presented at the Annual Meetings of both Financial Management Association (Paris 2001) and European Financial Management Association (London 2002) and at the Euro Working Group on Financial Modelling Meeting (Capri 2002). Please note that this chapter is a joint research paper with E. Altman, forthcoming in European Journal of Operational Research, Volume 163, Issue 16 May 2005, pp 65-82.

1. Introduction

The 1988 Basle Capital Accord, which created a minimum risk-based capital adequacy requirement for banks, marked a major step forward in introducing risk differentiation into the regulatory framework but did not represent the “optimum” solution.

The 1996 amendment to the Capital Accord aimed to correct some of the issues concerning the original accord, but it did not change the section regarding credit risk. As a consequence, regulatory capital was still not an appropriate basis for the capital allocation process. Financial institutions started developing their own internal credit risk systems using economic capital.

In April 1999, regulators proposed a document, *Credit Risk modelling: current practice and applications*, which aimed at assessing the potential applications and limits of credit risk models² for supervisory and regulatory purposes, in sight of the foreseeable amendment to the Capital Accord. The Basle Committee proposed the amendment in June 1999: A new capital adequacy framework. This event marked a breakthrough as regards the credit risk concept for capital adequacy. The committee’s proposal dramatically modified the standard approach for capital requirements imposed by the 1988 Accord. A more realistic approach was introduced based on internal rating systems. Moreover, it presented the reclassification of securities taking into account credit risk in all its aspects: default, migration, recovery rates, credit spreads, aggregation and concentration risk.

Over the last decade risk managers, regulators, academics and software vendors are devoted to define a sound credit (and market) risk measurement and management system. The main objective of this study is to define a general framework to price risky debt. The pricing of any risky security must reflect the return on a risk-free asset plus a risk margin. The risk margin must compensate the investor for the risk assumed which is represented by: the transition and recovery (by seniority of debt) risk, liquidity risk, credit exposure risk, default correlation risk, collateral risk and concentration risk. Moreover, in an integrated framework, another source of risk has to be considered, namely: market and credit correlation risk³.

² The document issued by the Basle Committee analyses in particular four widespread credit risk models: Creditmetrics™ (JP Morgan 1997); KMW Portfolio Manager™ (Kealhofer 1998), CreditRisk+™ (Credit Suisse First Boston 1997), Credit Portfolio View™ (Wilson 1997 and 1998). As the recent literature shows Koyluoglu and Hickman (1998), these models fit within a single generalised framework..

³ Consequently, all the correlations among the market factors (which drive the price of securities) and credit risk factors-also called *background factors*- which affect the creditworthiness of obligors in the portfolio need to be identified and modelled. For an illustrative description of the background factors within the credit risk model framework refer to Koyluoglu and Hickman (1998) and Wilson (1997 and 1998).

There are several approaches, which may be used to jointly model interest rates and credit spreads. In this paper reduced⁴ form models and in particular the ones directly modelling credit spread components (transition and recovery risk) are considered.

Our approach is a generalization of the Das & Tufano (DT) (1996) model, which is an extension of the Jarrow-Lando-Turnbull (JLT) (1997) model, and uses credit ratings to characterize the transition risk. Unlike the JLT model, the DT model makes the recovery rate in the event of default stochastic, and provides a two-factor decomposition of credit spreads. In this paper we generalise this approach by considering different set-ups for different seniority classes of debt. Therefore, its main features are those of being mark to market (MTM)⁵ and that the spread term structure by rating class (STSRC) is adjusted for a spread term structure which is contingent on the seniority of debt (SSD) within an arbitrage free framework.

Summarising, in our integrated pricing model we take into account the risk coming from both interest rates variations and credit event verifications⁶.

The remainder of this paper is organised as follows. Section 2 defines default and transition risk and it is aimed at illustrating both the most common methods to estimate them and the inherent empirical evidence. Section 3 analyses the same issues for the recovery rate by seniority of debt risk. Section 4 is the core of the paper. In this section the theoretical integrated pricing model is illustrated. The bulk of the suggested approach relies on the modelling of both the stochastic spread term structure by rating class and the spread contingent to the seniority of debt within a unified arbitrage-free framework. Section 5 illustrates possible applications of the integrated pricing model. Section 6 concludes.

2. Default and Transition Risk

Credit pricing and risk models attempt to measure credit losses. Losses are the consequence of the firm's financial position and asset quality deterioration, which then leads to the degradation of its creditworthiness (credit migration). The determination of the creditworthiness of the issuer is difficult as it is driven by many factors such as general economic conditions, industry

⁴ Reduced form models are so called to contrast them to structural models (Merton 1974). The difference between structural and reduced form model is outlined in section 2 and 4. For a complete review of the inherent literature refer to the work of Acharya, Das and Sundaram (2002).

⁵ "In contrast to the default mode paradigm, within the mark to market (or, to be more accurate, mark to model) paradigm a credit loss can arise in response to deterioration in asset's credit quality shorts of default". For a detailed discussion on the topic please refer to the work of Basle Committee, pag.21.

⁶ If a credit event occurs, the credit quality of the issuer changes. Examples of credit events are given in section 2.

trends and specific issuer factors like the issuer's financial wealth, leverage, market value, equity value, asset value, capital structure and less tangible things such as reputation and management skills.

The probability of a customer migrating from its current risk-rating category to any other category, within a pre-defined time horizon, is frequently expressed in terms of a rating transition matrix⁷. Rating migration probabilities are therefore collected in the transition (migration) matrices and describe the probability of migrating from any given credit rating to another one. Moreover, estimates of transition probabilities often suffer from small samples, either in the number of rated firms or in the number of events; in particular this happens when considering transition towards the most "distant" rating classes. This often results in biased estimates of these type of transition probabilities that has led the Basle Committee on Banking Supervision to impose a lower minimum probability of 0.0003 for rare events. In our analysis we will cluster obligors into obligor rating classes. In this matrix, the last column represents the worst internal grade (the worst state) which corresponds to the the default state⁸.

Let us assume to deal with K rating classes (the K th being the default state), then, the transition matrix is the collection of one-step transition probabilities of migrating from any class- i to any class- j at the given time- m , including the probabilities of remaining in the same class (corresponding to the off-diagonal values). The statistical, or actual, probabilities matrix can be represented as:

$$\bar{Q}(m) = \{\bar{q}_{i,j}(m)\} = \begin{bmatrix} \bar{q}_{11}(m) & \dots & \dots & \bar{q}_{1K}(m) \\ \bar{q}_{21}(m) & \bar{q}_{22}(m) & \dots & \bar{q}_{2K}(m) \\ \dots & \dots & \dots & \dots \\ \bar{q}_{K,1}(m) & \bar{q}_{K,2}(m) & \dots & \bar{q}_{K,K}(m) \end{bmatrix} = \begin{bmatrix} \bar{q}_{11}(m) & \dots & \dots & \bar{q}_{1K}(m) \\ \bar{q}_{21}(m) & \bar{q}_{22}(m) & \dots & \bar{q}_{2K}(m) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

Once the default state is reached, no other rating classes are possible to exist at the next time step, therefore all the transition probabilities are defined to be null, with the exception of the probability of staying in the K -state i.e. default. As a consequence of the definition of transition probability, another relevant property of the migration matrix is that the sum over the elements of the same row must equal one.

⁷ See Altman (1998) for a complete discussion on the topic.

⁸ In a two state default process, within the considered time horizon, there are only two possible events: "no default" and "default".

$$\sum_{j=1}^K \bar{q}_{i,j}(m) \equiv 1; \quad \forall i$$

In the integrated model, the statistical transition probabilities are derived from empirical data by using the mortality approach briefly described in the next section.. The process determining customer defaults or rating migrations can be modelled through two approaches: actuarial based methods and equity based methods.

2.1 Actuarial based method

The basic actuarial approach uses historical data on the default rates of borrowers to predict the expected default rates for similar customers. The actuarial (also called empirical) model is based on the estimation of statistical (or actual) transition probabilities⁹. This method uses historical data to evaluate the migration probabilities. In this model the inputs are represented by empirical data and the output will be the statistical estimation concerning these data. One of the most important criticisms to the empirical approach is the apparently static nature of the resulting average historical probabilities. In reality, actual transition and default probabilities are very dynamic and can vary quite substantially over one year, depending on general economic conditions and business cycles¹⁰.

In our integrated pricing model we will estimate the actual transition probability by using the Altman (1998) mortality rate approach. This method determines the (expected) default rate, using an empirical method. An important element that needs to be observed is the aging effect¹¹, that is, the time between instruments' issue up to valuation time. This approach implies lower default probabilities in the first year than in the next years. The question is not whether a bond or another credit derivative is going to default or not but when it is going to default and what will be the likely recovery, given its original rating and its original seniority. Credit instruments are classified for issue time and rating classes. After this classification it is possible to calculate the default probability. In order to do this, it is necessary to calculate the marginal mortality rate and the cumulative mortality rate.

The marginal mortality rate (MMR) is the probability that a credit instruments defaults over the first year, over the second and so on. The MMR can be expressed both in term of number and

⁹ Here we refer to actual (statistical or empirical) transition probability in contrast to risk-neutral probability

¹⁰ A real dilemma concerns the private companies that are neither rated by the agencies nor publicly traded. In fact a substantial proportion of these portfolios do not have very clear benchmark for estimating default and transition probabilities

¹¹ Altman's method (1998) is different from other methods determining the "aging effect". In fact Moody's and S&P use static pools (including all credit instruments), while Altman makes a distinction among instruments according to the issue date. The (actual) transition probability matrix in the integrated pricing model can be easily inferred from the migration matrix, which is estimated using the Altman mortality rate approach. For a more detailed illustration of the mortality (default) rate by rating and by age approach please refer to Altman (1998)

value. In the last case MMR is equal to the ratio between the total value of the corporate bonds included in a specific rating class defaulting over the planning horizon and the total value of corporate bonds included in the same rating class, at the beginning of the time horizon.

$$MMR_i = \frac{\text{Total value of defaulted bonds in the } i_{th} \text{ year}}{\text{Total value of bonds issued at the beginning of the } i_{th} \text{ year}}$$

where $i = 1 \dots NN = \text{number of years}$ ¹²

Consequently the survival rate (SR_i) is equal to $SR_i = 1 - MMR_i$. One can measure the cumulative mortality rate (CMR_T) during a specific time period, subtracting the product of the surviving populations over the previous time, that is $CMR_T = 1 - \prod_{i=1}^T SR_i$. Altman derives the migration probabilities for each rating class from CMR_T , which represents the default probability.

2.2 The equity based method

The *equity-based* approach, often associated with the Merton model, is mainly used for estimating the Expected Default Frequencies (EDF)¹³ of large and middle-market business customers, and is often used to crosscheck estimates generated by actuarial-based methods. This technique uses publicly available information on a firm's liabilities, the historical and current market value of its equity and the historical volatility of its equity to estimate the level and the rate of change of the volatility (at an annual rate) of the economic value of the firm's assets. There are at least three practical limitations to implement the option (Merton model) approach:

1. It is necessary to know the market value of firm's asset. This is rarely possible as the typical firm has numerous complex outstanding debt contracts traded on an infrequent basis.
2. It is necessary to estimate the return volatility of the firm's asset. Since the market prices cannot be observed for the firm's assets, the rate of return cannot be measured and volatilities cannot be computed.

¹² If, for example, the par value outstanding of high-yield debt in 1997 was 335.400 (\$ millions) and the par value defaults was 4.200 (\$ millions), the MMR (or alternatively the default rates) was 1.252%

¹³ Expected default probabilities can be inferred from the option models under the assumption that default occurs when the value of a firm's assets falls below its liabilities. See Crosbie (1998) for a detailed description of how the EDF are estimated within the KMV model.

3. It is necessary to simultaneously price all the different types of liabilities senior to the corporate debt under consideration. Most corporations have complex liabilities structures¹⁴.

Summarizing, the key ingredient of credit asset pricing and risk modelling is the default (or, in a multi-state framework, transition) risk, which is the uncertainty underlying a firm's ability to service its debts and obligations. Prior to default, there is no way to discriminate unambiguously between firms that will default and those that will not. At best we can only make probabilistic assessments about the default possibility. In practice, we use transition probabilities basically for two main reasons:

1. In the trading book, to price credit sensitive instruments adjusting it through the risk neutral transition probability
2. In the banking book, to measure the credit risk of portfolio losses of loans

3. Recovery by seniority of debt risk

The default is one of the main types of credit events which determines loss amount occurring once a credit has defaulted. This credit loss, also called loss given default (LGD), is defined as the difference between the (current) bank's credit exposure and the present value of the future net recoveries (cash payments from the borrower less workout expenses). Therefore the recovery rate (RR) is equal to the ratio between 1-LGD and the initial exposure¹⁵. LGD depends on a limited set of variables characterising the structure of a particular credit facility¹⁶. These variables may include the type of product (e.g. business loan or credit card loan), its seniority, collateral and country of origination.¹⁷

Reduced form model assume either a constant (Litterman and Iben, JLT, for example) or a stochastic recovery rate (Duffie and Singleton and DT for example). These models assume zero correlation among the LGDs of different borrowers, and hence no systematic risk is due to

¹⁴ For an interesting and detailed analysis of the limitations of the equity approach see Jarrow and Turnbull (2000)

¹⁵ The estimation of LGD depends on the availability of historical loss data that may be retrieved by the following possible sources: bank's own historical LGDs records, samples by risk segment; trade association and publicly available regulatory reports; consultants' proprietary data on client LGDs, and published rating agency data on the historical LGDs of corporate bonds.

¹⁶ For portfolios characterised by distributions of exposure sizes that are highly skewed, the assumption that LGDs are known with certainty may tend to bias downwards the estimated tail of the PDF of credit losses

¹⁷ In the Credit Risk+™ model the LGD is treated as a deterministic variable while in the other structural models is treated as a random variable. In these models the LGD probability distribution is assumed to take the form of a beta distribution because this result in a type of distribution whose shape is typically skewed to the right as shown in the empirical works of Altman and Kishore (1998), Carty and Lieberman (1996), Duffie and Singleton (1996), Castle, Keisman and Yang (2000)

LGD volatility¹⁸. Moreover, the empirical evidence shows that the recovery rates are both state¹⁹ and structure²⁰ dependent. In particular, our analysis, based on the last three decades default and recovery data on US corporate bonds shows that the recovery rate changes depend on the seniority of debt. In fact, comparing senior secured and unsecured bonds one can see that the recovery distribution for the latter is more spread out and has a longer lower tail (see table 1)²¹.

Recovery Rates by Seniority and Original Bond Rating, 1971-2001					
Seniority	Number of Observations	Recovery Rate			
		Average Price	Weighted Price	Median Price	Standard Deviation
Senior Secured					
Investment Grade	35	\$62.00	\$66.00	\$56.88	\$19.70
Non-Investment Grade	113	38.65	32.89	30.00	29.46
Senior Unsecured					
Investment Grade	159	\$53.14	\$55.88	\$50.00	\$26.14
Non-Investment Grade	275	33.16	30.17	31.00	25.28
Senior Subordinate					
Investment Grade	10	\$39.54	\$42.04	\$27.31	\$24.23
Non-Investment Grade	283	33.31	29.62	28.00	24.84
Subordinated					
Investment Grade	10	\$35.64	\$23.55	\$35.69	\$32.05
Non-Investment Grade	206	31.73	28.87	28.00	22.06

Table 1 Recovery rates by seniority of debt (1971-2001)

¹⁸ Furthermore, they assume independence among LGDs associated with the same borrower. The assumption that LGDs between borrowers are mutually independent may represent a serious shortcoming when the bank has significant industry concentrations of credits. Furthermore, the independence assumption is clearly false with respect to LGDs associated with similar (or equally ranked) facilities to the same borrower. The assumption of default intensities independence may contribute to an understatement of losses to the extent that LGDs associated with borrowers in a particular industry may increase when the industry as a whole is under stress.

¹⁹ Some evidence consistent with the state-dependence of recovery rates is presented in the analysis, based on recovery rates, compiled by Moody's for the period 1974 through 1996 (Carty and Lieberman, 1996). However, even for senior secured bonds, there was substantial variation in the actual recovery rates. Although these data are also consistent with cross-sectional variation in recovery, that is not associated with stochastic variation in time of expected recovery, Moody's recovery data also exhibit a pronounced cyclical component. There is equally strong evidence that the level of LGD of corporate bonds vary with the business cycle (as is seen, for example, in Moody's data). Speculative-grade default rates tend to be higher during recessions, when interest rates and recovery rates are typically below their long-run means.

²⁰ See Castle, Keisman and Yang (2000)

²¹ Also Duffie and Singleton found similar results.

The analysis also ranks the results in investment grade and non-investment grade.

When evaluating an instrument at the first steps of its life the type of guarantee rate on the underlying security is an extremely relevant characteristic, which – according to historical data - implies that the credit is subject to a global lower risk. This is shown in Table 1, where the rating class, at the time of issuance, non investment grade in particular, is not influencing the average values as much as the seniority class does (see the senior secured and senior unsecured investment grade case).

The most relevant issue is that the dominant factor influencing the evaluation of the security is the composition of both the recovery rate and default probability. Actually, low default rates do not assure that in case of default the recovery rate is low as well; on the other hand high recovery rate is not a credit low-risk index alone, since the security might be highly defaultable, implying the elevated investment risk.

In the integrated model we will estimate spread term structure by rating class, to explicitly consider the credit rating (risk) transition and will adjust them by means of spread term structure of recovery rate by seniority of debt, both in an arbitrage free framework, in order to get a risk-neutral price of the financial instruments.

4. The theoretical integrated pricing model

There are two main approaches to pricing credit risky instruments: the structural²² and the reduced form approach. It is argued that the structural approaches are of limited value when applied to price interest rate and credit sensitive instruments and, consequently, in measuring and managing market and credit risk in an integrated fashion. Rosen (2002) shows that, since the main focus of structural model is the measure of the counterparty exposure risk, they assume deterministic market risk factors, such as interest rate risk. In contrast to structural models, which assume a specific microeconomic process generating customers' default and rating migrations, reduced-form models attempt to directly describe the arbitrage free evolution of risky debt values without reference to an underlying firm-value process. Acharya, Das and

²² In fact, the structural approach assumes some explicit microeconomic model for the process that determines defaults or rating migrations of any single customers. A customer might be assumed to default if the underlying value of its assets falls below some specified threshold, such as the level of the customers' liabilities. The change in the value of a customer's assets in relation to various thresholds is often assumed to determine the change in its risk rating over the planning horizon. Structural approach models are Merton type models.

Sundaram (2002)²³ show how this class of model have resulted in successful conjoint implementations of term-structure models with default models. The objective pursued within the suggested integrated pricing model is that of deriving a general framework for pricing risky debt, both plain vanilla (as for example corporate bond) and (credit) derivative. Present values of all cash flows are calculated by using both stochastic interest rate term structure (market risk) and stochastic credit spread term structure (credit risk). This last term can be decomposed in the following risk sources:

- 1) stochastic recovery rates by seniority,
- 2) correlation between interest rate term structure and stochastic recovery rates (correlation of market and credit risk)
- 3) multi state transition probability at the m -th time step for the M -period process.

In this set up the proposed integrated pricing model may be considered a multi-period mark to model framework.

As for all reduced-form models, also in our integrated model we start modelling the risk free term structure by considering an underlying process for the evolution of risk-less rates. The objective is to build a lattice of risky rates on top of the risk-less rate process in an arbitrage-free manner by directly modelling credit spread components (transition and recovery risk).

We generalise the Das & Tufano (DT) model, where the spread term structure by rating class is modelled through three main components: risk neutral probability matrix, stochastic recovery rate and its correlation with interest rates. In the DT model the first component is aimed at estimating the transition risk; the second one, the recovery risk; the last one, the correlation between market and credit risk. In the integrated model different set-ups for different seniority classes are introduced within an arbitrage free framework.

As the empirical evidence shows (see section 2 and 3) the mean recovery is mainly contingent on the seniority of debt rather than on the rating class alone (investment grade vs. non investment grade in our analysis) as it is almost invariably assumed in all reduced model. The major contribution of this work it is to correct each STSRC through a spread contingent on the seniority of debt (SSD) within a unique arbitrage free framework. As a result, this model allows

²³ Reduced-form models may differ depending on the procedure that is used, the input information required, the use of ratings-matrix and the recovery assumptions. As pointed out by Das and Sundaram (2000)^[21] There are three commonly used assumptions concerning recovery rates in the event of default: recovery of par, where the recovery amount is specified as a fraction of par value due at maturity; recovery of treasury, where recovery amount is specified as a fraction of value of a default-free bond with the same maturity; recovery of market value, where the recovery amount is specified as a fraction of the immediately-preceding market value.

more variability in the spreads of risky debt. Moreover, by choosing different recovery rate processes for instruments within the same credit rating class, it allows variability of spreads to be instrument specific rather than rating class specific.

4.1 *The stochastic interest rate term structure model*

In the integrated model an interest term structure is assigned at each rating class i (where $0 \leq i \leq K$, if we consider K rating classes). The i -th interest rate f_i at which cash flows are to be discounted is composed of forward risk-free interest rate plus a (forward) spread s associated to the same rating class as shown below:

$$f_i(t) = \text{forward curve for rating class-}i = \text{forward risk free}(t) + \text{spread-}i \quad (4.1.1)$$

In this context, the risk-free forward interest rate (stochastic) process, can be modelled by using any interest rate term structure model like, for example, the Heath-Jarrow-Morton (1992) or the Black-Derman-Toy [1990] model. It is not the purpose of this paper to enter details of the risk neutral set up model formulation for the evolution of the interest rate free term structure, for which specialised literature may be addressed. More relevant to the present paper purposes is to illustrate how the spread is modelled for which the following paragraphs are devoted to.

4.2 *The stochastic spread term structure model*

Recovery rates, risk of default and the seniority type are relevant parameters for assessing credit risk. Therefore, in the integrated model, the spread is decomposed in its two main determinants: recovery rates and default (transition) risk. In order to price the credit spread component of the interest rate term structure, both recovery rate and default variables need to be modelled. Let d be the default rate (i.e., the rate at which default occurs). This rate may be either constant, or function of time-to-maturity of the security or of any other factor in the economy. The recovery rate represents the fraction of the face value of the security that is recovered in case of default and is denoted ϕ ²⁴.

Considering the influence of recovery and default rates on credit instruments, it is possible to analyse a first simple relationship between these parameters and interest rate spreads. Let r be the one period risk-less rate of interest, then the risk-neutral value of a credit risky bond, maturing in a single period from now, must be equal to the discounted value of the expected cash flows in the future:

²⁴ It results by construction $0 \leq \phi \leq 1$.

$$B = \frac{d\phi + (1-d)}{1+r} \quad (4.2.1)$$

where the parameters d and ϕ have been set to their risk-neutral values. On the other hand the price of the risky bond B off the spread curve is given by:

$$B = \frac{1}{1+r+s} \quad (4.2.2)$$

By equating the right hand sides of Eq.(4.2.1) and Eq.(4.2.2) the required relationship between the spread s , which is the observed market spread for the generic security I, and the determinants of the spread may be derived. Solving by s we obtain:

$$s = \frac{d(1-\phi)(1+r)}{1-d(1-\phi)} \quad (4.2.3)$$

In general, the actuarial estimation of the default rate is different from its risk neutral value d , because the way through which the actuarial value is estimated is independent from the market price of that security. If, recalling eq. 2.1, the actuarial default rate is $q_{i,k}$, we have:

$$s_{act} = \frac{q_{i,k}(1-\phi)(1+r)}{1-q_{i,k}(1-\phi)} \quad (4.2.4)$$

where s_{act} differs from s . To calibrate the statistical value of the default rate one can use spread market data. Given s , it is possible to render $q_{i,k}$ risk neutral by summing to $q_{i,k}$ the adjustment factor π .

$$s = \frac{(q_{i,k} + \pi)(1-\phi)(1+r)}{1-(q_{i,k} + \pi)(1-\phi)} \quad (4.2.5)$$

Consequently $d = q_{i,k} + \pi$.

This approach allows coupling the model of the stochastic process for the interest rate term-structure to the market data (i.e., theoretical and empirical data).

From Eq.(4.2.3) it is possible to see that:

a) the spread increases proportionally to the default rate d increase; this has a financial implication: as the default rate d increases the possibility of getting values far apart from the expected average value is higher. On the contrary, in the limiting case of default risk

approaching zero ($d \rightarrow 0$; for $d=0$ the recovery rate loses its meaning) the spread tends to zero ($s \rightarrow 0$), allowing the certain value equal to the average

b) the spread decreases proportionally to the recovery rate ϕ increase, which means that - in case of default - the higher is the chance of getting back the invested amount, the more limited fluctuations from the average price are got; in other words high recovery rates assure low credit risk. In the limiting case, approaching total recovery ($\phi \rightarrow 1$) the spread still tends to zero ($s \rightarrow 0$), in the ideal limiting case $s=0$ representing the evolution of a risk-less process

c) when the default rate tends to one ($d \rightarrow 1$) and the recovery rate tends to zero ($\phi \rightarrow 0$) the spread tends asymptotically to become infinite.

Of course limiting cases are never reached but their study helps visualising the trend of the functional dependence of the spread from the default risk and recovery rate. In fact, as pointed out by Das (2001), Eq. (4.2.3) expresses the spread as a function of the composite variable $d(1-\phi)$, for this reason the above formulation does not allow expressing the spread as a function of default risk and recovery rate independently. Therefore a more elaborated interest rate spread modelling is needed. Considering, for example, the HJM model, it is possible to observe that its structure allows for the required effective two-factor decomposition of credit spreads. Under the risk-neutral measure, the expected risky cash flows discounted at risk-less rates must be equal to the value of expected risk-less cash flows discounted at risky discount rates:

$$\sum_{m=\frac{t}{\Delta}+1}^{\frac{T}{\Delta}} \left[E \left[\exp \left[-\Delta \sum_{j=\frac{t}{\Delta}}^{m-1} f_i(t, j\Delta) \right] * C_d(m) \right] \right] = \exp \left[-\Delta \sum_{j=\frac{t}{\Delta}}^{\frac{T}{\Delta}-1} (f_i(t, j\Delta) + s_i(j\Delta)) \right] * 1 \quad (4.2.6)$$

where $C_d(m)$ is the expected cash flow of the risky bond in case of default at the time step- m before maturity and 1 is the cash flow in case of non-default.

In order to render $C_d(m)$ in explicit form it is necessary to define the cumulative and one-period default probabilities associated to the rating class of the instrument at any given time, and to consider the recovery rate at the corresponding default time.

Including the default risk, the recovery rate and the credit seniority information in $C_d(m)$, by means of Eq.(4.2.3) it is possible to estimate the determinants of the interest rate term structure spread associated to the rating class I contingent to the seniority of debt.

In order to develop a consistent framework - since for the interest rate term structure model a risk-neutral world is assumed, the actual transition probabilities (estimated by using the mortality approach²⁵ described in section 2) have to be risk-neutral adjusted. After having obtained the risk-neutral set up for the evolution of the term structure of interest rates, the integrated model derives the risk-neutral probabilities of the transition process to default.

Summarising, we will first correlate the interest rate term with one recovery rate structure, then, we will generalise the results by considering s seniority type thus including the spread correction due to the recovery dependence on the seniority.

Following this set up a new stochastic framework for the arbitrage-free pricing of risky debt is depicted. This framework is illustrated through the following three steps:

- 1) construct a one period risk neutral probability matrix for each seniority type
- 2) extend to a multi-period framework through the definition of a cumulative risk-neutral transition matrix which allows the obligor to default at any point in time
- 3) estimate the STSRC contingent to the seniority of debt

4.2.1 Risk-neutral probability transition matrix

One of the key points in which the integrated model departs from other models is in the spread dependence assumption of both the recovery rate on seniority s and of the rating class. In general, in the integrated model, the recovery rate is assumed to follow any “reasonable” distribution. We suggest calibrating the model by using a beta distribution in accordance to the empirical evidence described in the second section. In practice any value for the recovery rate is possible with a non zero probability. The probability density function of the beta distribution, $g_{s,m}$, is given by:

$$g_{s,m}(\phi_{s,m}, \alpha_{s,m}, \beta_{s,m}) = \begin{cases} \frac{\Gamma(\alpha_{s,m} + \beta_{s,m})}{\Gamma(\alpha_{s,m}) \cdot \Gamma(\beta_{s,m})} \cdot \phi_{s,m}^{\alpha_{s,m}-1} \cdot (1 - \phi_{s,m})^{\beta_{s,m}-1} & \text{for } 0 < \phi_{s,m} < 1 \\ 0 & \text{for } \phi_{s,m} < 0 \text{ and } \phi_{s,m} > 1 \end{cases} \quad (4.2.1.1)$$

with $s = 1, \dots, 5$

where $\phi_{s,m}$ represents the fraction of recovery at time m associated to the seniority type s, $g_{s,m}(\cdot)$ represents the probability associated to that recovery rate (belonging to the s seniority

²⁵ The statistical migration matrix is an input in this pricing model. One can also use other approach, like the S&P or Moody's method of estimating the migration matrix

class); the seniority type range is between 1 and 5 because the considered seniority classes are 5: senior secured, senior unsecured, senior subordinated, subordinated, junior subordinated. Moreover Γ stands for gamma distribution and $\alpha_{s,m} \in \beta_{s,m}$ are two generic parameters which depend from both the seniority (s) and the time (m). Equation 4.2.2.1 has the required property that the recovery is bounded by 0 and 1²⁶. If $\mu_{s,m} = 31,73\%$ and $\sigma_{s,m} = 22,06\%$, as for the subordinated non investment grade bond (see table 1), the pdf of the beta distribution is depicted in figure 1.

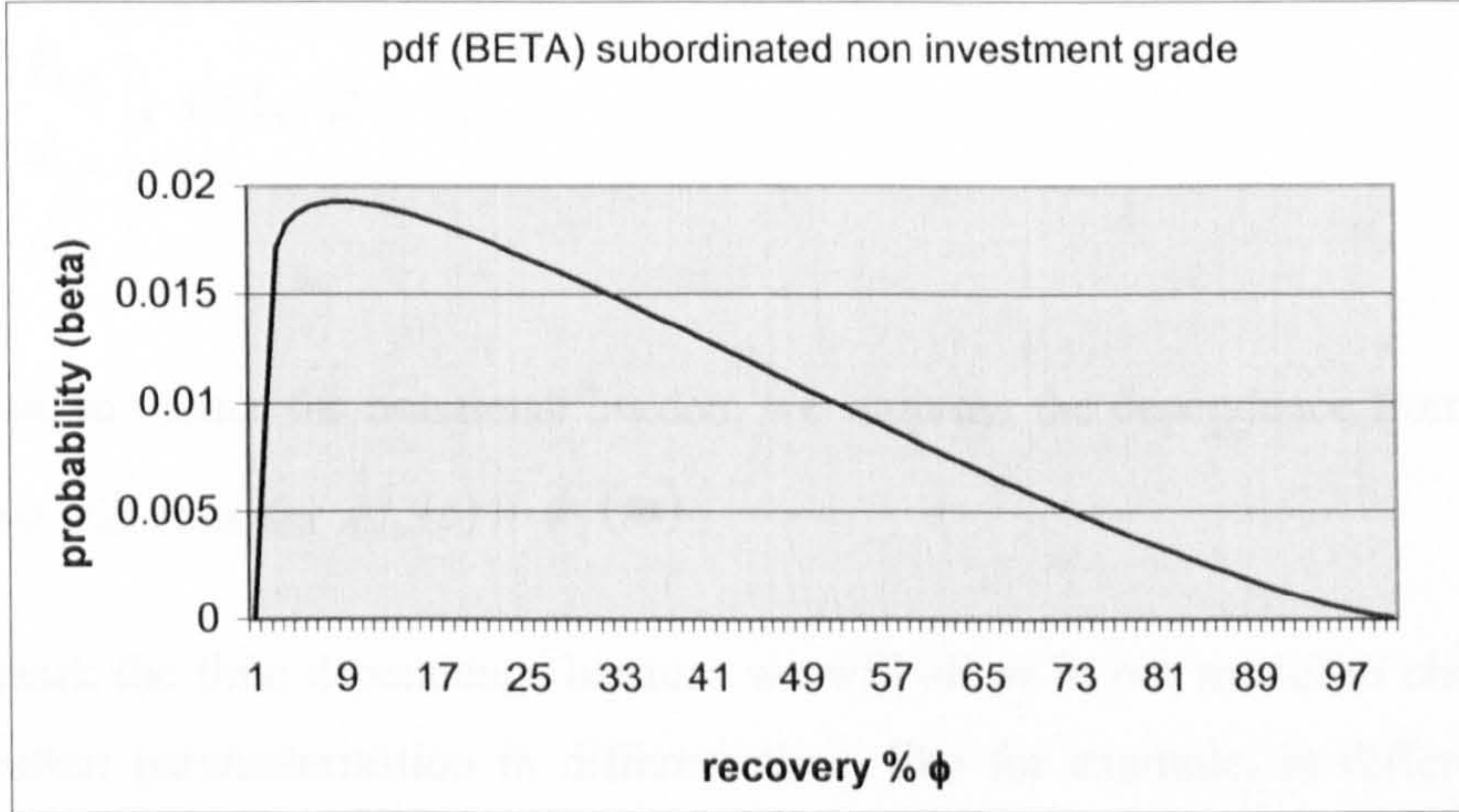


Figure 1 pdf beta distribution with $\mu_{s,m} = 31,73\%$ and $\sigma_{s,m} = 22,06\%$,

The figure illustrates the high degree of randomness of the recovery rates.

The model objective is to develop a risk neutral lattice for pricing risky debt. In order to render the forward interest and recovery rate process tractable numerically, the corresponding state space²⁷ should be discretised consistently through both the (discretised) time structure, m, and through the “shock” vectors v ²⁸, for the risk free term structure process, and z, for the recovery rate process. To implement the model, we make the standard discrete time assumption that v and z are binomial random variables. In particular, we assume that both v and z takes on the value ± 1 with probability 0,5:

²⁶ The parameters can be computed since the mean $\mu_{s,m}$ and variance $\sigma_{s,m}^2$ of a beta distribution are given by:

$$\mu_{s,m} = \frac{\alpha_{s,m}}{\alpha_{s,m} + \beta_{s,m}} \quad \text{and} \quad \sigma_{s,m}^2 = \frac{\alpha_{s,m}\beta_{s,m}}{(\alpha_{s,m} + \beta_{s,m})^2(\alpha_{s,m} + \beta_{s,m} + 1)}$$

where $\mu_{s,m}$ and $\sigma_{s,m}$ are the mean and standard deviation of the actual empirical distribution of credit recovery belonging to seniority debt type-s.

²⁷ The state-space is defined as the ensemble of all possible states related to the stochastic process.

²⁸ If, for example, the HJM model is used to build the risk neutral set up for the estimation of the risk free term structure, v represent the random variable of the underlying stochastic process, i.e.: $f(t+m, T) = f(t, T) + a(t, T)m + \sigma(t, T)v\sqrt{m}$, where a and σ represent respectively both the drift and the volatility of the process. In a discrete time set up, periods are taken to be of length $m > 0$, thus a typical time point, t, has the form lm for integer l.

$$v = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix}; \quad z = \begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} \quad (4.2.1.2)$$

Consequently, “dcretising”, the recovery rate vector, function of each seniority type s at the given time m , becomes:

$$\phi_{s,m}(z) = \begin{bmatrix} +\phi_{s,m} \\ -\phi_{s,m} \\ +\phi_{s,m} \\ -\phi_{s,m} \end{bmatrix}; \quad s = 1, \dots, 5 \quad (4.2.1.3)$$

From now on, to reduce the notational burden, we suppress the dependence from z and in the remainder we will consider $\phi_{s,m}(z) = \phi_s(m)$.

We will remark the time dependence because we will allow in our model to choose different beta distribution parameterisation in different time, like for example, in different economic cycle. In a discrete time set up, in order to consider a consistent and integrated risk-neutral framework, it is necessary to correlate the state space recovery rates structure with the forward rates term structure at any given time.

Let us define ρ as the (empirical) correlation between the term interest rate structure and the recovery rate²⁹, the assumed joint distribution is:

$$f(v, z) = \begin{cases} (+1, +1), & \text{with prob. } \left(\frac{1+\rho}{4}\right) \\ (+1, -1), & \text{with prob. } \left(\frac{1-\rho}{4}\right) \\ (-1, +1), & \text{with prob. } \left(\frac{1-\rho}{4}\right) \\ (-1, -1), & \text{with prob. } \left(\frac{1+\rho}{4}\right) \end{cases} \quad (4.2.1.4)$$

Moreover, let us define ρ' as the risk neutral probability vector³⁰ collecting the states probabilities of each branch of the lattice:

²⁹ The definition of the parameter ρ allows having one more degree of freedom, which enables to perform the proper recovery rates and interest rates correlation choice according to the overall economy time-scale considered in the model.

³⁰ The vector is risk neutral by construction having assumed that v and z takes on the value ± 1 with probability 0,5.

$$\rho' = \begin{bmatrix} \frac{1+\rho}{4} \\ \frac{1-\rho}{4} \\ \frac{1-\rho}{4} \\ \frac{1+\rho}{4} \end{bmatrix}$$

For computational needs (and for notation ease) it is useful to introduce another concept before to get the final explicit functional form for the forward spreads: the state-prices³¹. The state price (denoted by the variable $w(m)$) at time $m+1$ evaluated at time- m is defined as the price at time- m times the risk-neutral probability ρ' of being in that state at the time- m discounted at the risk-free interest rate, i.e.:

$$w(m\Delta) = \rho' \cdot w[(m-1)\Delta] \cdot \frac{1}{1+f_{m-1,m}} \quad (4.2.1.5)$$

where the state prices are considered as four dimensional vectors (corresponding to the four possible states defined by the double stochastic structure) for each seniority and $f_{m-1,m}$ is the forward rate between time $t=(m-1)\Delta$ and time $t=m\Delta$.

Both the interest rate term structure and the recovery rate structure are implied in the definition of the state price, track of them can be found in the discount and probability factors, respectively. The cash-flow at time step- m is a function of recovery rates as well as transition rates. While recovery rates are correlated to the risk-neutral interest rate structure, the transition probabilities have to be rendered risk-neutral in order to preserve the overall framework consistency.

For this purpose it is necessary to introduce the adjusting factors $\pi_i^s(m)$ to the empirical transition probabilities defined for any time step- m . Let us consider the one-period transition from a generic time- m to time $(m+1)$; this is performed by defining the unknown quantities

³¹ As pointed out by Das and Sundaram (2000) "State prices are the current value of a security that pays off a dollar in a single specific state in the future and zero in all other states. For example, if there are only two possible states ("up" and "down") at the same time in the future, then the state price of the "up" state would be the value of a dollar received in that state times the risk neutral probability of that state, discounted to the present, using a risk-less discount rate. State prices are useful since they allow computing the price of any security by multiplying the payoffs of the security by state prices in each node (state), and then adding these values up. Of course at time 0 the state price is simply unity. i.e. $w(0)=1$ "

$\pi_i^s(m)$ referred to the i -th rating class and to the s -th seniority type³² of the credit instrument at time step- m .

$$Q^s(m) = \{q_{i,j}^s(m)\} = \begin{bmatrix} q_{11}^s(m) & \dots & \dots & q_{1K}^s(m) \\ q_{21}^s(m) & q_{22}^s(m) & \dots & q_{2K}^s(m) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 - (1 - \pi_1^s(m)) \cdot \bar{q}_{11}(m) & \bar{q}_{12}(m) \cdot \pi_1^s(m) & \dots & \bar{q}_{1K}(m) \cdot \pi_1^s(m) \\ \bar{q}_{21}(m) \cdot \pi_2^s(m) & 1 - (1 - \pi_2^s(m)) \cdot \bar{q}_{22}(m) & \dots & \bar{q}_{2K}(m) \cdot \pi_2^s(m) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (4.2.1.6)$$

where $Q^s(m)$ is the risk neutral representation of $\bar{Q}(m)$, when incorporating the seniority type effect in the transition matrix by rating class, as shown in the generic element $q_{i,j}^s(m)$. The element $q_{i,j}^s(m)$, by construction, explicitly consider the adjustment factor $\pi_i^s(m)$. Invoking the definition of state price, for the credit instrument of seniority type- s being in class- i at time- m , the following condition, in a risk neutral world, must be satisfied:

$$w(m\Delta) \cdot E[C^s(m)] = \frac{1}{1 + f_{m,M}^{act} + s} \quad (4.2.1.7);$$

where $f_{m,M}^{act}$ is the actual forward interest in the period between time- m and maturity (time- M), s is the market spread and the expected cash flow at time- m for the bond of rating class- i and seniority type- s is determined by:

$$\begin{aligned} E[C^s(m)] &= \\ &= [q_{i,1}^s(m), q_{i,2}^s(m), \dots, q_{i,2}^s(m), \dots, q_{i,K}^s(m)] \cdot [1, 1, \dots, 1, \dots, \phi_s(m)] = \\ &= [\bar{q}_{i,1}(m)\pi_i^s(m), \bar{q}_{i,2}(m)\pi_i^s(m), \dots, 1 - (1 - \bar{q}_{i,i})\pi_i^s(m), \dots, q_{i,K}(m)\pi_i^s(m)] \cdot \\ &\quad \cdot [1, 1, \dots, 1, \dots, \phi_s(m)] \end{aligned} \quad (4.2.1.8)$$

³² One reason behind the choice of K rating class and s seniority type is that there are well studied tables of default frequencies for standard ratings, and there is not enough data to distinguish between different seniority types. Another reason is that while rating is subject to random changes the seniority class remains unchanged during the life of an asset

Eq (4.2.1.7) and Eq. (4.2.1.8) provide the solutions for the unknown π_i^s associated to the rating class- i and seniority type- s by calibrating those equations with the (average) market spread of the considered risky debt. In fact, making use of simple algebra, it is possible to show that Eq (4.2.1.7) is the generalisation of Eq (4.2.5) when considering the assumptions of the suggested integrated pricing model.

Applying the above-mentioned market price calibration it is possible to find all the adjustment factors to get the risk neutral transition matrix for seniority type- s at time t .

Five transition matrices for each rating class correspondent to the five seniority types are generated. Therefore the model can be split into five parallel models yielding specific information on the seniority for any rating class, at any time step. This information is then embedded in the final expression of the spread related to the seniority type. At this stage it is important to observe that, according to the data in table 1, default rates are not affected by the credit instrument dependence on seniority, while the recovery rate does. Within this unified risk neutral framework is possible to measure the contribution of the seniority type to the risk neutral spread curve.

4.2.2 Multiple time horizon

Up to now the attention was focussed on those variables, assumptions and parameters that have a direct impact on credit risk, without explicitly considering the time at which those quantities have been evaluated or defined. Another basic managerial aspect of credit risk is the time horizon of the risk measure. This measure of risk of a financial instrument is a critical issue. In fact, in this case, the problem of extrapolation, or interpolation, has to be faced in order to achieve the correct estimation of migration probabilities in the multiple time-horizons.

Let us consider the one-period probabilities $q^{s^0}_{ij}(m)$ as the probability of migrating from rating class- i to rating class- j in the time interval between time step $(m-1)$ and step m with respect to a generic recovery rate structure s . It is important to point out that the probabilities previously considered in the transition matrix elements at the generic time step- m are regarded as cumulative probabilities. The (actual) cumulative probabilities are obtained by the one step probabilities through a recursive procedure.

Under the assumption that the one period migration probabilities at subsequent time steps are independent, it is possible to obtain the actual rating class transition probabilities, from any class- i to any class- j , at the subsequent time step, by multiplying the actual migration matrix at time- m with the one at time $m+1$. This procedure may be applied recursively yielding for the

actual transition matrix at time T . Applying the risk-neutral adjustment procedure at any time step, as outlined in previous paragraph, the risk-neutral transition matrix at time period $m+1$ is directly derived. Notice that this structure allows embedding in any transition probability at the given time- m all the information on transition probabilities at previous time steps (maintaining probabilities independence). Therefore the single one-period transition probability $q_{ij}^s(m)$ keeps the information on $q_{kl}^s(n)$ for all states k,l and for all times steps- n ($n < m$). In particular, the default probabilities $q_{i,K}^s(m)$ contain the information on the previous time step transition probabilities.

This feature distinguishes the integrated model from the other reduced models outlined in section 1,2 and 4. Transition probability can change significantly over time. An investment grade has a higher chance of downgrade than of upgrade and viceversa³³. This means that, in the high rated firms, transition risk (and default probability in particular) increases over time and, by contrast, high yield risky debt that do not default, are more likely to improve than deteriorate in credit quality, thus showing a decreasing default probability over time.

4.2.3 One-period and cumulative transition probabilities

Before deriving the formula for the spread curve it is interesting to focus on transition probabilities. Provided that K rating classes are considered, $q_{iK}^{s,0}(m\Delta)$ is defined as the one-period default probability, i.e. the probability of migration from the rating class i to class K (corresponding to the default state), over the period from $[(m-1)\Delta, m\Delta]$, associated to the state i and generic seniority s ,

The cumulative probability of default at the time period- m ($t=m\Delta$) is defined as $q_{iK}^s(m\Delta)$, and it is a function of the previous-time cumulative probability (probability of having got default until the previous time step) and the one-period default probability (probability of getting to default between $(m-1)\Delta$ and $m\Delta$) as follows:

$$q_{iK}^s(m\Delta) = q_{iK}^s((m-1)\Delta) + [1 - q_{iK}^s((m-1)\Delta)] \cdot q_{iK}^{s,0}(m\Delta) \quad (4.2.3.1)$$

Conversely, the one period probability of default in the period indexed by m may be expressed

$$q_{iK}^{s,0}(m\Delta) = \frac{q_{iK}^s(m\Delta) - q_{iK}^s((m-1)\Delta)}{1 - q_{iK}^s((m-1)\Delta)} \quad (4.2.3.2)$$

³³ This phenomenon is known as mean reversion in credit ratings

These definitions are useful to compute expected cash flows over time for a zero coupon risky bond. Since $q_{iK}^s(m\Delta) > 0$, and the cumulative probability of default must be increasing:

$$q_{iK}^s(m, \Delta) - q_{iK}^s((m-1)\Delta) > 0 \quad (4.2.3.3)$$

then, default probabilities lie in the range $[0,1]$ as required. In this formalisation it is important to point out that by means of the procedure outlined above, the risk-neutral adjusted transition probabilities to default carry on the information of all the actual transition probabilities.

At this stage all the information required for deriving the spread structure as a function of its determinants has been derived and may be embedded into the cash flow evaluation.

With reference to Eq.(4.2.6) and Eq.(4.2.1.8), the expected cash flow at the m -th time period for the given seniority class- s in its explicit form is:

$$E[C_d^s(m)] = 1 \cdot [1 - q_{iK}^s((m-1)\Delta)] \cdot q_{iK}^{s0}(m\Delta) \cdot \phi_s(m\Delta, n) \quad (4.2.3.4)$$

which also generalise Eq.(4.2.1).

In the integrated model, within the multiple time horizon approach:

- a) the one-period probabilities are given by the first transition probability
- b) the cumulative probabilities are derived recursively.

The philosophy of the integrated model appears evident also at this stage since the strict correlation between the underlying model structure and the empirical data is assured at each step of the formulation: theory and actual data are interwoven in order to assure adherence between the theoretical process and the market dynamics. Recalling Eq.(4.2.3.4) it is possible to rewrite Eq.(4.2.6) in the following way:

$$\sum_{m=\frac{t}{\Delta}+1}^{\frac{T}{\Delta}-1} \left[E \left(\exp \left[-\Delta \sum_{j=\frac{t}{\Delta}}^{m-1} f_i(t, j\Delta) \right] C^s(m) \right) \right] = \exp \left[-\Delta \sum_{j=\frac{t}{\Delta}}^{\frac{T}{\Delta}-1} (f(t, j\Delta) + sp^s_i(j\Delta)) \right] ; \quad \forall i \quad (4.2.3.5)$$

Making use of both the definition of state prices and cash-flow in case of default (see Eq. 4.2.1.7) at any time-step- m Eq.(4.2.3.5) becomes:

$$\sum_{m=\frac{t}{\Delta}+1}^{\frac{T}{\Delta}-1} \left[\sum_{n=1}^m w(m\Delta, n) [1 - q_{iK}^s(m-1)\Delta] q_{iK}^{s0}(m\Delta) \phi_s(m\Delta, n) \right] = \exp \left[-\Delta \sum_{j=\frac{t}{\Delta}}^{\frac{T}{\Delta}-1} (f(t, j\Delta) + sp^s_i(j\Delta)) \right] \forall i \quad (4.2.3.6)$$

4.3 Spread term structure by rating class contingent to the seniority of debt

The term structure of forward credit spreads estimation is the problem to be solved in the last step of the process. For any rating class and seniority type the following spread, $\{sp^s_i(t)\}$ is evaluated

$$\{sp^s_i(t)\} \quad i=1,\dots,K; \quad s=1,\dots,5; \quad t < T \quad (4.3.1)$$

In order to give the spread curve in its explicit form it is necessary to consider its integral formulation. The spread curve evaluated at time t for a given rating class- i is defined by all spreads computed at consequent time steps within the bond life-span, specifically in the time interval $[t, T]$. Let us consider Eq.(4.2.3.6) and define the integral spread curve $S(\mu, M)$ between the μ -th and the M -th period (corresponding to any given time $\tau \in [t, T]$ and maturity $t=T$, respectively)

$$\begin{aligned} SP^s_i\left(\frac{\tau}{\Delta}, \frac{T}{\Delta}\right) &\equiv SP^s_i(\mu, M) \equiv \sum_{j=\frac{\tau}{\Delta}}^{\frac{T}{\Delta}-1} sp^s_i(j\Delta) \equiv \\ &\equiv -\frac{1}{\Delta} \cdot \text{Ln} \left\{ \sum_{m=\tau/\Delta+1}^{T/\Delta} \left[\sum_{n=1}^m w(m\Delta, n) [1 - q^s_{ik}(m-1)\Delta] q^{s_0}_{ik}(m\Delta) \phi_s(m\Delta, n) \right] \right\} - \sum_{j=\frac{\tau}{\Delta}}^{\frac{T}{\Delta}-1} f(\tau, j\Delta); \end{aligned} \quad (4.3.2)$$

In order to derive the spread curve at time step- μ the following differential relation is used

$$sp^s_i\left(\frac{\tau}{\Delta}\right) \equiv sp^s_i(\mu) \equiv SP^s_i(\mu, M) - SP^s_i(\mu-1, M) \quad (4.3.3)$$

Finally, referring to Eq. (4.3.2) the forward interest rate spread is determined as:

$$\begin{aligned} sp^s_i(\mu) &\equiv -\frac{1}{\Delta} \cdot \text{Ln} \left\{ \frac{\sum_{m=\tau/\Delta+1}^{T/\Delta} \left[\sum_{n=1}^m w(m\Delta, n) [1 - q^s_{ik}(m-1)\Delta] q^{s_0}_{ik}(m\Delta) \phi_s(m\Delta, n) \right]}{\sum_{m=\tau/\Delta}^{T/\Delta} \left[\sum_{n=1}^m w(m\Delta, n) [1 - q^s_{ik}(m-1)\Delta] q^{s_0}_{ik}(m\Delta) \phi_s(m\Delta, n) \right]} \right\} - \\ &\quad - \left[\sum_{j=\frac{\tau}{\Delta}}^{\frac{T}{\Delta}-1} f(\tau, j\Delta) - \sum_{j=\frac{\tau}{\Delta}-1}^{\frac{T}{\Delta}-1} f(\tau-\Delta, j\Delta) \right]; \quad i=1,\dots,K; \quad s=1,\dots,5; \quad t \leq \tau \leq T \end{aligned} \quad (4.3.4)$$

The last-period forward spread between time T and time $T+\Delta$ relative to the i -th rating class and adjusted for the seniority s is denoted by $sp^s_i(T) \equiv sp^s_i(M)$, by computing node- T on the tree of the interest forward rate structure, the spread is derived considering the last period expected cash-flow in case of default without considering previous cash-flow events. Referring to Eq.(4.3.4) it is straightforward to derive the last-period spread as follows

$$\exp[-\Delta sp^s_i(T)] = E \left[1 - \frac{q^s_{iK}(T+\Delta) - q^s_{iK}(T)}{1 - q^s_{iK}(T)} (1 - \phi_s(T+\Delta)) \right], \quad \forall i \quad (4.3.5),$$

which gives

$$\exp[-\Delta sp^s_i(T)] = E \left[1 - \frac{q^s_{iK}(T+\Delta) - q^s_{iK}(T)}{1 - q^s_{iK}(T)} (1 - \phi_s(T+\Delta)) \right], \quad \forall i \quad (4.3.6)$$

5. Applications

Our model requires only easily available information as input, namely: the risk-free yield curve, the term structure of credit spreads for each rating class, the statistical transition matrix and both mean and standard deviation of the recovery rate by seniority of debt. The most important and useful resultant model information are: risk neutral transition matrix and risk neutral spread term structure both contingent to the seniority of debt. Moreover, the bivariate lattice, thorough which the STSRC and SSD has been estimated, was built by correlating risk-less interest rates and recovery rates thus considering the integration between market and credit risk. Using this information, the following products, among others, are priced by generating the necessary cash flows at each node on the lattice and discounting the cash flows back by multiplication by the state prices to obtain present values: plain-vanilla risky debt for any rating class and any seniority of debt; rating-sensitive debt, in particular, our model perform quite well to price such debt since the rating transition matrix provides risk-neutral information on rating changes (adjusted to the seniority) which can be directly used to generate cash flows at each node on the tree; spread-adjusted notes: the coupon may also be indexed to the spread at each node, this is achieved by computing the forward spread at each node on the lattice and, since the price of risky debt is known at each node, and so is the risk-less rate, it is quite simple to compute the credit spread at each node as well; spread option: given the computed spread at each node, it is possible to price spread option since cash flows may be generated at each node by comparing the spread at the node with the strike rate; total return swaps: since the price of any underlying risky bond is computable at each node on the tree, the total return on the bond may also be easily calculated. Although the model is rich and flexible enough to price many credit assets, both plain vanilla and derivative, we think is particularly appropriate to price defaultable loans and corporate bonds.

6. Conclusions and further research

The stochastic spread structure model considered within the *integrated model* allows taking into account effects due to rating transitions (including default events) and recovery rates depending on seniority. The overall procedure allows discriminating the effects of the credit instrument belonging to a specified rating class at any given time; actually fixing the time step in the forward interest rate term structure, $k-1$ spreads corresponding to the defined rating classes are derived. More specifically, this model is aimed at computing the spread for credit instruments belonging to a defined rating class and having a specified seniority, so that to discriminate the information relative on the given seniority. This framework allows depicting the effects on spread curves due to the rating class and -for any given rating class- the effects due to the different seniority types using the risk neutral arbitrage set-up.

Further research on this area will be devoted both on considering the influence of the economic cycle on the estimation of recovery contingent to seniority and analysing the structural interdependencies between recovery rate and default probability. Moreover the issue of default correlation and its impact on pricing risky debt should also be investigated. Finally, from a practical point of view, there are at least two other relevant issues that need to be carefully taken in consideration in future work, namely both liquidity risk and parameter calibration. Our intuition is that we need an integrated pricing and risk model to exploit in a coherent framework the risk and capital management banking problem.

Chapter 2

A Framework for Risk Measures and Risk Adjusted Performance¹

We might decide that in one context one basic set of principles is appropriate, while in another context a different set of principles should be used.

Markowitz (1959)

¹ I wish to thank G. B. Adesi, E. Altman, A. Saunders, S. Hodges and seminar participants at Fointenbleau INSEAD, Eltville Frankfurt Center for Financial Studies, Bundesbank Training Department and Milan, Italian Banking Association, for helpful discussion of this study. Earlier versions of this chapter were presented at the Euro Working Group on Financial Modelling Meeting (Trondheim and New York 2000), at the Annual Meetings of both Financial Management Association (Toronto 2001) and European Financial Management Association (Helsinki 2003).

1. Introduction

The comparison of expected returns with the standard deviation of portfolio returns, suggested by the mean-variance portfolio theory of Markowitz (1952) and Sharpe (1964), has long been the benchmark for performance measurement, both because of its tractability and because of the clear intuitions it provides about diversification and the efficiency of portfolio allocation. However, the strong assumptions and substantial practical limitations of mean-variance portfolio theory are also widely known. Financial returns exhibit asymmetry (skewness), fat tails (kurtosis) and are all too often subject to liquidity problems that markedly increase the level of downside risk in stressed situations. Traditional performance measures based on mean-variance theory, such as the Sharpe ratio, Treynor ratio, and the Jensen alfa, generate paradoxical results when applied to portfolios with non-normally distributed returns.

Investors also exhibit a huge range of attitudes towards risk, typically either penalising extreme losses and/or valuing substantial gains to a much greater extent than would be predicted by a quadratic preference function. In theory, non-traditional performance evaluation techniques (based on other models of investor preferences and capital market equilibrium) can still be applied to non-normal returns, incorporating a range of different utility assumptions. Nevertheless, in practice, utility is not directly observed. Simple, intuitively tractable, measures of risk and performance are needed in order for portfolio managers and investors to communicate with each other and decide their investment policies.

The limitations of mean-variance measures and the inapplicability of utility theory to practical decision making, explain the interest of portfolio managers in a great variety of other measures of risk, used to assist them with their investment decisions. Downside measures of risk such as expected shortfall, conditional tail expectation, or lower-partial moments are attractive as a tool that emphasises the potential for loss. Upside potential can be captured by such measures as the Sortino ratio (Sortino, 1991) and the Upside Potential ratio, Sortino and Forsey (2001). However, how are portfolio managers to compare these different measures, embodying wholly different approaches to summarizing risk? Since there exist many (and conflicting) objectives, choosing a single “correct” measure of risk is unrealistic. There are also great differences in return distributions so that a measure of risk appropriate for one type of exposure and return distribution – e.g. equity risk -- may be much less appealing when applied to another – e.g. credit risk².

² To complete the picture, it is worth to mention that recent research in the behavioural finance area describes how investors say they want to behave. In general, investors do not seek the highest return for a given level of risk, as

This chapter confronts these issues. It reviews the wide range of risk measures currently used in portfolio management, assessing them both from the perspective of stochastic dominance and from the coherence axioms of Antzner et. Al. (1997). It then proposes the use of a particular class of risk measures – the distortion risk measures – for portfolio management.

Distortion risk measures have been developed by actuarial scientists (Wang 1995), providing an elegant mathematical solution to the problem of summarizing risk. One particular attraction of distortion risk measures is the clarification of conditions under which a risk measure satisfies both the axioms of coherence and the different orders of stochastic dominance. Distortion risk measures can be chosen that are coherent and satisfy at least the first two orders of stochastic dominance. Moreover, as we will emphasise, a portfolio manager can make a choice within a “family” of distortion measures, so as to achieve an appropriate balance of downside risk (potential losses) and upside risk (potential gains). Distorted measures of risk do not provide a single risk-measure applicable to all investment situations. What they do provide is a single treatment of risk-measurement that can incorporate varying preferences towards both upside and downside risks within a single measurement framework, whilst also ensuring that investment choices are consistent with both stochastic dominance and coherence³.

The paper is organized as follows. Section two provides an overview of the literature. Section three is aimed at briefly describing the most important appealing risk measures features: coherence and stochastic dominance. Section 4 reviews the most widespread used risk measures in the banking, insurance and asset management industry. In this section investors is considered to approach decision-making under uncertainty with different risk definitions, risk measures and risk risk-reward criteria. For this purpose through section 4.1, 4.2 and 4.3 risks are categorized in four classes:

- Risk as maximum potential loss (given a set of possible states of nature), typically quantile risk measure such as Value at Risk (VaR);
- Risk as the magnitude of deviation from a target (two sided and one side), typically conditional measure of shortfall risk such as expected shortfall (ES), conditional tail expectation (CTE) and tail VaR (TVaR);

traditional portfolio theory assumes. Investors seek upside potential with downside protection. Investors desire consistency of return and therefore choose decision processes that preserve appropriate future financial flexibility. Rather than maximize the expected return, they want to maximize a “satisfiable” strategy.

³ From a shareholder and management perspective, for the sake of portfolio optimization and sound risk management, it is essential for a risk measure to properly reflect the risk differentials in alternative strategies or portfolios. In this case a risk measure, although being coherent, ignores useful information in a large part of the loss of the distribution, and consequently lacks incentive for mitigating losses below the quantile VaR. A good risk measure for portfolio management must therefore go beyond coherence.

- Risk as the magnitude of deviation from the expected utility;
- Risk as maximum potential loss defined as the difference between the wealth today and the end-of period wealth.

In section five a risk measure used in actuarial science, namely the distortion risk measure with increasing and strictly concave distortion function, firstly introduced by S. Wang (1995), that are shown to be coherent and stochastic dominance compliant, is analysed. In section six the most widespread risk-adjusted performance used in the asset, banking and insurance industry are discussed, namely: Sharpe Ratio; Information Ratio; Jensen Alfa; Treynor Ratio (based on traditional equilibrium portfolio theory); ROVaR, Return on Value (or Capital) at Risk; Generalised Sharpe Ratio; Sortino Ratio; Kappa; Omega and the Upside Potential Ratio. Section 7 will illustrate the key features, strengths and weaknesses of these RAPMs by considering an end user perspective within three different frameworks: quadratic utility theory, mean variance (normally distributed returns), deviation from investor target risk-return. Section 7.3 presents the main contribution of the paper. A RAPM based on distorted risk measure is introduced aimed at taking advantage, in term of pro-active risk-return management, of all information provided by the whole distribution of portfolio value (return). Section 8 concludes and summarises the main findings of this chapter.

2. Overview of the literature

Financial portfolio optimisation is a mature field that grew out of the Markowitz's mean-variance theory, and the theory of expected utility. Both theories rely on the numerical representation of the preference relation investors have for assets with random outcomes. It is also assumed that investors are averse to the variability of random outcomes (or risk). Once a numerical representation of the investors' behaviour is obtained, it is possible, in practice, to use different optimisation methods to compute the optimal allocation of assets for a particular investor.

The process of performing an optimal asset allocation basically deals with the problem of finding a portfolio that maximizes the expected utility of the portfolio manager. As long as it is supposed that the returns of the portfolio assets follow a normal distribution, the return distribution of any portfolio considered will also be normal. In this case, as it is done throughout the traditional portfolio theory introduced by Markowitz (1952) and Sharpe (1964), the problem of finding an expected, utility-maximizing portfolio for a risk-averse portfolio manager, represented by a concave utility function, can be restricted to finding an optimal

combination of the two parameters mean and variance. This method dramatically simplifies the whole asset allocation process, and is known as mean-variance analysis. It is the aim of the portfolio manager to find a portfolio that maximizes his expected return under a given risk level or a portfolio that minimizes his risk under a given return level. Risk, in this case, is measured by the variance of the portfolio return. Unfortunately, selection rules based on the two parameters mean and variance are of limited generality as they are optimal only if the utility function is quadratic or the return distribution is normal.

In deriving the CAPM, Sharpe, Lintner and Mossin assume expected utility (EU) maximization following the approach proposed by Markowitz, normal distributions and risk aversion. Kahneman & Tversky suggest Prospect Theory (PT) and Cumulative Prospect Theory (CPT) as an alternative paradigm to EU theory. They show that investors distort probabilities, make decisions based on change of wealth, exhibit loss aversion and maximize the expectation of an S-shaped value function that contains a risk-seeking segment. Employing change of wealth rather than total wealth contradicts EU theory. The subjective distortion of probabilities violates the CAPM assumptions of normality and homogeneous expectations, and the S-shaped value function violates the risk aversion assumption. Prospect Theory claims characteristics of investors' behavior which contradict the expected utility theory in general, and the classical assumptions of the CAPM in particular, but unfortunately it does not suggest any equilibrium pricing model which can substitute the existing expected utility model and in particular the CAPM. The Sharpe-Lintner-Mossin CAPM is derived by assuming that investors are risk-averse, that they maximize expected utility of total wealth, and that the returns are normally distributed with homogeneous expectations regarding these distributions⁴. Experimental studies cast doubt on the foundations of the CAPM. Based on experimental findings, Prospect Theory (PT) and Cumulative Prospect Theory (CPT) (see Tversky and Kahneman (1992))⁵, were developed as an alternative paradigm to expected utility. On the one hand, PT and CPT have become a cornerstone in economic research and are the foundation of behavioral finance and behavioral economics.

On the other hand, the CAPM is still the most popular asset-pricing model. PT asserts that probabilities are distorted. This violates two assumptions of the CAPM: first, the normality

⁴The normality assumption can be relaxed by adding the assumption of quadratic utility functions. Because the quadratic utility has two severe drawbacks ($U' < 0$ from some critical value, and increasing absolute risk aversion) researchers generally are not willing to assume this utility function. There are other justifications of the CAPM. Merton (1973) assumes continuous portfolio revisions which leads to end of period lognormal distributions of returns and to an instantaneous CAPM. The CAPM can be obtained also as a special case of the Arbitrage Pricing Theory (APT), see Ross (1976). In this paper I use the classical Sharpe-Lintner-Mossin CAPM assumptions, i.e., normal distribution is assumed.

⁵ They prove in this paper that although CPT (and PT) is in conflict to EUT, and violates some of the CAPM's underlying assumptions, the security market line theorem (SMLT) of the CAPM is intact in the CPT framework.

assumption is violated, and secondly, as each investor has his/her subjective probability distortion, investors face heterogeneous probability distributions of returns, even if before the distortion they all face the same normal return distributions. Thus, the normality and the homogeneous expectation CAPM assumptions are violated. PT asserts that investors make decisions based on change of wealth that violates EUT asserting that decision-making should be based on total wealth rather than change of wealth.

Moreover, no traditional (that is, based on capital market equilibrium models) performance evaluation technique applies when assessment the performance of investments in presence of non-normal return distribution and/or lacking liquidity. Traditional performance measures such as the Sharpe ratio, Treynor ratio, Jensen alfa usually generate paradox results when facing with non-normally distributed returns. This can be traced back to the definition of risk used by those measures. Variance, for example, is simply inappropriate for non-symmetrical return distribution. In addition, lacking market liquidity implies lacking market or benchmark index, lacking measures of co-movements (like co-variance) and lacking comparability between investments. Unfortunately, the Sharpe ratio is prone to manipulation – particularly by strategies that can change the shape of probability distribution of returns.

For example, Dybvig and Ingersoll (1982) show that non-linear payoffs limit the applicability of the Sharpe ratio to the problem of performance evaluation. More recently, Altman, Onorato and Pastorello (2000) show that Sharpe ratios are particularly misleading when the shape of the return distribution is far from normal. Other researchers, recognizing the limitations of the Sharpe ratio and its relatives, have sought alternatives to the reward-to-variability approach. These include stochastic-discount factor based performance measures (c.f. Chen and Knez (1996)).

Moreover, the Sharpe ratio and other related reward-to-risk measures may be manipulated with option-like strategies. A detailed analysis of this topic can be found in W. Goetzmann and al. (2002)⁶. In this paper they derive the general conditions for achieving the maximum expected Sharpe ratio and static rules for achieving the maximum Sharpe ratio with two or more options, as well as a continuum of derivative contracts. The optimal strategy rules for increasing the Sharpe ratio. Their results have implications for performance measurement in any setting in which managers may use derivative contracts. In a performance measurement setting, they suggest that the distribution of high Sharpe ratio managers should be compared with that of the optimal Sharpe ratio strategy. This has particular application in the hedge fund industry where

⁶ Nber Working Paper Series, Sharpening Sharpe Ratios, William Goetzmann, Jonathan Ingersoll, Matthew Spiegel, Ivo Welch, Working Paper 9116, <http://www.nber.org/papers/w9116>, national bureau of economic research, August 2002.

use of derivatives is unconstrained and manager compensation itself induces a non-linear payoff. The shape of the optimal Sharpe ratio leads to further conjectures. Expected returns being held constant, high Sharpe ratio strategies are, by definition, strategies that generate regular modest profits punctuated by occasional crashes. Their evidence suggests that the "peso problem" may be ubiquitous in any investment management industry that rewards high Sharpe ratio managers

As an alternative, risk measures based on moments of the underlying return distributions became very popular. The fundamental aspects of downside risk approaches were published in Roy (1952). His paper can be seen as the starting point for safety first and particularly for portfolio theory. An important and not very well known paper was written by Telser (1955). This paper examines shortfall restrictions, i.e. it restricts the universe of efficient portfolios to those, which can be characterized by reaching a certain minimum return with a minimum level of probability. The minimum return is called threshold return, the minimum probability for failing to reach the threshold return is usually referred to as shortfall probability. Both of them are concerned with the probability k for the stochastic portfolio return \tilde{r} lacking a certain level of benchmark return r^* : $k = P(\tilde{r} \leq r^*)$. The abbreviation P represents "probability". In Roy (1952) this case is investigated for probability distributions characterized by the mean and the standard deviation by using the Tschebyscheff inequality.

A straightforward, brief introduction to downside risk management for practitioners can be found in Sortino and van der Meer (1991). Pelsser and Vorst (1990) present an extremely advanced and significant working paper. It shows how the probability distributions of optioned portfolios can be estimated and how the consequences for the shortfall risk of optioned portfolios must be implemented. Although this paper provides insights especially for determining the probability distributions of optioned stock portfolios the authors analyze how to meet the needs of investors by imposing shortfall risk restrictions for an optioned stock portfolio. They use a similar method to specify those probability distributions as is the case in Bookstaber's and Clarke's (1983) approach¹. However, in contrast to Bookstaber and Clarke they use continuous time stochastic processes to describe the underlying stock prices.

This allows them to specify the process of the stock prices in accordance with the assumptions of the Black and Scholes model for pricing options and excludes that negative stock prices can occur with positive probabilities. Pelsser and Vorst (1990) use a compounded Wiener process for the stock prices which provides a systematic and an unsystematic risk component. Then they identify optimum portfolios by maximizing the end-of-period wealth subject to shortfall restrictions. Although the shortfall risk of optioned portfolios is not based on normal

distributions they transform the shortfall restriction in order to receive linear restrictions. For a solution of the maximization problem a linear program has to be solved.

The Leibowitz and Henriksson (1989) approach was constructed by paying special attention to the management of pension fund assets. In this framework the threshold return is typically assumed to be deterministic. This approach is known in the literature as the surplus management approach. In practice, the relevant threshold benchmark for pension fund managers may, for instance, be the growth rate of economic gross wages which randomly fluctuates. A more appropriate modeling of this fact can be found in Leibowitz and Kogelman and Bader (1991) where a two asset case is considered with stochastic liabilities.

An extension of the surplus shortfall constraint concept is provided by Wolter (1993). He gives a generalization for a portfolio consisting of n assets where the variance of surplus returns, i.e. of the return differential between assets and liabilities, is minimized. As in the case of the classical mean/variance optimization the resulting function is characterized by a hyperbola. The safety first approach provides a method which characterizes adequately the decision situations of pension fund managers. However, a serious disadvantage of the shortfall probability k as a measure of risk is that the extension of failing to attain the threshold return does not play a considerable role: The extent of falling short relative to a threshold return is not respected. For instance, whether the threshold return is missed by 0.1% or 10% does not influence the decision, since merely the probability of falling short of a threshold return is relevant rather than the extent of the shortfalls. Therefore, a generalization of the safety first approach was provided with lower partial moments (LPM). Harlow (1991) gives an easily understandable introduction to lower partial moments. He defines lower partial moments as downside risk measures as follows

$$LPM_l = \int_{-\infty}^{r^*} \frac{P(\tilde{r} = r)}{P(\tilde{r} \leq r^*)} (r - r^*)^l dr$$

A lower partial moment of order l is defined as the expected value of the downside deviations to the power of l of a portfolio return from a certain level of threshold return r^* .

For instance, if $l=0$ the lower partial moment LPM_0 exactly equals the shortfall probability k , the base of the exponential term strictly equals one, consequently the extent of deviations of the portfolio return from the threshold return is not respected. If l equals one, then LPM_1 may be interpreted as the expected portfolio return below the threshold return. Furthermore, for $l=2$ the Harlow definition is similar to the general definition of the variance except for the fact that the threshold return does not necessarily equal the expected portfolio return, and except for the fact that the integral is limited at the top by the threshold return. Thus, characterization as downside variance and in the case of the square root as downside volatility would be an appropriate description. The downside volatility is also characterized in Wolter (1992). Harlow shows how to

construct a portfolio selection program, where the LPM_1 is minimized instead of the variance of the portfolio. He points out that lower partial moments are useful if the portfolio returns are not significantly normally distributed, especially if they are skewed. The latter effect is typically provoked by using options in stock portfolios or by using credit asset in credit portfolio risk management.

Since lower partial moments characterize portfolios whose return distributions are skewed, downside risk measures and probability distributions of optioned stock portfolios are in tight context. Moreover, the literature concerning the (credit asset and optioned portfolios) density functions of both LPMs and probability distributions cannot be separated. A more advanced application of the lower partial moments is in Harlow and Rao (1989). They provide an asset pricing model in a very general mean/lower partial moment framework (MLPM). In their paper they refer to Bawa (1975 and 1978), where it is shown that a MLPM framework is consistent with a very general set of utility functions and asset return distributions. Especially the HARA (hyperbolic absolute risk aversion) class of utility functions is completely consistent with LPM_1 , whereas LPM_2 is consistent with all risk averse utility functions where the first and the third derivatives are positive and the second derivative is negative, i.e. for all utility functions displaying a skewness preference. For positive third derivatives of the utility functions, maximizing the expected utility in

$$E[U(W)] = EU(\bar{W}) + \left[\frac{1}{2!} U''(\bar{W}) \right] \sigma_w^2(\tilde{W}) + \sum_{i=3}^{\infty} \left[\frac{1}{i!} U^{(i)}(\bar{W}) \right] m_i(\tilde{W})$$

requires maximizing $\sum_{i=3}^{\infty} \frac{1}{i!} U^{(i)} m_i$ which implies maximizing $\frac{1}{6} U^3 m_3$. Since m_3 represents the skewness of the probability distribution, maximizing expected utility implies maximizing the skewness and hence truncating the downside potential of portfolio returns and maintaining the upside potential. It follows that the skewness preference of investors is obtained by constructing utility functions with positive third derivatives, i.e. for investors with decreasing absolute risk aversion.

The authors also refer to Bawa and Lindberg's (1977) model where the threshold return is assumed to be the riskfree rate of interest. Harlow and Rao (1989) provide an asset pricing model where the expected portfolio return is linearly dependent on changes in the market portfolio in the same manner as in the CAPM. The beta factor is modified and becomes a ratio where the generalized co-lower partial moment between the portfolio and the market portfolio is divided by the generalized lower partial moment of the market portfolio. Special attention has to be paid for the tests of their empirical findings. Harlow and Rao (1989) provide evidence that the new model

cannot be rejected whereas the CAPM is rejected. They specify different levels of target returns and obtain the best results when the target return is chosen as the average market return. This shows that the market participants are merely concerned with the downside risk of their holdings subject to the expected market return and less with the beta risk.

As was already discussed before, this is an extremely interesting approach to downside risk, due to the fact that options are included in stock portfolios here. A technique for estimating the return distribution of such portfolios is given and then applied to arbitrary portfolios consisting of stocks, call and put options. The selection of assets is optimized subject to certain shortfall constraints. An extremely important issue in the context of shortfall risk is the role of the investment time horizon effect. In this context Harlow (1991) showed that the shortfall risk is not an appropriate measure of risk if using put options skews the probability distribution of the portfolio. Furthermore he points out that the shortfall risk, given a constant annualized threshold return, is a function of the investment time horizon. The practical implication for pension funds is that the deviation of investment time horizon, which is typically very long termed, and the reporting horizon, normally one year, force suboptimal portfolios to be chosen. That is, portfolios, which are not volatile enough, provide too low returns.

One way to overcome these paradoxes is define what are the appealing features a risk measure should have and what rule of rationality should obey when selecting investment alternatives especially in the most general case of non-normal return distributions. In this thesis two main attributes are believed to be appropriate for measure of performance that cope with non normal distributed returns, and therefore show asymmetry and kurtosis⁷: stochastic dominance and coherence.

Stochastic dominance principles were first discussed in the seminal paper of Bawa (1975). Extensive research has been done to derive concepts for ordering uncertain prospects resulting in principles such as the stochastic dominance of order 1, 2 or 3 (see, e.g., Bawa 1975 or Martin et al. 1988). Bawa argues that it is quite reasonable to assume that the average investor makes a decision that is consistent with a finite, increasing and concave utility function with decreasing, absolute-risk aversion

A performance measure is coherent (first introduced by Artzner and al. (1997/99)) if shows the following properties: subadditivity, positive homogeneity, decreasing monotonicity and

⁷ Skewness measures the asymmetry of a return distribution. Positive skewness denotes a distribution where more occurrences are located below the mean but in a limited range, while there is significant upward potential in the return distribution. Preference for positive skewness means that from two return distributions, other relevant factors being equal, investors prefer the positively (or right) skewed one. Preference for positive skewness also implies that investors are willing to give up some expected return in order to participate in chances of much larger gains. Kurtosis describes the degree to which the distribution is weighted toward its tails. Aversion toward kurtosis means that investors dislike dispersion in terms of larger fat tails of the probability distributions, that is, occurrences on both ends of the probability distribution receive more weight than in the definition of the variance.

translation invariance. End users/decision makers when ranking investment alternatives implicitly accept these properties. The issue is that well known and widespread used risk adjusted performance measures like the Return on Value at Risk (ROVaR) do not comply with these properties and therefore decision making can be heavily biased.

From a portfolio management perspective, it is possible to observe that during the last few years some academics have started to address the problem of incorporating VaR and other related measures in the optimization of portfolios. Wang (2000) proposed a Mean-Variance-VaR approach that uses both variance and VaR as a double risk measure simultaneously. Grootveld & Hallerbach (2000) investigated the properties of a Mean-VaR optimization problem when applied in practice analyzing the effect of estimation risk on portfolio composition. Acerbi & Tasche (2002) proposed the use of the Expected Shortfall as an alternative to VaR in order to overcome the flaws that VaR has as an incoherent risk measure.

The axiomatic approach to risk measures is an important and very active subject, which applies to different topics of actuarial and financial interest like premium calculation and capital requirements. Besides the coherent risk measures by Artzner et al., one is interested in the distortion risk measures were introduced in the actuarial industry by Wang (1995/96) and Wang et al.(1997). Under certain circumstances, distortion risk measures are coherent risk measures (e.g. Wang et al.(1997), Theorem 3). For this reason, they can be used to determine the capital requirements of a risky business, as suggested by several authors including Wirch and Hardy (1999), Goovaerts et al.(2002) and Wang (2002). On top of each class of risk measures risk adjusted performance are built up by means a classical risk-reward approach. In this chapter, strengths and weaknesses of each of them are analysed in light of coherence and stochastic dominance assumptions, on one hand, and the use in portfolio management, in the other hand.

3 Appealing features of risk measures: stochastic dominance and coherence⁸

Consider a decision maker faced with a number of risks. A risk measure is defined as a mapping from the set of random variables representing the risks at hand to the real line. Note that in the sequel, I will always consider random variables, for notation easy from now on $\tilde{X} = X$, as losses⁹. In the next sections the following definitions and notations will be used:

$\mathbf{X} = (X_1, X_2, \dots, X_n)^T$: n-dimensional random vector

$\rho(X)$ general risk measure associated with the loss random variable with $\rho(X) \in \mathfrak{R}$

$F_X(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$: cumulative distribution function (cdf)¹⁰

Φ denotes the standard normal cdf

Φ^{-1} denotes the inverse of the standard normal cdf

$\bar{F}_X = 1 - F_X$: tail function¹¹;

$f_X(x_1, x_2, \dots, x_n) = \frac{\partial^n F_X(x_1, x_2, \dots, x_n)}{\partial x_1 \dots \partial x_n}$: probability density function (pdf)

$\Phi' = \phi(z)$ denotes the pdf of the standard normal distribution¹²

x_p | p^{th} quantile with $\bar{F}_X(x_p) = 1 - p$

The p-quantile risk measure of F_X can also be written as

$$F_X^{-1}(p) = Q_p(X) = x_p = \inf\{x \in \mathfrak{R} | F_X(x) \geq p\} \quad p \in (0,1)$$

The expectation of X , $E[X]$, if it exists, can also be written as¹³

⁸ These section is based on the following research papers: McCarl, (1996) Stochastic Dominance Notes, working paper. R. D. MacMinn (1987) Notes on Stochastic Dominance, working paper. A. Diederich (1998) Conflict and the Stochastic Dominance Principle of Decision Making Indiana University Cognitive Science Program Tech Report 224. Jay R. Lund (2003) Probabilistic Design and Optimization working paper, Spring, 2003 University of California, Davis. H. Grootveld, W. Hallerbach, (1999), Variance vs downside risk: Is there really that much difference? European Journal of Operational Research 114, 304-319. J. Dhaene, R.J.A. Laeven, S. Vanduffel, G. Darkiewicz, M.J. Goovaerts, (2004) Can a coherent risk measure be too subadditive?, working paper, University of Amsterdam. Goovaerts, De Vylder & Haezendonck (1984), or also chapter 5 in Kaas, Goovaerts, Dhaene & Denuit (2001). F. Delbaen (2000), Coherent Risk Measures, working paper, University of Pisa. Artzner, P., Delbaen, F., Eber, J.-M. and D. Heath (1997). Thinking coherently. RISK 10(11), 68-71. Artzner, P., Delbaen, F., Eber, J.-M. and D. Heath (1999). Coherent measures of risk, Mathematical Finance 9(3), 203-28.

⁹ A negative outcome for the loss variable means that a gain has occurred

¹⁰ A cdf is, by construction, both a monotone increasing and right-continuous function.

¹¹ Also, obviously, results: $\bar{F}_X = 1 - F_X$ and $\bar{F}_X(x_1, x_2, \dots, x_n) = P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n)$

¹² $\Phi' = \phi(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{z^2}{2}}$ with z standard normal variable. $\phi(z)$ therefore denotes the density function of the standard normal distribution. For example let $F_X^{-1}(p) = \Phi^{-1}(p)$, where F is standard normal, and $p=0.05$ i.e. consider an interval confidence of 95%. Substituting results: $F_X^{-1}(0.05) = \Phi^{-1}(0.05) = Q_{0.05}(X) = 1.644 = z$ and $\Phi' = \phi(1.644) = 0.1031$, where Φ' represents the probability mass function evaluated at $z=1.644$.

$$E[X] = - \int_{-\infty}^0 [1 - \bar{F}_X(x)] dx + \int_0^{\infty} \bar{F}_X(x) dx, \text{ with } \bar{F}_X(x) = Pr[X > x]$$

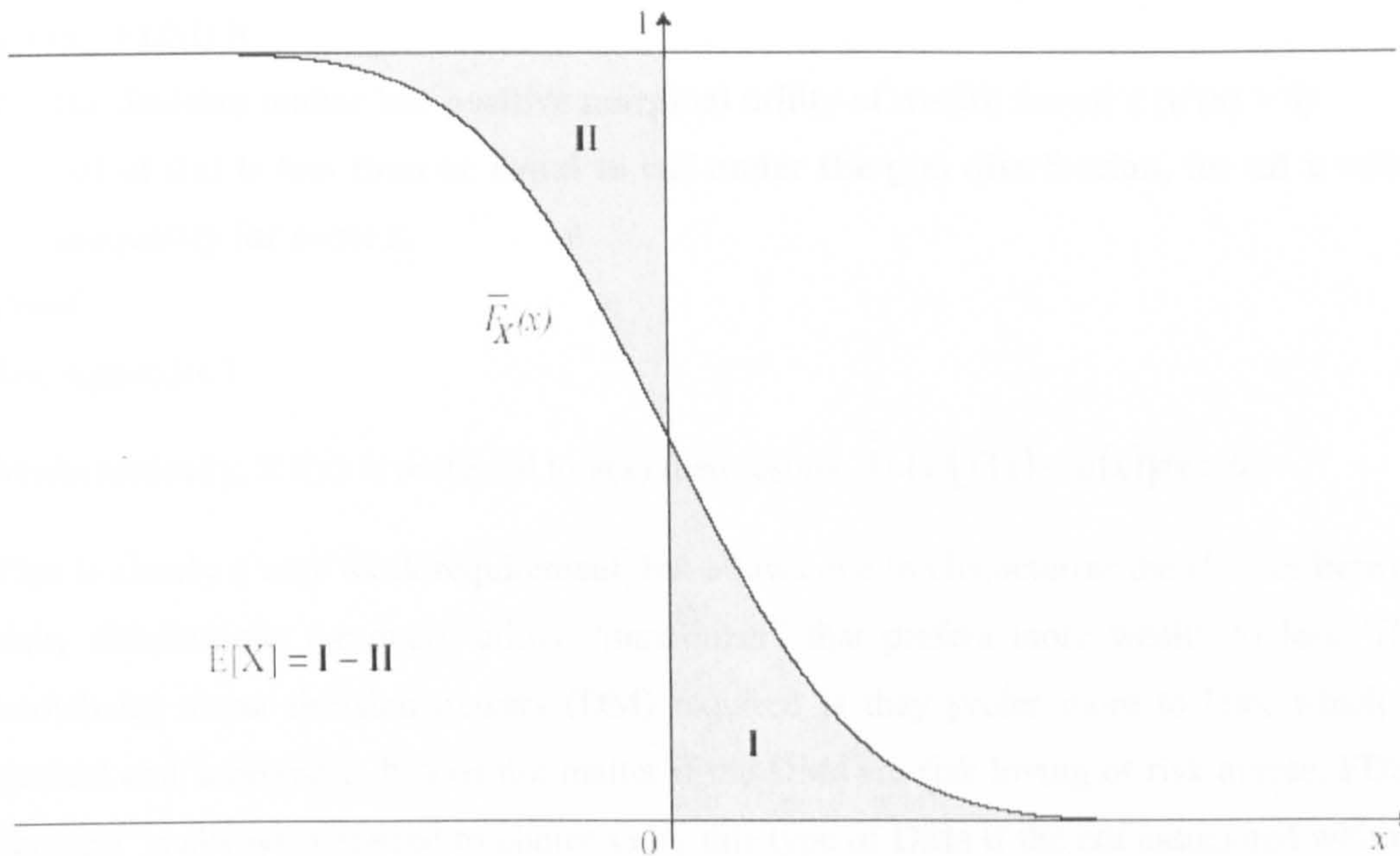


Figure 1 The expected value integral representation

3.1 Stochastic Dominance

Decision rules are all based on known risk preference of a decision maker that has to be quantifiable. However, in reality, the preference is usually not known and difficult to acquire through elicitation. Rather than simply assuming the preference, Stochastic Dominance (SD) rules provide a way to make decisions based on little or limited knowledge of a decision maker's risk preference.

There are three important assumptions at the core of the SD paradigm:

1. Individuals are expected utility maximizers.
2. Two alternatives are to be compared and these are mutually exclusive¹⁴.
3. Analysis is developed based on population probability distributions.

3.1.1 First Degree Stochastic Dominance (FDSD)

As previously observed, SD assumes expected utility of wealth maximization, $\text{Max } E[U(x)]$, with x the level of wealth.

¹³ This integral representation of the expectation of a random variable and its geometric interpretation (see figure 1) is described in "Risk Measures and Optimal Portfolio Selection (with applications to elliptical distributions); J. Dhaenex, E. A. Valdezz and T. Hoedemakers, Samos 2004 Workshop, Risk Measures and Optimal Portfolio Selection, pag 17. Proof of this result when considering a discrete random variable can be found in P. Bremaud (1994) pag. 56 and pag. 76 or in V. Berger (2001) pp. 345-350.

¹⁴ This means that one or the other must be chosen not a convex combination of both.

Proposition 1 The FSD rule

Given two probability distributions $f(x)$ and $g(x)$, distribution $f(x)$ dominates distribution $g(x)$ by FSD if:

1. the decision maker has positive marginal utility of wealth for all x ($u'(x) > 0$)
2. cdf of $f(x)$ is less than or equal to cdf under the $g(x)$ distribution, for all x with strict inequality for some x .

Proof.

See Appendix 1

Mathematically, if $f(x)$ is preferred to $g(x)$ then results $\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx > 0$.

This is clearly a very weak requirement, but allows one to characterize the choices between two risky distributions for every utility “maximizer” that prefers more wealth to less. The only knowledge about decision makers (DM) required is they prefer more to less, which is very general and acceptable. It does not matter if the DMs are risk loving or risk averse. FSD says choice F is always preferred to choice G by this type of DMs if the cdf associated with G is not less than that with F for any outcome level and is greater than it for at least one outcome level¹⁵.

3.1.2 Second Degree Stochastic Dominance (SDSD)

The above SD development while theoretically elegant is not terribly useful. This means that when comparing two return distributions one always outperforms the other. This may not always be the case. The next development in SD involves making an assumption about risk aversion.

Proposition 2 The SDSD rule.

If

- 1) An individual has positive marginal utility (non-satiation), $u'(x) > 0$
- 2) An individual has diminishing marginal utility of income $u''(x) < 0$
- 3) $[F_2(x) - G_2(x)] \leq 0$ for all x , with strict inequality for some x

where $F_2(x)$ and $G_2(x)$ are the second integral of F and G with respect to x , i.e.:

$$F_2(x) = \int_{-\infty}^x \int_{-\infty}^x f(x)dx = \int_{-\infty}^x F(x)dx$$

$$G_2(x) = \int_{-\infty}^x \int_{-\infty}^x g(x)dx = \int_{-\infty}^x G(x)dx$$

with

¹⁵ In many cases only limited information about the behaviour of the investor (for example it is known that only he or she is risk averse and non-satiable) is available. Still, it is possible to determine some conditions in which unambiguously one risky asset will be preferred over another, even if it is not possible to establish a complete order among risky assets.

$$F(x) = \int_{-\infty}^x f(x) dx$$

$$G(x) = \int_{-\infty}^x g(x) dx$$

Then

f(x) dominates g(x) by a SDSD.

Proof

See Appendix 2

Mathematically, if f(x) is preferred to g(x) then results $\int_{-\infty}^{\infty} u'(x)[F(x) - G(x)] dx$

SDSD reveals decision-making principles for DMs who prefer more to less and are risk averse. SDSD says choice F is always preferred to choice G by this type of DMs if the **integral** of cdf associated with G is not less than that with F for any outcome level and is greater than it for at least one outcome level. If the condition for FDSD is satisfied, it must be satisfied for SDSD, not vice versa¹⁶. Therefore, SDSD applies to a wider range of situations. However, again, there are cases that SDSD is not sufficient to lead to a decision¹⁷.

3.1.3 Remarks on Stochastic Dominance

In the more than half-century since modern utility theory was first developed and despite its theoretical appeal and long-standing academic calls for its use, it has been used only very rarely in practice¹⁸. From a pure technical perspective, it is also important to note that:

1. there is a trade off in the strictness between the DMs preference and the outcome distribution. The more one knows about the investors preferences the less information one requires for the distribution. The final aim of the SD rule is to identify an “efficient set” of investment alternatives. If there is only one choice in the efficient set, it is the decision. Nevertheless, usually there are multiple elements in the efficient set, even using high degree SD rules. In this case the bottom line is that there is no single decision to be recommended.

¹⁶ This means the conditions on the distribution of the outcomes is lenient for the second degree comparing to that for the first degree

¹⁷ It is also possible to define rules of decision making based on Higher Degree Stochastic Dominance, such as Third Degree Stochastic Dominance (TDSD). TDSD requires the decision maker to be decreasing absolute risk averse, i.e., DARA Decision-Makers. Whitmore and Hammond made up a third degree stochastic dominance rule by extending this approach once more: the implied decision rules is optimal for every individual who prefers more to less, who exhibits risk aversion and who displays decreasing absolute risk aversion. They find that if one assumes that the first derivative is positive, the second derivative negative and the third derivative positive and that the third integral of the probability function of f is always smaller than that of g(x) then f(x) dominates g(x). The logical intuition is that a risk averter with diminishing absolute risk aversion is found.

¹⁸ See also section 4.3 to investigate reasons explaining why utility-based risk measures have been rarely used in practice.

2. A second difficulty to use SD is that each pair of actions has to be compared for all possible outcomes. This is extremely hard if both decision choice and outcome are continuous.
3. A third disadvantage is there is no welfare measure in SD to compare alternatives.
4. Another important complication is the lack of ability to discriminate among cases with low crossings. SD requires the dominant distribution to always have a greater minimum than the dominated distribution. If the distribution shows a vast improvement under all the observations but the lowest one, then SD will not hold in any form. The real question is how risk adverse individuals are. SD assumes that the individuals fall in the class of all risk averters that includes infinitely risk adverse individuals. It assumes someone can possess a risk aversion parameter that is so large that the utility of the small difference at the lowest observation is extraordinary important. Furthermore it assumes that the alternatives are mutually exclusive. When one performs this type of analysis one ignores the possibility that the alternatives could be diversified. This is perfectly reasonable when dealing with two mutually exclusive alternatives but not with portfolio of assets. A problem in SD involves potential presence of a portfolio.
5. Finally a stochastic dominance selection rule is subject to sampling error. In fact, if distribution means and variances get close together then the probability of improper dominance conclusions can become quite high¹⁹.

Summarising, SD analysis is used to identify conditions under which one risky outcome would be preferable to another. Unfortunately, the question of how much better action F is than action G remains unanswered. Moreover, SD is not a necessary condition. In fact, if an appropriate SD condition is met, it is a sufficient condition for proving that Alternative A is better than Alternative B. However, if a SD condition is not met, Alternative A might still be better than Alternative B. While SD cannot always yield selection of the best alternative, these criteria are easy to apply and can sometimes yield a single optimal solution or a smaller sub-set of solutions without the need to specify a complete utility function.

3.2 Coherence

Another desirable set of attributes a reasonable risk measure should possess have been firstly described by Artzner and al. (ADEH) (1997; 1999a). The set up in (ADEH) is the following: consider a capital market with a finite number K of outcomes or states of the world. Let x and y be k -dimensional vectors representing the possible state-contingent payoffs of two different portfolios, $\rho(x)$ and $\rho(y)$ the portfolios' risk measures, a and b arbitrary constant (with $a > 0$),

¹⁹ Of course, the smaller the sample size the more likely one is to have errors

and r the risk-free interest rate. ADEH argue that any reasonable risk measure should satisfy the following four properties:

$$\text{i. } \rho(x + y) \leq \rho(x) + \rho(y) \quad (\text{sub-additivity}) \quad (3.2.1)$$

$$\text{ii. } \rho(\alpha x) = \alpha \rho(x) \quad \forall \alpha \geq 0 \quad (\text{positive homogeneity}) \quad (3.2.2)$$

$$\text{iii. } \rho(x) \geq \rho(y), \quad \text{if } x \leq y \quad (\text{decreasing monotonicity}) \quad (3.2.3)$$

$$\text{iv. } \rho(x + b) = \rho(x) + b \quad (\text{translation invariance}) \quad (3.2.4)$$

The first property says that the risk measure of an aggregate portfolio must:

1. Be less than or equal to the sum of the risk measures of the smaller portfolios that constitute it
2. Ensure that the risk measure should reflect the impact of hedge or offsets.

If condition (i) is not satisfied, then one can reduce the risk of portfolio by splitting it into two or more parts²⁰. For a sub-additive measure, portfolio diversification always leads to risk reduction, while for measures which violate this axiom diversification may produce an increase in their value even when partial risks are triggered by mutually exclusive events. Therefore, though one can perfectly think of possible alternative axiomatic definitions of risk measure, it is hard to believe that no sensible set of axioms could in any case admit any sub-additivity violations²¹.

The second property says that the risk measure is:

1. Proportional to the scale of the portfolio
2. Independent of scale changes (e.g. currency) in the unit in which the risk is measured. For example, halving the portfolio halves the risk measure.

The first two properties together imply that the risk of a diversified portfolio must be less than or equal to the appropriate weighted average of the risk of the instruments or sub-portfolios that make up the diversified portfolio. Risk measures that do not satisfy these conditions fail to capture the benefits of diversification²².

Property (iii) says that if the portfolio with payoff y has greater losses than (i.e. dominates stochastically) another risk with payoff x , then x is riskier because of the higher loss potential.

²⁰ In a banking context, this would imply that it is possible to reduce the capital requirement by splitting a bank into parts. This condition also permits decentralized calculation of risk, since the sum of the risk measures of sub-portfolios provides a conservative estimate of the risk of the aggregate portfolio

²¹ Sub-additivity is an essential property also in portfolio-optimisation problems. This property in fact is related to the convexity of the risk surface to be minimized in the space of portfolios and the risk minimization process will always pick-up a unique, well-diversified optimal solution. Only if the surfaces are convex they will always be endowed with a unique absolute minimum and no fake local minima

²² For example, if the pay-off of a diversified portfolio is $z = wx + (1 - w)y$, where w and $1 - w$ are the weights of two sub-portfolios with payoffs x and y , then properties (i) and (ii) require that $\rho(z) \leq w\rho(x) + (1 - w)\rho(y)$.

This also means that if $x \leq y$, then each element of y is at least as large as the corresponding element of x .

Property (iv) says that adding a risk-free instrument to a portfolio decreases the risk by the size of the investment in the risk-free instrument. For example, if adopting a regulatory perspective, this means that there is no additional capital requirement for an additional risk for which there is no uncertainty. This property expresses the fact that a portfolio made of sub-portfolios will risk an amount that is at most the sum of the separate amounts risked by its sub-portfolios. This axiom captures the essence of how a risk measure should behave under the composition/addition of portfolios. It is the key test for checking whether a measurement of a portfolio's risk is consistent with those of its parts.

So far, two frameworks for consistently measure the risk of a portfolio, namely the SD and Coherence have been discussed. Aim of the next sections is to analyse possible (and desirable) interdependencies between these frameworks thus providing superior information:

- to regulators, rating agencies and bondholders when evaluating the risk of default and the size of the shortfall
- to shareholders, investors risk and portfolio mnagers when evaluating the risk-return-value profile of the portfolio.

4 Taxonomy of risk measures: an end user's perspective

The global idea about investors is that they can be characterized by their non-satiability, (i.e., investors always prefer more money to less) and their risk behavior (investors can be risk-averse, risk-neutral or even risk-seekers). Unfortunately, in reality, risk is not universally defined, and each investor may approach decision-making under uncertainty with different risk definitions, risk measures and risk risk-reward criteria²³. In this paragraph the most widespread used risk measures in the banking, insurance and asset management industry will be reviewed.

Throughout this section four definitions of risks will be considered:

1. risk as maximum potential loss (given a set of possible states of nature)
2. risk defined as the magnitude of deviation from a target

²³ Risk adjusted performance measurement (RAPM) are examples of risk-reward criteria

3. risk defined as the magnitude of deviation from the expected utility²⁴
4. risk as maximum potential loss defined as the difference between the wealth today and the end-of period wealth

In this section I will show that a correct interpretation of the end user's perspective, on one hand, and the identification of the possible consequences of axioms violations with respect of both the SD and coherent paradigm, on the other hand, can produce more valuable investment decisions and simplify the decision-making process.

Moreover, some interesting technical remarks, namely the problem of ranking investment alternatives and of capital requirements decided by regulators when the underlying probability distribution is non-normal and shows fat and long tail, will be discussed.

4.1 Risk as maximum potential loss

Aim of this section is to define a typology of risk measures which are based on the idea of maximum potential losses (the most popular being the Value at Risk, VaR) that can occur under specific circumstances. The VaR for a portfolio both in the financial and actuarial literature is defined as the p -quantile of the loss distribution. Therefore, for any p in $(0, 1)$, the p -quantile risk measure for a random variable X , which will be denoted by $Q_p(X)$, is defined by

$$Q_p(X) = \inf \{x \in \mathbb{R} \mid F_X(x) \geq p\}, p \in (0, 1)^{25} \quad (4.1.1)$$

The calculation of VaR can be performed in different ways²⁶. Independent of the used calculation method, VaR gives an answer to the following question: given time horizon T and confidence level k , what is the maximum loss L in market value (over the pre-specified time horizon T) that is exceeded only with probability k ? To answer this question the end user has to specify first the time horizon and the confidence level. Basically, this choice is arbitrary. The confidence level depends on the risk aversion of the institution and on the VaR-model performance. The time horizon is determined from the portfolio structure and the possibility to liquidate (or hedge) the risky positions. Under the multivariate normal distribution assumption, it correctly aggregates risk and reflects the benefit of diversification²⁷.

²⁴ The distinction between two-sided and one-sided measurement of risk will be described in the next section.

²⁵ The quantile function $Q_p(X)$ is a non-decreasing and left-continuous function of p .

²⁶ There are two main methods that can be used to estimate the value at Risk: the parametric (variance-covariance approach) and non-parametric approach (historical and Monte Carlo simulation). The advantages and disadvantages of the various methods must be carefully considered

²⁷ A risk measure is pseudo-coherent (denoted by ρ) if it has the homogeneity and the risk-free condition properties, but not the sub-additivity property; the VaR is one example of a pseudo-coherent risk measure. Value-at-risk is a

Apart from all the glory reserved to this risk measure in the last ten years it is worthwhile to highlight that Value-at-Risk satisfies (ii), (iii), and (iv) coherence properties, but is not a coherent risk measure because it fails to satisfy the axiom of sub-additivity²⁸. Moreover it fails to satisfy also the SDD principle because this type of risk measure is calculated by using only tail loss distribution information thus disregarding all other information incorporated in the remaining part of the loss distribution²⁹.

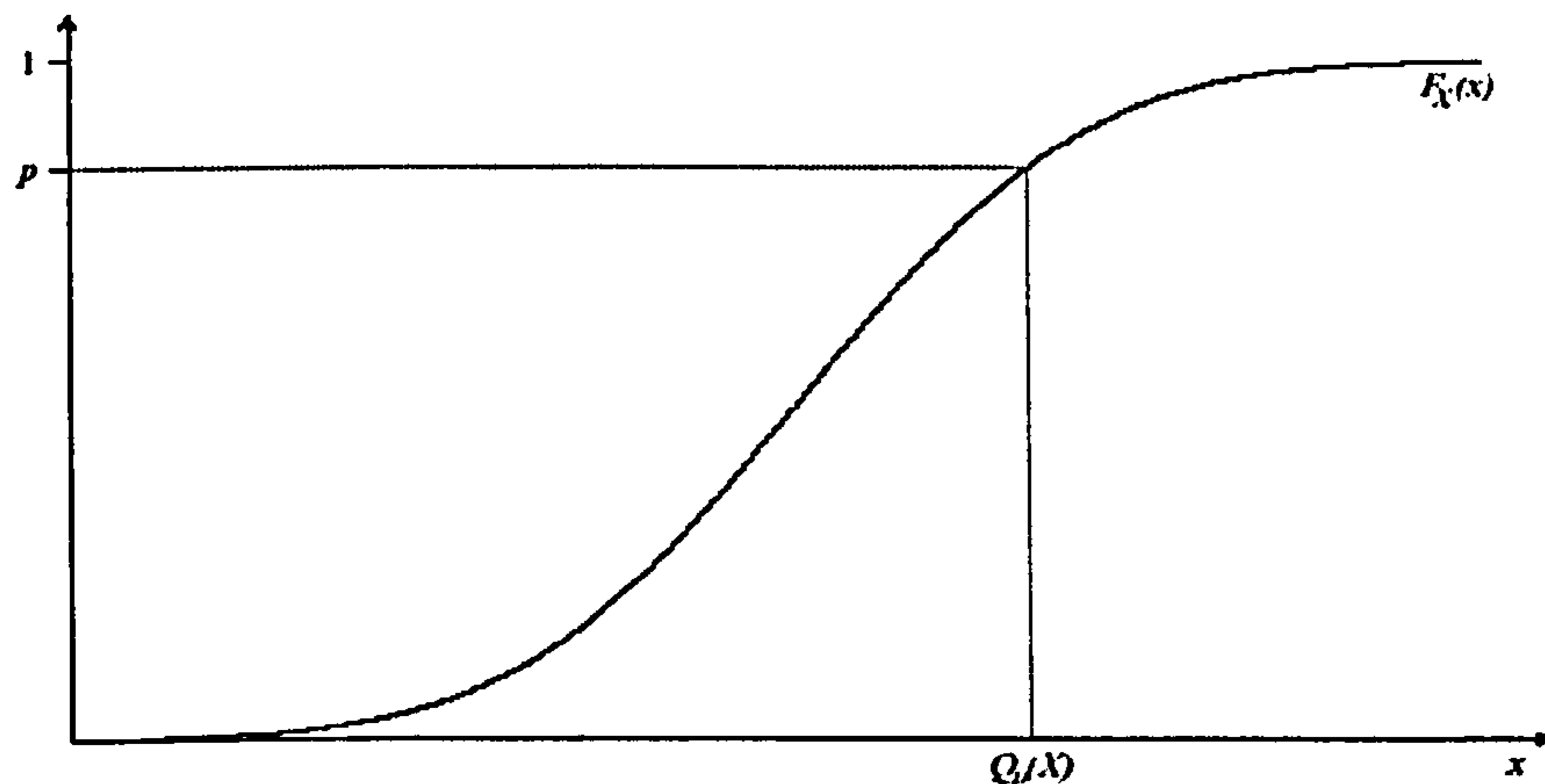


Figure 2 Quantile Risk Measure

A closer inspection of the property of this risk measure may help in understanding why VaR can be used for capital requirements purposes.

Regulators, bondholders and rating agencies care is to calculate the capital requirement with respect to a random loss with the aim of avoiding insolvency³⁰, like the VaR measure³¹. Strictly speaking, for this class of market participants, therefore, the only worry is downside risk and not active portfolio risk management³².

Unfortunately VaR is a risk measure that only concerns about the frequency of default and not the size of default. If X denotes the aggregate claims of a generic portfolio of a financial institution over a given reference period and P denotes the aggregate premium (or the provision) for this portfolio, then $K = Q_p(X) - P$ is the smallest “additional capital” required such that the insurer, broadly speaking, becomes technically insolvent, i.e. $X > P + K$, with a (small)

coherent risk measures if the possible change in portfolio value is described by the normal distribution, but not generally

²⁸ For a detailed analysis of the sub-additivity broken axiom violation consequences see Dowd K.(2002) pp.27-28.

²⁹ See section 4.2.5.4 for an illustration of this result.

³⁰ This quantile reserving principle is used in practice in the insurance industry and it is a well known concept for actuaries

³¹ For a risk which may result in a gain or a loss, it is appropriate when pricing the risk to take in consideration the potential gains. In capital adequacy, this can lead to unacceptable results - for example, a negative capital requirement. The portfolio management perspective will be analysed in more detail in section 7.

³² Strictly speaking, for them risk is associated only to adverse event

probability of at most $1 - p$. Using the p -quantile risk measure for determining a solvency capital is meaningful in situations where the default event should be avoided, but the size of the shortfall is not important. Loosely speaking, the size of shortfall is more important for the bondholders than for the rating agencies.

From a shareholders and bondholders perspective one might think that quantile VaR at the company level is a meaningful risk measure since the default event is of primary concern, and the size of shortfall is only of secondary importance. But from a proactive risk and portfolio management perspective these types of “truncated” risk measure do not provide any sensible insights simply because they are built in a manner that will not allow different risk profile of scenarios and states of the nature that can be faced behind the considered quantile or, stated alternatively, behind the truncation value. After this value all states of the world are assumed to be equally likely and worth nothing!

The same argument can be considered for all risk adjusted performance measurement (RAPM) built on top of this quantile risk measure. The same criticism therefore applies also to the asset management industry when the portfolio manager uses VaR as a discriminating tool (or technique) to select among different alternative investment strategy. In theory, if the asset manager can avoid to care about the tail risk behind the quantile value and the shape of the underlying portfolio distribution is normal, VaR can still provide interesting insights. In practice, as it will be described in section 6, portfolio manager typically use RAPM based on the idea that the risk should be measured as “deviation from a target”. This topic is introduced in the next section.

Summarising, a single quantile risk measure of a predetermined level p does not give any information about the thickness of the upper tail of the distribution function. For example, a regulator is not only concerned about the frequency of default, but also about the severity of default. Furthermore, shareholders and management should be concerned about the question “how bad is bad?” when they want to evaluate the risks at hand in a consistent way. Best practices in both the banking and insurance industry often uses another risk measure the so called the Tail Value-at-Risk ($TVaR_p$) at level p . This risk measure will be discussed in section 4.2.5.

*4.2 Risk as the magnitude of deviation from a target: two sided and one sided risk measures*³³

Two sided risk measures calculate the magnitude of the distance (in both directions) from the realizations of a random outcome x to its expected value $E(x)$ ³⁴. Looking for example at quadratic deviations (volatility) this leads to the risk measure variance (or standard deviation respectively by taking the root). Both risk measures have been the traditional risk measures in economics and finance since the pioneering work of Markowitz. These risk measures exhibit a number of nice technical properties. For instance, the variance of a portfolio return is the sum of the variances and covariances of the individual returns. Furthermore, the variance is used as standard optimisation function³⁵ aimed at nicely solving the portfolio management problem. These properties are particular helpful in solving the normative selection problem, since the optimal portfolio is not yet known.

Two sided risk measure contradicts the notion of risk that only negative deviations are dangerous, it is downside risk that in this framework matters. In addition, variance does not account for fat tails of the underlying distribution and for the corresponding tail risk. This leads to the proposition to include higher (normalised) central moments (as e.g. the third and the fourth moment, respectively, skewness and kurtosis) into the analysis to assess the risk more properly. As the devil's advocate, Markowitz (1959) has recognized and stressed the limitations of mean-variance analysis.

³³ This section is based on the following research papers: Huang Linzebrgh (1988); Ingersoll (1987); A. Plantinga and S. de Groot (2001); Benninga, S. Z., 1997, 'Financial modeling', MIT Press, Cambridge, MA.; Elton, Edwin J. and Martin J. Gruber, 1995, Modern portfolio theory and investment analysis, John Wiley & Sons; Pratt, John W., 1964, "Risk aversion in the small and the large", *Econometrica*, Vol. 32, pp. 122-136; Sharpe, William F., 1966, "Mutual fund performance", *Journal of Business*, No. 1, Vol. 2, pp. 119-138.; Sharpe, William F., 1994, "The Sharpe ratio", *Journal of Portfolio Management*, Vol. 21, No. 1, pp. 49-58; Sortino, Frank A. and Robert van der Meer, 1991, "Downside risk", *Journal of Portfolio Management*, Vol. 18, summer, pp. 27-31; Sortino, Frank A., Robert van der Meer, and Auke Plantinga, 1999, "The Dutch triangle", *Journal of Portfolio Management*, Vol. 26, No. 1, pp. 50-59; Sortino, Frank A., and Lee N. Price, 1994, "Performance measurement in a downside risk framework", *Journal of Investing*, Vol. 3, No. 3; Tversky, A., and D. Kahneman, 1992, 'Advances in prospect theory: cumulative representation of uncertainty', *Journal of risk and uncertainty*, pp. 297-323; Optimal Algorithms And Lower Partial Moment: Ex-Post Results by David N. Nawrocki, College of Commerce and Finance, Villanova University; Two paradigms and nobel prizes in economics: A contradiction or coexistence?, By Haim Levy, Enrico De Giorgi and Thorsten Hens, working paper, June 2003; Sharpe, W.F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425-442, Sharpe W. F. (1991), "Capital Asset Prices With and Without Negative Holdings," *The Journal of Finance*, 46, 2, 489-509; Decomposing Portfolio Value-at-Risk: A General Analysis Winfried G. Hallerbach, *The Journal of Risk* 5/2, Febr. 2003; Jorion, Ph., 2001, "Value at Risk: The Benchmark for Controlling Market Risk", McGraw-Hill, Chicago Ill., 2nd ed. (1st ed. 1997); Litterman, R., 1997a, "Hot Spots and Hedges (I)", *RISK* 10/3, March, pp. 42-45; Litterman, R., 1997b, "Hot Spots and Hedges (II)", *RISK* 10/5, May, pp. 38-42.

³⁴ In the following, without losing generality, the expected value will be considered as the relevant target for the end-user.

³⁵ Such as, for example, the quadratic optimisation functions.

From a general perspective, the variance as a risk measure may miss its link with an investor's preference structure or with the distributions of security returns and portfolio returns. Information concerning mean and variance then is not sufficient to adequately discriminate between return distributions. For example, typically a risk manager or an asset manager is interested in characteristics of distributions over the whole range of the returns (full domain)³⁶. These characteristics can be captured by full domain distribution measures surpassing the second statistical moment. In the resulting general setting of multi-moment portfolio analysis, the investor, in theory, should be able to specify preferences for the first four moments. In practice, this is a not easy exercise.

Full domain (risk) measures can be contrasted with partial domain measures, which provide information for some distribution over some part of its domain. Of special relevance is “downside risk”, focusing on returns falling below some critical level. If the end user only care about downside risk, then this risk measure can provide very useful and reliable information. The global idea about downside risk is that the left-hand side of a return distribution involves risk while the right-hand side contains the better investment opportunities.

A measure of shortfall risk is one-sided risk measure and measures the shortfall risk relative to a target variable. This may be the expected value, as assumed so far but, in general it can be an arbitrary (fixed or variable) deterministic target τ ³⁷ or even a stochastic benchmark $\tilde{\tau}$. A more general (and more sophisticated) approach to downside risk specifies risk in terms of probability-weighted functions of deviations below some target return. One example is the semi-variance, introduced in Markowitz (1959)³⁸ and which measures the variability of returns below the mean. The semi-variance is a special case of the more general **lower partial moments (LPM)** risk measure, which form the partial domain analogous of variance and higher moments³⁹. Seminal references are Bawa (1975) and Fishburn (1977). These authors introduced a general definition of downside risk in the form of LPM and developed the “ (α, τ) - model”, respectively. The LPM of order α around τ ⁴⁰ with respect to the investment return X , can be interpreted, in a statistical

³⁶ Return distributions may exhibit asymmetry or other form characteristics different from normal distributions

³⁷ Target gain, target return, minimal accepted return to give some appropriate examples.

³⁸ Markowitz pointed out: “Since an investor worries about underperformance rather than over-performance, semi-deviation is a more appropriate measure of investor's risk than variance”. Sharpe (1963): “Under certain conditions the mean-variance approach leads to unsatisfactory predictions of investor behavior”

³⁹ See Harlow and Rao (1989) for recent references in this area.

⁴⁰ In practice, a benchmark return, a short term interest rate or a required return given some liabilities (minimal acceptable return) can serve as the target return. The parameter α reflects the investor's feeling about the relative consequences of falling short of τ by various amounts. Fishburn (1977) has shown that $\alpha=1$ (which suits a risk-neutral investor) separates risk-seeking ($0 < \alpha < 1$) from risk-averse behaviour ($\alpha > 1$) with regard to returns below the target τ . In practice, values of α ranging from less than 1 to greater than 4 were observed by Fishburn. By changing the parameters α and τ most downside measures used in practice can be formed. For example, as shown later in the paragraph, $\alpha=2$ yields the target semi-variance, while adding the restrictions that $\tau = E(X)$ produces the semi-variance (or lower partial variance).

sense, as non-central moments of an underlying distribution, where the interest is focused only on the lower part of the whole distribution. How much of the distribution is relevant for the computation of the LPM depends on an ex ante defined target return τ . Formally, this relation can be specified, for a continuous random variable X , as follows:

$$LPM_{\alpha}(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^{\alpha} f(X) dX = \int_{-\infty}^{\tau} (\tau - X)^{\alpha} dF(X), \quad (4.2.1)$$

or in normalised form

$$\alpha \geq 2 \quad R(X) = LPM_{\alpha}(\tau, X)^{\frac{1}{\alpha}} \quad (4.2.2)$$

where:

$F(X)$ is the cdf of the investment return X .

$f(x)$ is the pdf to quantify the probability of X taking on a certain value x .

It is remarkable that the relevant area to compute any LPM is limited by the target return τ , which defines the upper bound of the integral. An important characteristic lies in the choice of the variable α , which defines the order of the LPM. Basic probabilistic measures, within the previous described set-up, playing an important role in applications, are obtained for $\alpha = 0, 1$ and 2 and corresponding return density functions⁴¹.

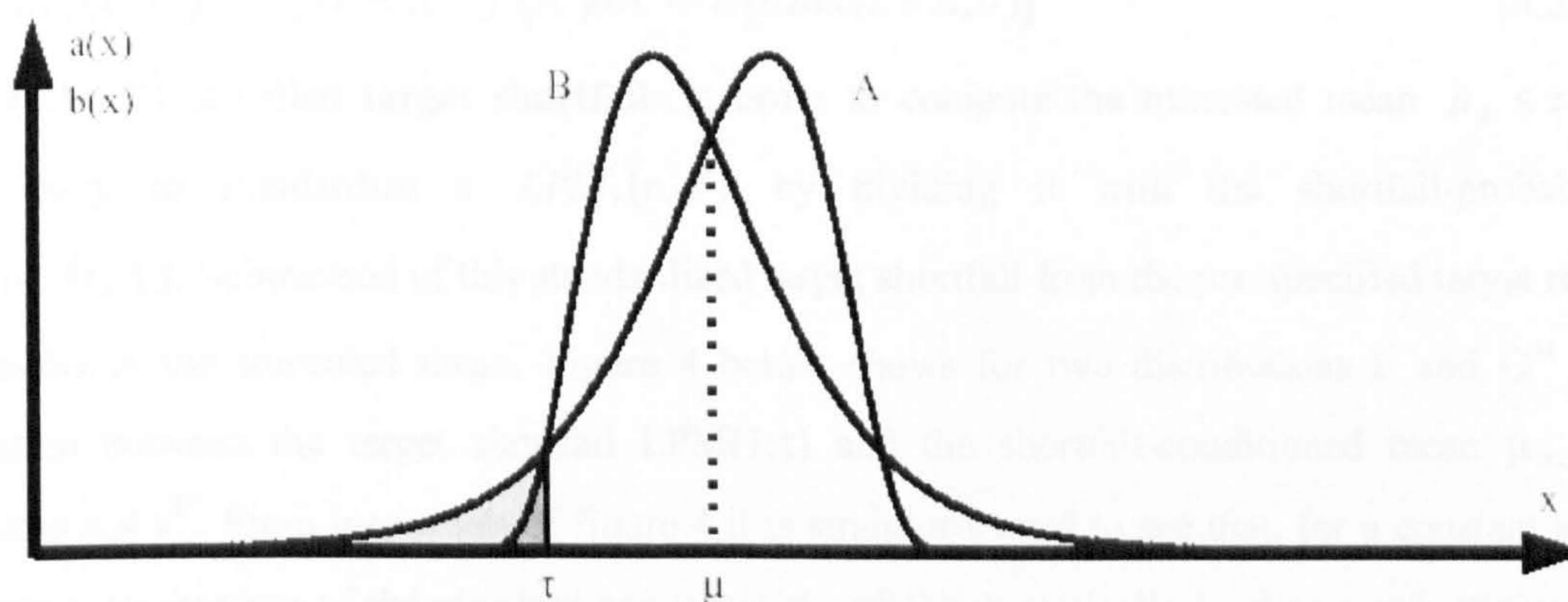


Figure 3 Shortfall-probabilities under a target return τ for two portfolios A and B with their variations as illustrated in the following sections.

4.2.1 LPM when $\alpha = 0$ and the expected value is the target value τ

For the case of $\alpha = 0$ the extent of a shortfall relative to τ remains irrelevant

$$LPM_0(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^0 f(X) dX = \int_{-\infty}^{\tau} f(X) dX = [F(X)]_{-\infty}^{\tau} = F(\tau) - F(-\infty) = F(\tau) \quad (4.2.3)$$

Any $LPM_0(\tau, X)$ is simply the part of the whole distribution *below* the pre-specified target return τ and is therefore commonly cited as **shortfall probability**. Figure 3 illustrates the

⁴¹ This graph is quoted from A. Peter, Risk Measures, Working paper, University of Mannheim, Jan 2003

shortfall-probabilities under a target return τ for two portfolios A and B with their corresponding return density functions. The two distributions are calibrated in order to receive identical means and variances respectively⁴².

Portfolio selection techniques based on the ideas of Markowitz will consider A and B equally desirable. But a $LPM_0(\tau, X)$ -based investor will undoubtedly prefer the right-skewed portfolio B, which has a much better protection against low returns without capping the probability of high gains. The return behaviour of portfolio B is represented by the log-normally distributed curve B, whereas portfolio A is described by the density curve A, which is just a reflection of curve B at its mean μ . Both A and B possess the same mean and variance. Therefore a μ - σ -portfolio selection in the sense of Markowitz is of no further help. Nevertheless, it seems obvious that an investor will prefer portfolio B with the smaller shortfall probability relative to τ ⁴³.

4.2.2 LPM when $\alpha = 1$ and the expected value is the is the target value τ

The first-order LPM computes the shortfall distance relative to the target return by integrating the probability-weighted differences between τ and every X with $X \in [-\infty, \tau]$.

$$LPM_1(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^1 f(X) dX = E[\text{Max}(\tau - X, 0)] \quad (4.2.2.1)$$

$LPM_1(\tau, X)$ is called **target shortfall**. In order to compute the truncated mean $\mu_{X \leq \tau}$ it is necessary to standardize a $LPM_1(\tau, X)$ by dividing it with the shortfall-probability $LPM_0(\tau, X)$. Subtraction of this standardized target shortfall from the pre-specified target return τ results in the truncated mean. Figure 4 below shows for two distributions F and G⁴⁴ the relation between the target shortfall $LPM(1;\tau)$ and the shortfall-conditioned mean $\mu_{x \leq \tau}$ for returns $x \leq \tau$ ⁴⁵. From inspection of figure 4 it is straightforward to see that, for a constant target return τ , an increase of the standardized target shortfall automatically implies a reduction of the shortfall conditioned mean.

⁴² The shortfall-probabilities in figure 3 are illustrated by the two differently shaded areas under the corresponding density functions.

⁴³ In section 7 other portfolio strategies based on downside risk measure will be analysed.

⁴⁴ Figure assumes the target return is the same for both distribution.

⁴⁵ It is important to note that $LPM(1;\tau)$ has to be divided by $LPM(0;\tau)$ in order to get a measure which is subtractable from τ and therefore calculate the truncated mean $\mu_{x \leq \tau}$

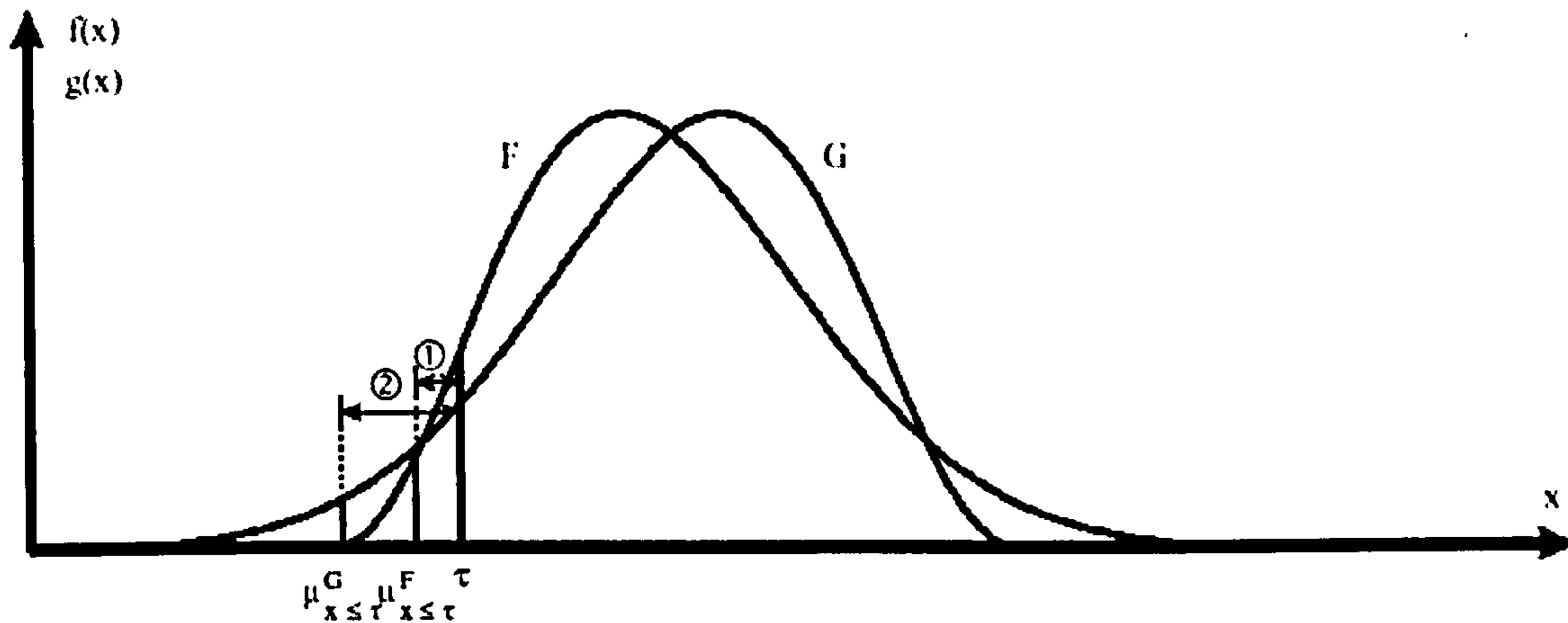


Figure 4 Relation between the target shortfall $LPM(1;\tau)$ and the shortfall-conditioned mean $\mu_{x \leq \tau}$ for returns $x \leq \tau$ two distributions F and G ⁴⁶

To see this, by construction, it is possible to show that the standardised target shortfalls are related through the following relationships:

$$\tau \mu_{x \leq \tau} = \tau - \frac{\int_{-\infty}^{\tau} (\tau - X) f(X) dX}{\int_{-\infty}^{\tau} f(X) dX}$$

$$\tau \mu_{x \leq \tau}^F = \frac{LPM_{1,F}(\tau, X)}{LPM_{0,F}(\tau, X)} < \frac{LPM_{1,G}(\tau, X)}{LPM_{0,G}(\tau, X)} = \tau \mu_{x \leq \tau}^G$$

Consequently, the relationship between downside $LPM_0(X, \tau)$ and maximum potential loss VaR based risk measure becomes:

$$VaR_K = L = W(1 - e^{-\tau}) = W(1 - e^{-F^{-1}(k)}) = W(1 - e^{-F^{-1}(LPM_0(X, \tau))}) \quad (4.2.2.2)$$

This means that a maximum potential loss risk measure such as the VaR can be calculated on the basis of a downside risk measure, such $LPM_0(X, \tau)$ and vice versa⁴⁷.

4.2.3 LPM when $\alpha = 2$ and the expected value is the target value τ

$LPM_2(\tau, X)$ Can be calculated by integrating the probability-weighted and squared downside deviations between τ and every X with $X \in [-\infty, \tau]$:

$$LPM_2(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^2 f(X) dX = E[\text{Max}(\tau - X, 0)^2] \quad (4.2.3.1)$$

This measure is therefore called **target variance** or - by taking its quadratic root - target standard deviation. As in the case of the classical variance, $LPM_2(\tau, X)$ overweighs realizations of X

⁴⁶ This graph is quoted from A. Peter, Risk Measures, Working paper, University of Mannheim Jan 2003

⁴⁷ The proof is in appendix 3

which are far below the target return⁴⁸. The asymmetrical LPM for $\alpha > 2$ can be derived analogous to the terms of the higher statistical moments⁴⁹. It is straightforward to show that the degree of risk aversion corresponds directly with the chosen order α of the LPM-measure.

4.2.4 LPM when $\alpha = 1, 2$ and the target value is the expected value $\tau = E(X)$

$LPM_1(E(X), X)$, also called **Lower semi-absolute deviation** is a generalisation of eq. (4.2.2.1) and can be calculated as follows:

$$LPM_1(E(X), X) = \int_{-\infty}^{\tau} (E(X) - X)^1 f(X) dX = E[\text{Max}(E(X) - X, 0)] \quad (4.2.4.1)$$

$LPM_2(E(X), X)$, also called **Semivariance deviation**, is a generalisation of eq. (4.2.3.1) and can be calculated as follows:

$$LPM_2(E(X), X) = \int_{-\infty}^{\tau} (E(X) - X)^2 f(X) dX = E[\text{Max}(E(X) - X, 0)^2] \quad (4.2.4.2)$$

A particular case of the Semivariance is the **Semistandard deviation**, or alternatively called target semi-deviation. Semivariance can be calculated by taking the square root of eq. (4.2.4.2):

$$\{LPM_2(E(X), X)\}^{0.5} = \left\{ \int_{-\infty}^{\tau} (E(X) - X)^2 f(X) dX \right\}^{0.5} = E[\text{Max}(E(X) - X, 0)^2]^{0.5} \quad (4.2.4.3)$$

$\{LPM_2(E(X), X)\}^{0.5}$ measures the standard deviation of below-target returns and therefore it only penalizes those events that are below the target. As all others previously described downside risk measure, it is an asymmetric risk measure.

The example in table 1 illustrates the behaviour of the LPM measure for different moment degrees (α). The target return is set to 15% for the two investments A and B⁵⁰. Note that:

- when $\alpha < 1$, the Investment A is considered to be less risky than Investment B, although the skewness number and the distribution information indicates that Investment B has less downside risk. Note that this is consistent with a risk-loving utility function.
- When $\alpha = 1$ the two LPM values are equal.
- When $\alpha > 1$, then Investment B is considered to be less risky than Investment A. Note that this is consistent with a risk-averse utility function. Also, as α increases, Investment A takes on a heavier risk penalty. In particular:
 - When $\alpha = 1.5$ Investment A is twice as risky as Investment B.
 - When $\alpha = 2$ Investment A is four times as risky as Investment B.

⁴⁸ It has to be emphasized that the deviations of X are neither computed relative to the first central moment μ of the underlying distribution (like the variance) nor to any truncated mean $\mu_{x \leq \tau}$, but to the target return τ

⁴⁹ These risk measures are termed as target skewness and target kurtosis

⁵⁰ Which for this example is the same as the mean return. Not necessarily the target return must be equal to the mean return

Chapter 2

- When $\alpha=3$ Investment A is sixteen times as risky as Investment B.

If τ is set to some other return, then the LPM values will depend on the degree of skewness in the return distribution. This demonstrates the importance of setting the correct values of α and τ when using the LPM downside risk measures.

	Company A		Company B	
	Return	Prob.	Return	Prob.
	-5.00	0.20	10.00	0.80
	20.00	0.80	35.00	0.20
Mean Return	15.00		15.00	
Variance	100.00		100.00	
Skewness	-1.50		1.50	
LPM $\alpha=0.0$ $\tau=15$	0.20		0.80	
LPM $\alpha=0.5$ $\tau=15$	0.89		1.79	
LPM $\alpha=1.0$ $\tau=15$	4.00		4.00	
LPM $\alpha=1.5$ $\tau=15$	17.89		8.94	
LPM $\alpha=2.0$ $\tau=15$	80.00		20.00	
LPM $\alpha=3.0$ $\tau=15$	1600.00		100.00	

Table 1 - Example of Degrees of the Lower Partial Moment.

Summarising, in order to provide investment advice, the use of an appropriate risk measure is imperative. Following the downside risk measure framework, the most important investor behaviour/dynamics affecting the choice of the risk measure are:

- Investors perceive risk in terms of below target returns.
- Investor risk aversion increases with the magnitude of the probability of ruinous losses.
- Investors are not static.

As the investor's expectations, total wealth, and investment horizon changes, the investor's below target return risk aversion changes. Consequently, in theory, investors need to be constantly monitored for changes in their level of risk aversion. This is a not easy and expensive exercise.

4.2.5 Conditional measures of shortfall risk

In this section the main measures of shortfall risk are described, namely: the Conditional Tail Expectation (CTE), Tail VaR (TVaR) and Expected Shortfall (ES). Moreover their main strengths and weaknesses are analysed.

4.2.5.1 Conditional Tail Expectation (CTE)

One of the most widespread measures of risk used in the insurance industry (and in banking industry as well, in the area of credit risk portfolio management) is the so called Conditional Tail

Expectation (CTE). CTE_p is defined for smooth distribution functions, given the parameter p , with $0 < p < 1$, as follows:

$$CTE_p(X) = E[X | X > F_x^{-1}(p)] \quad (4.2.5.1.1)$$

where $F_x^{-1}(p)$ is the inverse distribution function of the loss random variable, X . That is, $F_x^{-1}(p)$ is the 100α percentile of the loss distribution⁵¹.

4.2.5.2 Tail Value at Risk (TVaR_p(X)),

TVaR_p is the arithmetic average of the quantiles of X from p on and is defined as

$$TVaR_p(X) = \frac{1}{1-p} \int_p^1 Q_q(X) dq, \quad p \in (0,1) \quad (4.2.5.2.1)$$

Note that the TVaR_p is always larger than the corresponding quantile $Q_p(X)$. From (4.2.5.2.1) it follows immediately that the TVaR_p is a non-decreasing function of p . Let X again denote the aggregate claims of a financial institution portfolio over a given reference period and P the aggregate premium (or the provision) for this portfolio. Setting the amount of “additional capital” equal to $TVaR_p(X) - P$, it is possible to define “bad times” as those where X takes a value in the interval $[Q_p(X), TVaR_p(X)]$. Hence, “bad times” are those where the aggregate claims exceed the threshold $Q_p(X)$, but not using up all available capital. The width of the interval is a “cushion” that is used in case of “bad times”.

4.2.5.3 Mean Excess Loss (MEL) or Expected Shortfall (ES)

Another widespread used measure of risk in the banking industry when measuring credit risk is the so called Mean Excess Loss (MEL) or the Expected Shortfall (ES). ES measures the average shortfall under the condition a shortfall occurs. In practice ES can be considered as a kind of (average) worst case scenario. ES at level p will be denoted by $ES_p(X)$ (or $MEL_p(X)$) and is defined, for a loss random variable X , as

$$MEL_p(X) = ES_p(X) = E[(X - Q_p(X))_+], \quad p \in (0, 1) \quad (4.2.5.3.1)$$

Example

For example, let assume X is standard normally randomly distributed. Following example 3.9.1 in Kaas, Goovaerts, Dhaene and Denuit (2001), where it is shown how to calculate the Quantile VaR, CTE, TVaR and ES when X is standard normally distributed,

if

Φ denotes the standard normal cumulative distribution function

Φ' denotes the density function of the standard normal distribution⁵²

⁵¹ Loosely speaking, $CTE_\alpha(X)$ is equal to the mean of the top $(1-\alpha)\%$ losses.

Φ^{-1} denotes the inverse of the standard normal distribution function

$\Phi^{-1}(p)=z$ if the mean is 0 and the standard deviation is 1

The Quantile VaR $Q_p(X)$, the CTE_p TVaR_p ES_p at 95% (p=5%) confidence interval results

$$Q_p(X) = \mu + \sigma(\Phi^{-1}(p)) = 1.644 \quad (4.2.5.3.2)$$

$$ESF_p(X) = \sigma\Phi'(\Phi^{-1}(p)) - \sigma(\Phi^{-1}(p))(1-p) = 0.02 \quad (4.2.5.3.3)$$

$$CTE_p(X) = \mu + \sigma \frac{\Phi'(\Phi^{-1}(p))}{(1-p)} = 2.062 \quad (4.2.5.3.4)$$

$$TVaR_p = \mu + \sigma(\Phi^{-1}(p)) + \frac{\sigma\Phi'(\Phi^{-1}(p)) - \sigma(\Phi^{-1}(p))(1-p)}{1-p} = 2.062 \quad (4.2.5.3.5)$$

From inspection of eq. (4.2.5.3.4) and (4.2.5.3.5) it is easy to show that if the mean is 0 and the variance is 1 results that CTE is equal to TVaR.

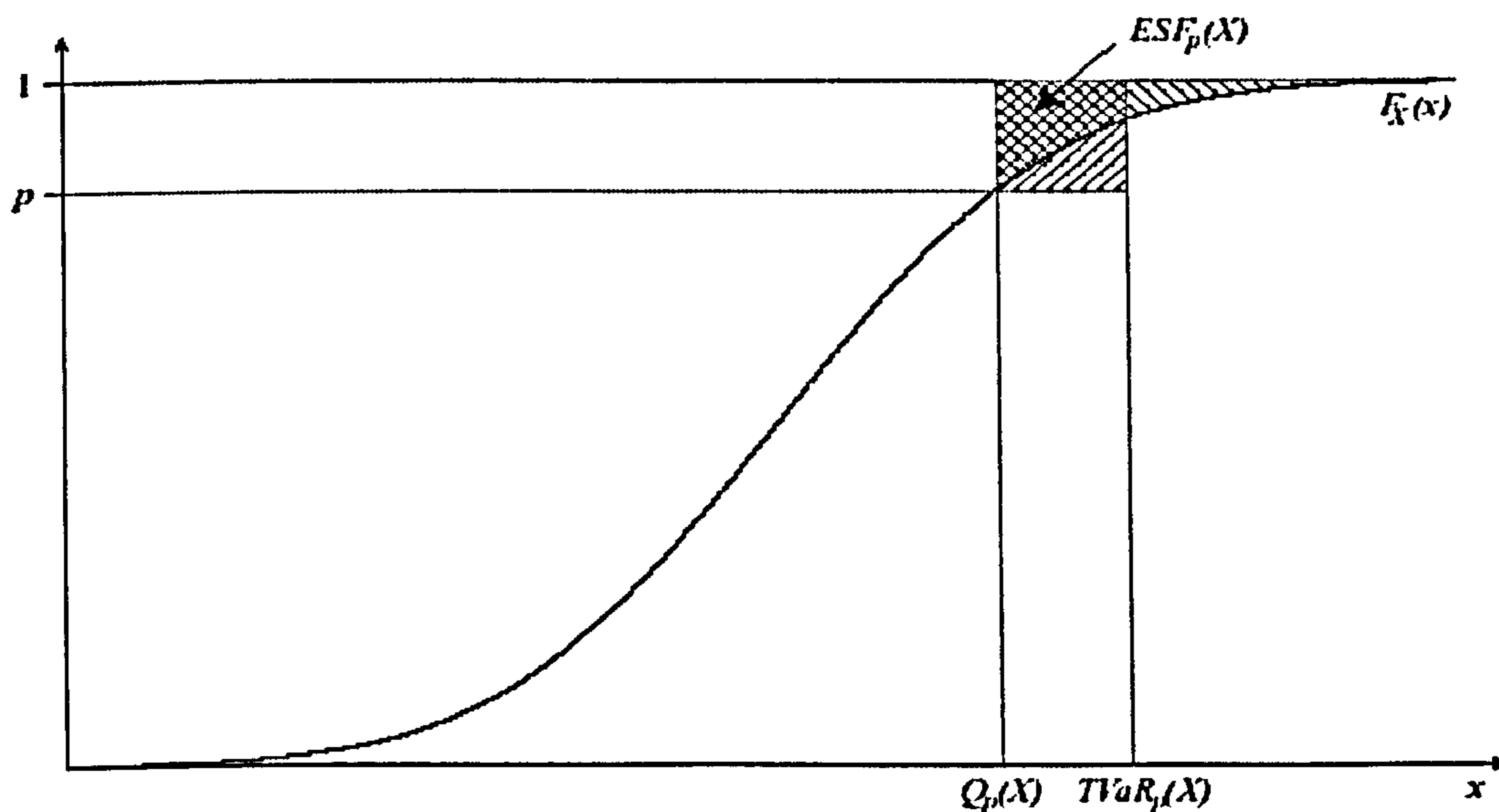


Figure 5 Expected Shortfall, Quantile and Tail VaR

4.2.5.4 Remarks

In this section some interesting relationships and risk measure properties are analysed:

1. Dhaenex, Valdezz and Hoedemakers (2004)⁵³ showed that the relations between Quantiles, TVaR, CTE and ESF can be summarised as follows:

⁵² $\Phi' = \phi(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2}}$ see also footnote 13

⁵³ See footnote 11

$$\text{TVaR}_p(X) = Q_p(X) + \frac{1}{1-p} \text{ESF}_p(X) \quad p \in (0,1) \quad (4.2.5.4.1)$$

$$\text{CTE}_p(X) = Q_p(X) + \frac{1}{1-F_X(Q_p(X))} \text{ESF}_p(X) \quad p \in (0,1) \quad (4.2.5.4.2)$$

$$\text{CTE}_p(X) = \text{TVaR}_{F_X[Q_p(X)]}(X) \quad p \in (0,1) \quad (4.2.5.4.3)$$

2. CTE and ESF are not sub-additive.
3. TVaR is sub-additive: in fact results $\text{TVaR}_p(X+Y) \leq \text{TVaR}_p(X) + \text{TVaR}_p(Y)$, $p \in (0, 1)$.
4. Despite the advantage of corresponding closer to an intuitive notion of risk, shortfall measures have the disadvantage that they lead to greater technical problems with respect to the disaggregation of portfolio risk, optimisation procedure and parameter statistical identification and estimation as will be discussed in detail in the next sections.
5. These type of measures ignore losses below the quantile being considered. As a result their use in portfolio management may lead to sub-optimal decisions, the main reason being that it doesn't comply with the SDST principle as the following example it will help to clarify⁵⁴. Consider the following two loss distribution X and Y:

$$X = \begin{cases} 0 & \text{with probability 0.95} \\ 50 & \text{with probability 0.025} \\ 100 & \text{with probability 0.025} \end{cases} \quad Y = \begin{cases} 50 & \text{with probability 0.975} \\ 100 & \text{with probability 0.025} \end{cases}$$

Clearly Y has the riskier distribution, and clearly $X \prec_{2nd} Y$. The CTE with parameter 97.5% is the expected value of the loss, given the loss lies in the upper 2.5% of the distribution. In both these cases, the CTE (97.5%) risk measure is 75. In fact, for any parameter $\alpha \geq 0.975$ the CTE of these two risks will be equal. The CTE risk measure cannot distinguish between these risks in general.

It is therefore argued that to use a risk measure for portfolio management purposes it is crucial that it should go beyond coherence. In fact, as previously described, although being coherent, CTE ignores useful information in a large part of the loss distribution, and consequently lack incentive for mitigating losses below the quantile VaR. The problem is that CTE doesn't properly adjust for extreme low-frequency and high severity losses, since it only accounts for the mean shortfall and not higher moments, as illustrated in the previous example. Same remark applies for all others downside risk measure

A measure aimed at providing useful information for asset allocation and risk management should go beyond coherence and accomplish with the stochastic dominance principle one and

⁵⁴ The example is draw from Wirch and Hardy (2001), working paper, "Ordering of Risk Measures for Capital Adequacy".

two. Section 5 will introduce another class of risk measure, namely the **distortion risk measure**, that are able to comply with both the coherence and the SD framework and, consequently, they are more appropriate when active risk management is the main purpose of the analysis.

4.3 Utility theory based risk measures⁵⁵

As expected utility theory is one of the standard theory for decisions under risk it will be interesting to investigate the relationship between investor preference functions and measures of risk built by using specific classes of utility functions. Formally, the corresponding preference function is of the form $risk(x) = -E[U(X)]$ where $U(X)$ denotes the utility functions specific to every investor. Risk therefore corresponds to the negative expected utility of the transformed random variable $X - E(X)$. Specific risk measures are then obtained for specific utility functions. By using the same framework illustrated in the previous paragraphs, it is possible to define the following utility-based risk measures:

- magnitude of deviation from the expected utility
- risk as maximum potential loss defined as the difference between the wealth today and the end-of period wealth

These classes of utility-based risk measures will be used when building risk adjusted performance measurement based on some arbitrary utility function which relates the reward value and the risk measure utility function such as the Generalised Sharpe Ratio (GSR) introduced by S. Hodges (1997).

Even if in theory it is possible to build utility-based risk measure that are coherent and complying with stochastic dominance principle in the practice of the banking, asset management and insurance industry, in the more than half-century (since modern utility theory was first developed and despite its theoretical appeal and long-standing academic calls for its use) it has been used only very rarely, for the following main reasons:

1. Utility estimation is by no means an easy or exact exercise and therefore it is almost impossible to build reliable utility function because:

⁵⁵ This section is based on the following research papers and books: A. Plantinga and S. de Groot, (2001), Risk Adjusted Performance Measures and Implied Risk Attitudes, working paper, University of Groningen. Z. Bercényi (2001), Accounting for illiquidity and non-normality of returns in the performance assessment, working paper, University of Munich. Pedersen, C.S. and Satchell, S.E. (1999), 'Choosing the Right Measure of Risk: A Survey', in: S.B. Dahiya, ed., The Current State of Economic Science, Vol. 2 (46) of Micro Economics, Macro Economics, Monetary Economics, Spellbound Publ. Rossi, G.A. and Tibiletti, L. (1996), Higher Order Polynomial Utility Functions: Advantages in Their Use, Technical report, Department of Mathematics and Statistics, University of Torino. J. E. Ingersoll. Theory of financial decision making. Rowman & Littlefield studies in financial economics. Rowman & Littlefield, 1987; C.F. Huang and R. H. Litzenberger. Foundations for Financial Economics. Prentice-Hall, 1988.

Chapter 2

- a) utility responses might vary between individuals and even vary for the same individual on different days.
 - b) many planning problems evolve over decade-long periods, during which utility functions are likely to change considerably.
 - c) People's valuations are likely to change as they learn more about a problem.
 - d) It is often difficult to decide who in a large organization or group should be interviewed for estimating a utility function.
 - i. For a private company, should the president, middle -manager, board members, or stockholders be interviewed?
 - ii. For public company, should the utility function be that of the agency head, private citizens, or elected officials?
 - e) High-level decision-makers are busy people; it is often hard to get their time for interviews.
 - f) Decision-makers are often unsure what their preference structure/utility function is
 - g) For newly controversial issues, these preferences may be highly unstable as new information develops.
 - h) Decision-makers, such as politicians, have political reasons to avoid revealing their true preference or utility functions.
 - i) Utility theory and processes for estimating utility are somewhat abstract and subjective. The process of estimating utility functions is somewhat artificial.
2. the maximisation of the expected utility is based on the rationality assumptions which also can lead to paradoxical results⁵⁶; in fact: experience and empirical evidence tell us investor are “not-rational”⁵⁷

⁵⁶ See Huang and Linzbergh (1998) for a closer examination of the rationality assumption

⁵⁷ If an investor is asked to choose between two possibilities: one wins 1000 dollar “for sure” (100% probability) another “lottery” being 100 dollar with 5% probability and 2100 dollar 95% probability, even if the two lotteries have the same expected value all investors will prefer the first lottery. Also the well known Allias paradox supports the intuition that investor do not behave “rationally”, which means that maximising the expected utility within the risk neutral framework is not consistent with the typical human behaviour. See also the second paragraph for a description of the difference between portfolio theory and post-modern portfolio theory

5 Distortion risk measures: definition, examples and properties⁵⁸

In this section I will consider the class of distortion risk measures, introduced by Wang (1995). A number of properties of these risk measures can be generalized to the class of distortion risk measures. Recalling the definition of the expectation of X , if it exists⁵⁹

$$E[X] = - \int_{-\infty}^0 [1 - \bar{F}_X(x)] dx + \int_0^{\infty} \bar{F}_X(x) dx \quad \text{with } \bar{F}_X = Pr(X > x) \quad (5.1)$$

Wang (1996) defines a family of risk measures by using the concept of distortion function. A distortion function is defined as a non-decreasing function $g: [0,1] \rightarrow [0,1]$ such that $g(0)=0$ and $g(1)=1$. The distortion risk measure associated with distortion function g is denoted by $\rho_g[\cdot]$ and is defined by

$$\rho_g[X] = - \int_{-\infty}^0 [1 - g(\bar{F}_X(x))] dx + \int_0^{\infty} g(\bar{F}_X(x)) dx \quad \text{for any random variable } X \quad (5.2)$$

With

⁵⁸ This section is based on the following research paper: Hürlimann, W. (2002), An economic capital allocation for lookback financial losses, Working paper. Hürlimann, W. (2002), Distortion Risk Measures And Economic Capital, Switzerland. Kaas R., Heerwaarden, van A.E. and M.J. Goovaerts (1994), Ordering of Actuarial Risks, CAIRE Education Series 1, Brussels. Goovaerts, M.J., Kaas, R. and J. Dhaene (2002), Economic capital allocation derived from risk measures. Working paper. Goovaerts, M.J., de Vijlder, f. and Haezendonck, J. (1984) Insurance Premiums. North-Holland Publishing Co., Amsterdam. Goovaerts, M.J., Kaas, R. and Dhaene, J. (2003) Economic capital allocation derived from risk measures, North American Actuarial Journal. Goovaerts M. J., R. Kaas, J. Dhaene, Qihe Tang A unified approach to generate risk measures. Goovaerts, M. J., R. Kaas, J. Dhaene, Qihe Tang (2004), Some New Classes of Consistent Risk Measures a Center for Risk and Insurance Studies (CRIS) Katholieke Universiteit Leuven. J. Dhaene, R.J.A. Laeven, S. Vanduffel, G. Darkiewicz, M.J. Goovaerts (2004) Can a coherent risk measure be too subadditive? Catholic University of Leuven, Center for Risk and Insurance Studies, Leuven, Belgium and University of Amsterdam. Dhaene, j., Denuit, m., Goovaerts, m.j., Kaas, r. and Vyncke, d. (2002a) The concept of comonotonicity in actuarial science and finance: theory. Insurance Math. Econom. 31(1), 3-33. Dhaene, j., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002b) The concept of comonotonicity in actuarial science and finance: application. Insurance Math. Econom. 31(2), 133-161. Wang, S. (1995), Insurance pricing and increased limits ratemaking by proportional hazard transforms. Insurance: Mathematics and Economics 17; pp. 43-54. Wang, S. (2000), A class of distortion operators for pricing financial and insurance risks. Journal of Risk and Insurance 67(1), 15-36. Wang, S. (2001), Equilibrium pricing transforms: new results using Bühlmann's 1980 model, Working paper. Wang, S. (2002). A risk measure that goes beyond coherence, Working paper <http://www.gloriamundi.org/picsresources/sw.pdf>. Wang, S., Young, V.R. and H.H. Panjer (1997), Axiomatic characterization of insurance prices, Insurance Mathematics and Economics 21, 173-183. Wang, s. and Young, V.R. (1998) Ordering risks: expected utility theory versus Yaari's dual theory of risk, Insurance Math. Econom. 22(2), 145-161. Wirth, J.L. and M.R. Hardy (1999). A synthesis of risk measures for capital adequacy, Insurance Mathematics and Economics 25, 337-47. Wirth J.L., and M.R., Hardy (1999), Proper Ordering for Risk Measures, working paper. Wirth J.L., and M.R., Hardy (2000), Distortion Risk Measures: Coherence and Stochastic Dominance, working paper. Wirth J.L., and M.R., Hardy (2000), Ordering of Risk Measures for Capital Adequacy, working paper. Yoshida, T. and Y. Yamai (2001). Comparative analyses of expected shortfall and value-at-risk: expected utility maximization and tail risk. Working paper; <http://www.gloriamundi.org/picsresources/tyyy.pdf>. J. Antoon Pelsser (2004), On the Applicability of the Wang Transform for Pricing Financial Risks, working paper, Erasmus University Rotterdam, www.few.eur.nl/few/people/pelsser. Artzner, Ph. (1999) Application of coherent risk measures to capital requirements in insurance, North American Actuarial Journal 3(2), 11-25. Embrechts, P., Mcneil, A.J. and Straumann, D. (2002) Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond, 176-223, Cambridge Univ. Press, Cambridge. Gerber, H.U. (1974) On additive premium calculation principles. ASTIN Bulletin 7, 215-222. Yaari, M.E. (1987) The dual theory of choice under risk. Econometrica 55(1), 95-115.

⁵⁹ A detailed proof of the integral representation of the Distorted risk measure can be found in "Empirical Estimation of Risk Measures and Related Quantities", Bruce L. Jones and Ricardas Zitkic, Department of Statistical and Actuarial Sciences, working paper University of Western Ontario, 2004. The geometric interpretation is quoted from "Risk Measures and Optimal Portfolio Selection (with applications to elliptical distributions); Jan Dhaenex, Emiliano A. Valdezz and Tom Hoedemakers; Samos 2004 Workshop, Risk Measures and Optimal Portfolio Selection, pag 26

$\rho_g[X]$ the distorted expectation of X

$g(x) \geq x$.

Figure 6 below helps clarify the relationship between the expected and the distorted expectation value of a random variable.

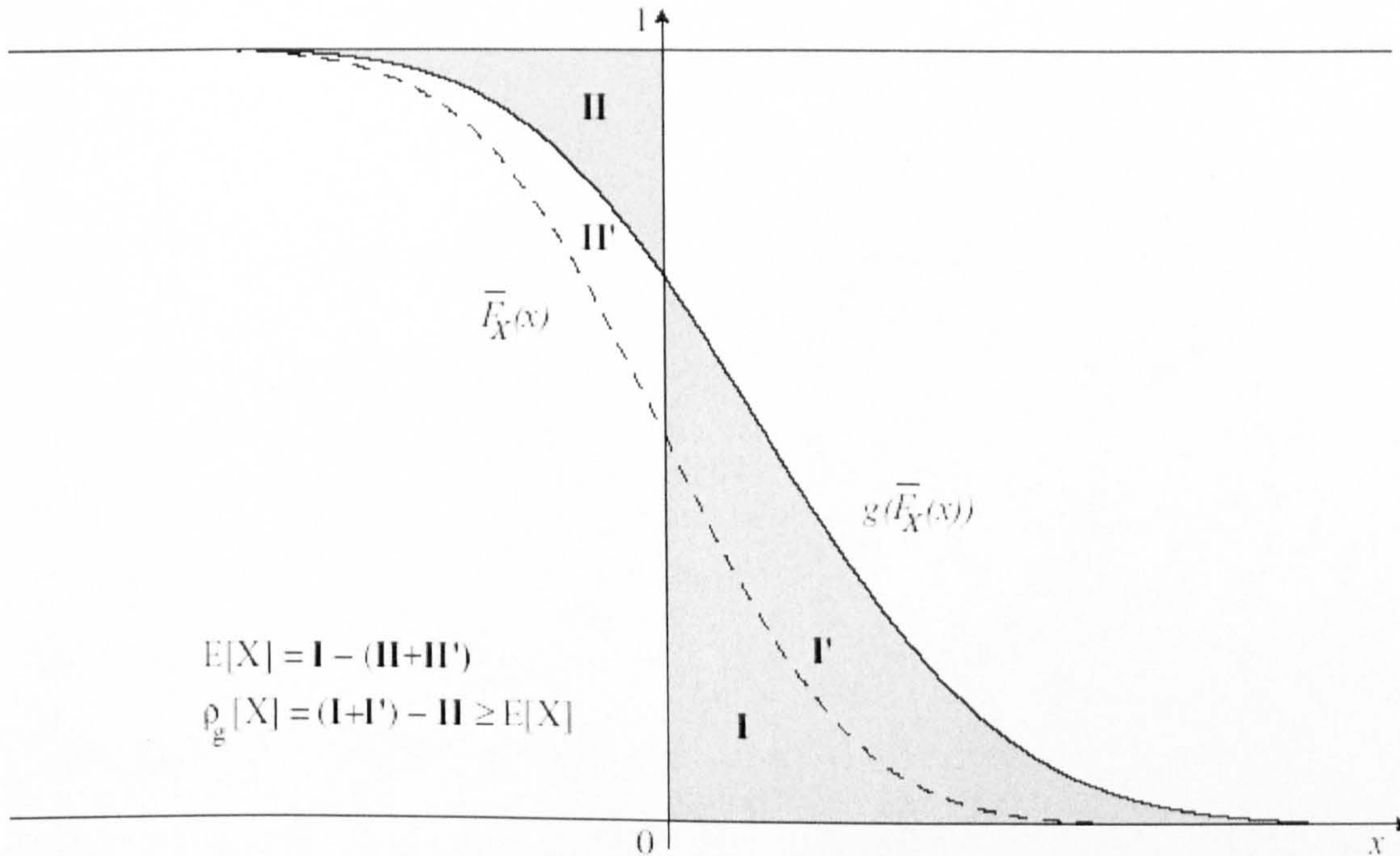


Figure 6 The relationship between the expected and the distorted value of a random variable

In figure 6 the dotted line represents $\bar{F}_X(x)$ and the solid line resembles $g(\bar{F}_X(x))$. By construction results $\rho_g[X] > E[X]$, i.e. the distorted expected value of x is greater than the expected one⁶⁰. For a general distortion function g , the risk measure $\rho_g[X]$ can be interpreted as a “distorted expectation” of X , evaluated with a “distorted probability measure”⁶¹.

Also note that $g_1(q) \leq g_2(q)$ for all $q \in [0, 1]$ implies that $\rho_{g_1}[X] \leq \rho_{g_2}[X]$. One immediately finds that $g(\bar{F}_X(x))$ is a non-increasing function of x with values in the interval $[0, 1]$. However $\rho_g[X]$ cannot always be considered as the expectation of X under a new probability measure, because $g(\bar{F}_X(x))$ will not necessarily be right-continuous.

Let now consider two examples of distorted risk measure: the Quantile and the TVaR distorted risk measure.

⁶⁰ Note that the distortion function g is assumed to be independent of the distribution function of the random variable X

⁶¹ See Denneberg (1994).

5.1 The quantile distorted risk measure

From equation (5.2), it is possible to see that the quantile $Q_p(X)$, $p \in (0, 1)$ corresponds to the distortion function

$$g(x) = I(x > 1 - p) \quad 0 \leq x \leq 1 \tag{5.1.2}$$

Where I represents the “distorted” VaR

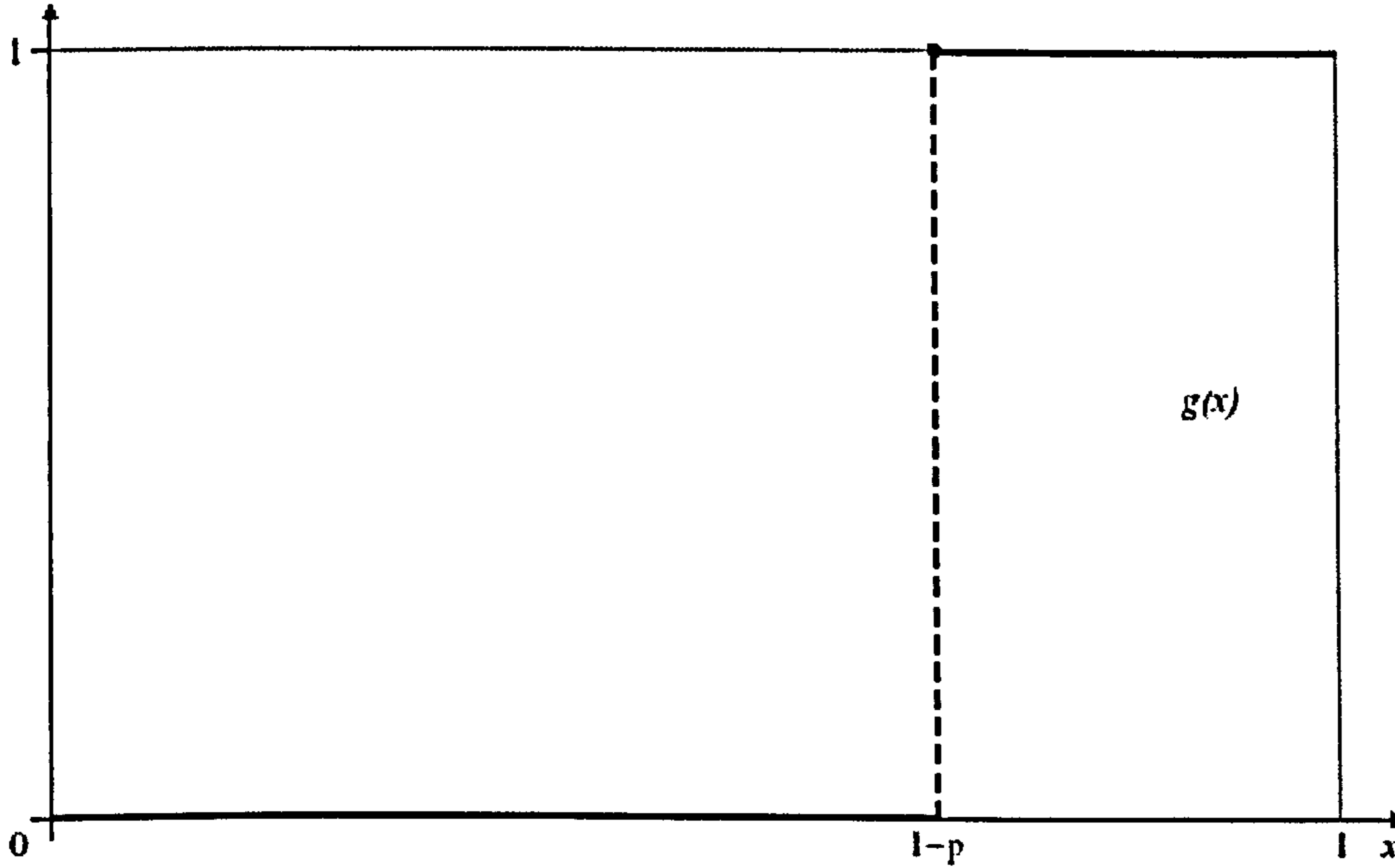


Figure 7 The Quantile Distorted Risk Measure

5.2 The TVaR_p(X) distorted risk measure

TVaR_p(X), $p \in (0, 1)$ as illustrated in section 4.2.5.2 corresponds to the distortion function

$$g(x) = \min\left(\frac{x}{1-p}, 1\right), \quad 0 \leq x \leq 1 \tag{5.2.1}$$

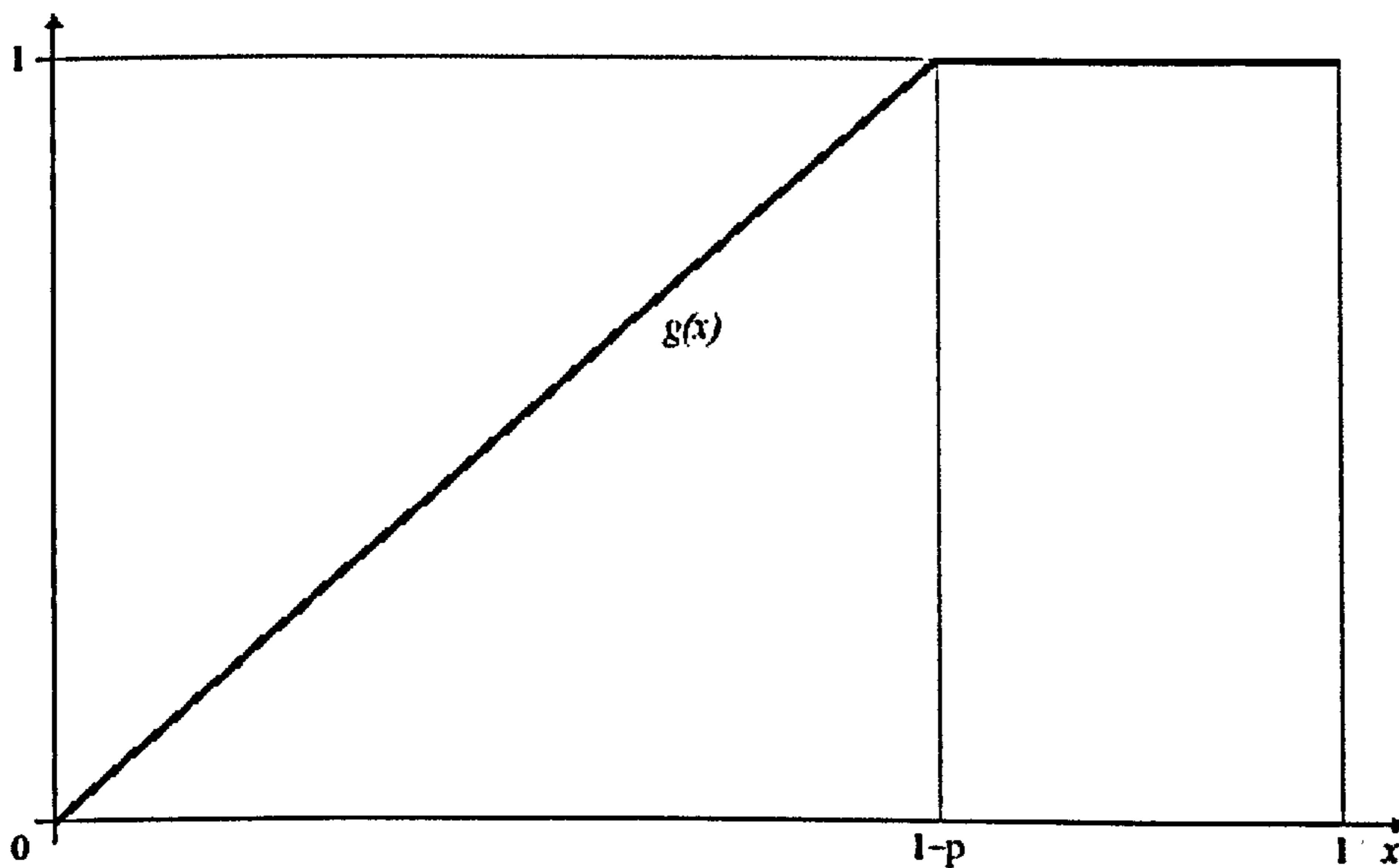


Figure 8 The Tail VaR Distorted Risk Measure

Summarising, by using TVaR_p there is no incentive for taking actions that increase the distribution function for outcomes smaller than Q_p . Moreover, if one uses the ESF, there is the possibility of not adjusting for extreme low-frequency, high severity losses. Also both ESF and CTE are not a distortion risk measure⁶². The Wang transform risk measure offers a possible solution to this problem

5.3 The Wang transform risk measure

From eq. (4.2.5.2.1) it is possible to see that the TVaR_p risk measure uses only the upper tail of the distribution. Hence, this risk measure does not create incentive for taking actions that increase the distribution function for outcomes smaller than Q_p . Also, from (4.2.5.3.4) TVaR_p only accounts for the expected shortfall and hence, does not properly adjust for extreme low-frequency and high severity losses. The Wang Transform risk measure was introduced by Wang (2000) as an example of a risk measure that could give a solution to these problems.

For any $0 < x < 1$, define the distortion function

$$g_p(x) = \Phi[\Phi^{-1}(x) + \Phi^{-1}(x_p)], \quad 0 \leq x \leq 1, \quad 0 < p < 1 \quad (5.3.1)$$

which is called the 'Wang Transform at level p '.

The corresponding distortion risk measure is called the Wang Transform risk measure and is denoted by $\text{WT}_p(X)$. For a normally distributed random variable X , results

$$1 - g_p(\bar{F}_X(x)) = \Phi\left[\frac{x - Q_p(X)}{\sigma}\right] \quad (5.3.2)$$

which implies that the Wang Transform risk measure is identical to the quantile risk measure at the same probability level in case of a normal random variable:

$$\text{WT}_p(X) = Q_p(X) \quad (5.3.3)$$

or, stated differently,

$$\rho_{g_p}(X) = Q_p(X).^{63} \quad (5.3.4)$$

Let illustrate an example to clarify how this calculation is performed

As previously described Wang introduces a new transform that can be written in the following

$$\text{form: } g_p(x) = \Phi[\Phi^{-1}(x) - \Phi^{-1}(p)] \quad \text{with } 0 \leq x \leq 1$$

Where

$\Phi^{-1}(x)$ is the standard normal percentile with respect of x

⁶² In fact the function g depends on the distribution function of X ; hence it is not possible to infer that $\text{CTE}_p(\cdot)$ is a distortion risk measure. For a complete analysis of the relationships between distorted and shortfall measure of risk and for a proof of this result see Risk measures and optimal portfolio selection, J. Dhaene, S. Vanduffel, Q. Tang, M. J. Goovaerts, R. Kaas, D. Vyncke, working paper, 2003

⁶³ Examples illustrating the fact that the WT risk measure uses the whole distribution and that it accounts for extreme low-frequency and high severity losses can be found in Wang (2001).

$\Phi^{-1}(p)$ is the inverse of the standard normal distribution at the required level of probability α .

It is possible to generalize for arbitrary $F(x)$. In this case, Wang suggests the following adjustment for parameter uncertainty $g_p(x) = Q(\Phi^{-1}(x))$ (5.3.5)

5.4 Concave distortion risk measures

A subclass of distortion functions that is often considered in the literature is the class of concave distortion functions. A distortion function g is said to be concave if for each q in $(0,1)$, there exist real numbers a_y and b_y and a line $l(x) = a_y x + b_y$, such that $l(q) = g(q)$ and $l(q) \geq g(q)$ for all q in $(0,1)$ ⁶⁴. A risk measure with a concave distortion function is then called a ‘concave distortion risk measure’. For any concave distortion function g , $g[F_X(x)]$ is right-continuous, so that in this case the risk measure $\rho_g[X]$ can be interpreted as the expectation of X under a ‘distorted probability measure’.

Concave distortion risk measures are coherent⁶⁵. For example, because results $\rho_g[X+Y] \geq \rho_g[X] + \rho_g[Y]$, concave distortion risk measures are sub-additive. This means that Q_p is not sub-additive whereas $TVaR_p$ is sub-additive.

5.5 The Beta distortion risk measure

Wirch & Hardy (2000) suggested to use for pro-active portfolio management concave distorted risk measure such as the Beta distribution function $g(x) = F_\beta(x)$, with $0 \leq x \leq 1$. This risk measure is defined as follows:

$$F_\beta(x) = \frac{1}{\beta(a,b)} \int_0^x t^{a-1} (1-t)^{b-1} dt \quad 0 \leq x \leq 1 \quad (5.5.1)$$

where $\beta(a, b)$ is the Beta function with parameters $a > 0$, $b > 0$, i.e.

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (5.5.2)$$

The Beta distortion function is strictly concave for any parameters $0 < a \leq 1$ and $b \geq 1$, with $a \neq 1$ and $b \neq 1$ ⁶⁶.

⁶⁴ A concave distortion function is necessarily continuous in $(0,1)$. For convenience, in the remainder I will always tacitly assume that a concave distortion function is also continuous at 0.

⁶⁵ In Artzner (1999) and Artzner, Delbaen, Eber & Heath (1999), as reported in section 3.2, a risk measure satisfying the four axioms of subadditivity, monotonicity, positive homogeneity and translation invariance is called “coherent”. Any concave distortion risk measure is coherent. Note that the quantile risk measure is not a concave distortion risk measure whereas $TVaR$ is a concave distortion risk measure. Therefore, the class of concave distortion risk measures is only a subset of the class of “coherent” risk measures. For proof see Wang & Dhaene (1998).

⁶⁶ For a detailed analysis of the technicalities related to the beta function see also chapter 1, section 4.2.1.

Note that the Beta distortion risk measure with the parameters $a = 0.1$ and $b = 1$ reduces to the PH-transform risk measure, which is considered in Wang (1995).

5.6 The relationship between stochastic dominance and ordered distortion risk measures

All distortion functions preserve first order stochastic dominance. This is because distortion function is defined as an increasing function. Unfortunately, as illustrated in section 3.1, not all risks can be ordered using second order stochastic dominance. In fact, any pair of risks with survival distributions that cross an even number of times cannot be compared, as the sign of the difference in integrals before the first crossing and after the last crossing are opposite. In order to rank random variables with second order stochastic dominance, an additional property is needed. Wirch and Hardy (2001) showed, based on a previous work of Wang (1996), that to preserve the SDSD a distortion function need also to be strictly concave. It is worthwhile to note that a risk measure derived from a distortion function that is concave, but not strictly concave, does not strongly preserve SDSD. A Proof of this result can be found in Wirch and Hardy (2001)⁶⁷.

Summarising, the class of **strictly concave distortion risk measures** are coherent and comply with FSDS and SDSD. As such, in theory, these are the most effective in proactive risk management as, for example, in setting up optimal credit risk portfolio management and economic capital allocation strategies. Portfolio strategies based upon risk adjusted performance that are built on top of distorted risk measure are discussed in section 7.3 and they constitute one of the most important research contribution of this chapter.

6 Risk adjusted performance measure

The risk-reward criteria can be represented through a two dimensional vector $h(\tilde{W}) = [reward(\tilde{W}), -risk(\tilde{W})]$ ⁶⁸, called performance vectors and composed of reward and risk measures of the random return \tilde{W} . Markowitz first proposed the use of a performance vector of

⁶⁷ Stochastic dominance can be also characterized in terms of ordered distortion risk measures. For any random pair (X, Y) , X is smaller than Y in stochastic dominance sense if and only if their respective distortion risk measures are ordered, i.e. $X \leq_{st} Y \Leftrightarrow \rho_g[X] \leq \rho_g[Y]$ for all distortion functions g . The proof follows immediately from Theorem 4.2 in Dhaene, Vanduffel, Tang, Goovaerts, Kaas, and Vyncke (2003).

⁶⁸ The negative sign assigned to the risk value is due to the fact that most investors want to minimize risk.

size two, in which each component is specifically designed to measure the risk-reward performance of a portfolio.

As usual, the most difficult task is to select the adequate risk and reward measures of a portfolio, to approximate the investors' behaviour accurately. In the next paragraphs the most widespread used RAPM are analysed. As for the risk measure, the end-user perspective will be the main element to consider when choosing to use a specific RAPM. This issue is of particular relevance for risk managers, asset managers and also for rating agencies. Of course this topic is of crucial importance for institutional and private investor as well. Unfortunately the answers provided in the literature are still confusing. It will be shown that both paradigms (coherence and stochastic dominance) jointly considered can provide more fine tuned guidelines for both portfolio managers and investors.

6.1 Sharpe Ratio

The Sharpe ratio is one of the most common measures of portfolio performance. William Sharpe developed it in 1966 as a tool for evaluating and predicting the performance of mutual fund managers. Since then the Sharpe ratio, and its close analogues the Information ratio, the squared Sharpe ratio and M-squared, have become widely used in practice to rank investment managers and to evaluate the attractiveness of investment strategies in general. The appeal of the Sharpe measure is clear⁶⁹. It is an affine transformation of a simple t-test for equality in means of two variables, the first variable being the manager's time series of returns and the second being a benchmark⁷⁰. The literature on performance evaluation is a large one (c.f. Brown, 2000 reference website), and much of it has focused on the limitations of standard measures. However, despite twenty years of academic understanding of the problems of benchmarking and performance measurement, the Sharpe ratio and its relatives remain fundamental tools in research and practice. In 1965 (revised 1994) W. Sharpe proposed the following simple measure of relative performance, namely the Sharpe Ratio:

$$SR_i = \frac{(R_i - R_b)}{\sigma_{R_i, R_b}} \quad (6.1)$$

where

R_i = investment return of asset i over some time horizon T

⁶⁹ Unfortunately, the Sharpe ratio is prone to manipulation – particularly by strategies that can change the shape of probability distribution of returns. For example, Henriksson and Merton (1981) and Dybvig and Ingersoll (1982) show that non-linear payoffs limit the applicability of the Sharpe ratio to the problem of performance evaluation. More recently, Bernardo and Ledoit (2000) show that Sharpe ratios are particularly misleading when the shape of the return distribution is far from normal.

⁷⁰ For a review of its history and use, see Sharpe (1994). For a current textbook discussion and applications of the Sharpe ratio, see for example, Bodie, Kane and Marcus (1999) p. 754-758. For applications in the mutual fund industry, see Morningstar (1993) p.24

Chapter 2

R_b = benchmark expected return over same T

$R_i - R_b$ = excess return over same time T

σ_{R_i, R_b} = standard deviation of excess return over T

The SR of any investment is defined as its return in excess of a benchmark return divided by the standard deviation of excess return⁷¹. If A and B are two mutually exclusive investments, an investor should invest in a mixture of A (and therefore not in B) and the risk free asset if and only if $SR_A > 0$ and $SR_A > SR_B$. The justification is based on 2 assumptions:

- Return on A and B are similarly distributed
- Investors prefer higher returns of the same risk level or lower risk levels for the same return (they are risk averse).

The optimality of the Sharpe rule can be seen from a simple heuristic argument. Suppose to consider a benchmark, b, and to face a choice between two mutually exclusive investments, A and B, where A has the higher expected SR. The investment choices are depicted in table 4.1. The idea behind the Sharpe rule is following: take the benchmark as a hypothetical initial investment, and then try to select an alternative that improves on the benchmark in risk-expected return terms⁷². However, applying the Sharpe rule to the choice of alternative investments can give misleading answer if:

- Returns on those investments are correlated with the existing portfolio return
- Returns are not normally distributed.

For example, through inspection of table 2 it is possible to observe that $SR_A=0.50$ and $SR_B =0.33$. The Sharpe rule would therefore recommend choosing A over B. Since A is positively correlated and B is negatively correlated with the existing position, the “naive” Sharpe rule fails to take account that A increases the overall risk while B tends to reduce it.

To take account of this effect, the Sharpe rule applies to the alternative portfolios resembling the investment (A or B) plus the existing position rather than the single alternative investments. From inspection of table 6.1 it is possible to note that:

- The portfolio of investment A plus the existing position produces a SR of 0.15
- The portfolio of investment B plus the existing position produces a SR of 0.20.

⁷¹ The benchmark very often is represented by a riskfree investment alternative, typical, a treasury bond maturing at T.

⁷² A high expected SR therefore represents a good departure from the benchmark because it implies a considerably higher expected return than the benchmark for relatively little extra risk, and a low expected SR is a bad departure because it offers little extra expected return relative to the greater risk entailed

The correct choice is therefore investment B⁷³.

In the next section a comparison with the analogous utility based risk measure, namely the Generalised Sharpe Ratio, and a practical example (reported in the appendix 4) to better clarify how one can go wrong when founding its decision making process on this popular RAPM is presented. For example, paradoxical results may arise if an asset manager uses SR to do stock picking or if a credit risk managers selects the optimal bank portfolio, with respect to a pre-determined risk-return profile, as it will be shown in section 7.

	Investment A	Investment B	Benchmark	Initial Position	Position + Investment	
					P+A	P+B
size	1	1		1	2	2
Exp. Return	11%	10%	7%	8%	9,50%	9,00%
St. Dev. Ret.	8%	9%		7%	13,4%	7,4%
Exp Exces Ret	4%	3%		0%	2,00%	1,50%
Std Dev Exces Ret	8%	9%			13,4%	7,4%
Corr. Port	0,6	-0,6	0			
SR	0,50	0,33			0,15	0,20

Table 2 Sharpe Ratio comparison: single investment vs portfolio

6.1.1 Information Ratio

The way the SR has been presented in the previous section is not the traditional one in which the benchmark is assumed to be the risk free-rate. In fact, by just considering instead of the asset i a generic portfolio (or fund) p the SR becomes the widely used risk adjusted performance measure (RAPM) Information Ratio IR_p . In this case, from the perspective of the asset management industry, the standard error represents the Tracking Error Volatility ($TEV_{p,b}$) of the fund.

Therefore:

$$SR_p = IR_p = \frac{r_p - r_b}{TEV} \quad (6.1.1.1)$$

where

R_p = investment return of asset i over some time horizon T

R_b = benchmark expected return over same T

$R_p - R_b$ = excess return over same time T

$\sigma_{R_p, R_b} = TEV_{p,b}$ standard deviation of excess return over T

⁷³Although the Sharpe ratio has become part of the canon of modern financial analysis, its applications typically do not account for the fact that it is an estimated quantity, subject to estimation errors that can be substantial in some cases. This important aspect will not be analysed in this chapter. For a detailed analysis of this topic refer to The Statistics of Sharpe Ratios Andrew W. Lo AIMR, 2002.

6.2 Treynor Ratio

The second RAPM, the Treynor Ratio, is also derived from the CAPM. In fact the CAPM poses a linear relationship between the excess return of an asset and excess return of the market to which the asset belongs

$$R_i - r_f = \alpha_i + \beta_i(R_m - r_f) \quad (6.2.1)$$

Where

R_i = investment return of asset i over some time horizon T

r_f = risk free return over same T

R_m = investment return on the market portfolio over some time horizon T

α_i = investment's excess return over the market over same time T

$$\beta_i = \frac{\sigma_{i,m}}{\sigma_m} \text{ CAPM measure of risk}$$

Treynor's Ratio (Treynor, 1966) TR is defined as the ratio

$$TR = \frac{\alpha_i}{\beta_i} \quad (6.2.2)$$

The investor that want to follow this rule has to choose investment A over B if:

$$\frac{\alpha_A}{\beta_A} > \frac{\alpha_B}{\beta_B} \Rightarrow \frac{R_A - r_f}{\beta_A} > \frac{R_B - r_f}{\beta_B} \text{ Moreover, substituting (6.1) in (6.2.1) and using } \beta_{i,m} = \frac{\rho_{i,m}\sigma_i}{\sigma_m},$$

$$\text{results } TR = \frac{SR_i}{\rho_{i,m}}.$$

Some obvious conclusions now follow: the Treynor rule is not reliable if we are working with any benchmark other than the risk-free asset. Even if we are working with a risk-free benchmark, the Treynor rule will only be reliable if the alternative investments have returns that are equally correlated with the market return (that is, if $\rho_{A,M} = \rho_{B,M}$). Because there is no particular reason to expect this correlation condition to hold, there is no reason to believe the Treynor rule to be reliable even if the risk-free asset is choosed to be the benchmark. Furthermore, the same criticism described for the SR rule applies to TR.

The TR can be also presented in another form $T_p = \frac{r_p - r_f}{\beta_p}$, which can be considered one of the many variants this RAP can be founded in the literature.

6.3 Jensen alpha J_α

Jensen (see Jensen, 1968) proposed to consider only the asset excess return over the market, that is α . An investor chooses investment A over B if $\alpha_A > \alpha_B$. The basic idea is to analyse the performance of an investment manager one must look not only at the overall return of a portfolio, but also at the riskiness of that portfolio. For example, if there are two mutual funds that both have a 12% return, a rational investor will want the fund that is less risky. Jensen's measure is one of the ways to help determine if a portfolio is earning the proper return for its level of risk. If the value is positive, then the portfolio is earning excess returns. In other words, a positive value for Jensen's alpha means a fund manager has "beat the market" with his stock picking skills.

Moreover, substituting (6.1) in (6.2.1) and using $\beta_{i,m} = \frac{\rho_{i,m}\sigma_i}{\sigma_m}$,

Results that

$$\sigma_A \left(\frac{R_A - r_f}{\beta_A} \right) - \sigma_B \left(\frac{R_B - r_f}{\beta_B} \right) > (\sigma_A \rho_{A,M} - \sigma_B \rho_{B,M}) \left[\frac{R_M - r_f}{\sigma_M} \right]$$

consequently

$$\sigma_A (SR_A) - \rho_{A,M} (SR_M) > \sigma_B (SR_B) - \rho_{B,M} (SR_M)$$

Jensen's alpha therefore provides the same preferences as the Sharpe Ratio only for assets with the same σ and the same ρ . In particular, for the Alpha rule to give the same decisions as the Sharpe rule, three conditions to hold are required: the same two conditions needed for the Treynor rule to be reliable, and the additional condition that the returns on the two investments have the same standard deviation. The alpha rule is therefore reliable only under conditions even more restrictive than those required for the Treynor rule.

Summarising the alpha rule is inferior to the Treynor rule, and it was already proved that the Treynor rule is inferior to the Sharpe rule.

Consider the following example to see how to consistently select the best investment alternative. Let A and B be two investment alternatives showing the returns, volatilities and correlations reported in table 3. From inspection of table 3 it is possible to see that:

- SR rule rank investment A superior than B
- TR and Alpha rank B superior than A

Because it has been shown that SR is a superior RAP it is better to avoid using TR and Jensen Alpha to avoid sub-optimal decision⁷⁴.

⁷⁴ As for the SR and the TR also the J_α can be founded in the literature under other forms like for example the following one: $J_\alpha = r_p - [r_f + \beta_p (r_M - r_f)]$. In any case the same criticism applies.

	Investment A	Investment B	Benchmark	mkt port	Position + Investment	
					P+A	P+B
size	1	1		1	2	2
Exp. Return	12,0%	8,0%	5,0%	9,0%	10,5%	8,5%
St. Dev. Ret.	12,0%	8,0%	0,0%	9,0%	21,0%	13,7%
Exp Exces Ret	7,0%	3,0%		4,0%	5,5%	3,5%
Std Dev Exces Ret	12,0%	8,0%		9,0%	21,0%	13,7%
Corr. Port	1,00	0,30	0,00	1,00		
Beta	1,33	0,27		1,00		
Alfa Jensen	0,017	0,019				
Treynor	0,013	0,073				
SR	0,583	0,375		0,444	0,262	0,255

Table 3 Sharpe Ratio, Jensen Alfa and Treynor Ratio comparison

6.4 The Return on Value at Risk (ROVaR)

The ROVaR is the relevant expected return divided by the VAR:

$$ROVaR_i = \frac{R_i}{VaR_i} \quad (6.4.1)$$

where R_i is the (expected) return on asset i and VaR_i is the Value-at-Risk (at the chosen confidence level).

The ROVaR rule tells us to choose the investment or portfolio with the highest (expected) return over VaR. One can assess the reliability of the ROVaR rule by comparing it to the Sharpe rule. To do so, first replace the VaR in the previous equation with $-(R_i + \alpha\sigma_{R_i})$. Here, α is the standard normal variate reflecting the confidence level on which the VAR is estimated. It follows that:

$$ROVaR_i = -\left(\frac{R_i}{R_i + \alpha\sigma_{R_i}}\right) \quad (6.4.2)$$

by making the additional assumption of a risk-free benchmark it is possible to show how ROVaR relate to SR:

$$RoVaR_i = -\left(\frac{R_f + SR_i\sigma_{R_i}}{\alpha\sigma_{R_i} + R_f + SR_i\sigma_{R_i}}\right) \quad (6.4.3)$$

Equation (6.4.3) indicates there is no reason to expect that the ROVaR rule and the Sharpe rule will give the same rankings of alternative investments. Indeed, this becomes obvious if one

makes the further simplifying assumption that the risk-free return is zero. In this case, (6.4.3) reduces to:

$$ROVaR_i = - \left(\frac{SR_i \sigma_{R_i}}{\alpha \sigma_{R_i} + SR_i \sigma_{R_i}} \right) \quad (6.4.4)$$

Since α can take any finite positive or negative value, the two rules will not give the same rankings.

Summarising, preference rankings set by this measure can be very different from those obtained from the Sharpe rule and highly dependent on the choice of α . In a static context (one time period) and for assets with similar distributed returns, among the proposed performance measures, the Sharpe rule is the only measure that is coherent with the mildest form of investor risk aversion. However, in practice, one has to compare the performance of wildly different investment strategies leading to very different return distributions. The Sharpe rule is therefore severely limited in its applicability. Obviously most of the criticism previously considered when discussing the VaR risk measure properties also applies to RoVaR.

6.5 Generalised Sharpe Ratio (GSR)

In an influential preprint, S Hodges (1997) proposes a generalisation of SR for investors with constant absolute risk aversion, which is investors who maximise $E[U(\tilde{W})]$ with $U(\tilde{W}) = -\exp(-\lambda W)$ for some $\lambda > 0$ reflecting absolute risk aversion. \tilde{W} Denotes the total wealth of the investor. For simplicity, assume $w_0 = 0$ so that $U(\tilde{W}) = -1$. Favourable investment opportunities taken in optimal quantities will lead to optimal expected utilities U^* in the range $(-1, 0)$. Consider an investment opportunity over a time horizon T leading to (forward annualised) mean returns μT and standard deviation σ , i.e. $returns \approx N(\mu T, \sigma^2 T)$ and maximise the expected utility of an investment of X units of capital in this opportunity as compared to the risk free investment yielding r . Mathematically,

$$MAX E[U(x)] = -\exp \left\{ - \frac{((\mu - r)Tx - 0.5\lambda \sigma^2 T x^2)}{\lambda} \right\} \quad (6.5.1)$$

We find

$$U^* = Max E[U] = -\exp \left\{ -0.5 * \frac{(\mu - r)^2 T}{\sigma^2} \right\} \quad (6.5.2)$$

which is independent of λ .

SR can be expressed in terms of the maximum expected utility as follows;

$$SR = \frac{(\mu - r)}{\sigma} = \sqrt{\left(\frac{-2 \ln(-U^*)}{T}\right)} \quad (6.5.3)$$

$$\text{Hodges (1997) defines the } GSR = \sqrt{\left(\frac{-2 \ln(-U^*)}{T}\right)} \quad (6.5.4)$$

It is a measure of performance for investors with constant absolute risk aversion, that is, with utility $U(\tilde{W}) = -\exp(-\tilde{W}\lambda)$, whatever their coefficient of risk tolerance λ ($\lambda = 1$ can be chosen without loss of generality). In the next section it will be shown that: GSR can be applied to compare any type of return distributions, GSR is always compatible with stochastic dominance and GSR could have many other applications than performance ranking.

Summarising, choices among various investment opportunities are bound to depend on some expression of risk attitude. In the simple case of similarly distributed opportunities, a mild expression of risk aversion suffices to justify the Sharpe ratio rule. For arbitrary distributions, a more precise expression of risk aversion is necessary. Choosing constant absolute risk aversion (of whatever degree) suffices to define a GSR with pleasant and intuitive properties. It would be particularly useful to use the GSR to compare performances of highly skewed returns such as with hedge funds. However caution must always be exercised when knowledge of the return distribution is based on ex-post data.

It is also worth to mention that utility theory based RAP have not been widespread used in practice, in the asset management industry in particular, as it will be discussed in section 7.1.

An example of how SR and GSR perform when the underlying return distribution is not symmetrical is illustrated in the appendix 4.

6.6 RAP measures based on downside risk

In addition to the SR this study also examines several risk-adjusted performance measures that use the so-called downside deviation with respect to a reference point. The reference point, which may also be called the minimal acceptable rate of return, is used to distinguish “risk” from “volatility”. According to Sortino and Van der Meer [1991], realizations above the reference

point imply that goals are accomplished and, therefore, are considered “good volatility”. Realizations below the reference point imply failure to accomplish the goals and should be considered “bad volatility” or risk. Based on this premise, this study investigates the Sortino ratio and the Upside-Potential Ratio.

6.6.1 Sortino Ratio

The Sortino ratio is probably the most well known downside risk adjusted performance measure, and it is calculated as follows:

$$SOR = \frac{E(R) - \tau}{LPM_2(\tau, X)} \quad (6.6.1.1)$$

where τ is the minimal acceptable rate of return and $LPM_2(\tau, X)$ is the downside risk with respect to the minimal acceptable rate of return. In the variance calculation no distinction is made between upside and downside deviation. For this reason, an investment with monthly returns of -5% and +5% will have the same variance as another investment that is flat one month and +10% the next. In accordance with the description above, the Sharpe Ratio is therefore using a non directionally biased measurement of volatility to adjust for risk. This concept has been criticized, as it may actually punish a fund for a month of exceptionally high performance. For many individuals, this type of deviation is not only acceptable, but also desirable. It is for this reason that the Sortino Ratio was developed. Instead of using standard deviation in the denominator, the Sortino Ratio uses downside semi-variance. This is a measurement of return deviation below a minimal acceptable rate. By utilizing this value, the Sortino Ratio is only penalizing for “harmful” volatility. It is a measurement of return per unit of risk on the downside.

Although there are arguments in favor of both ratios, the use of the Sharpe has been more widespread. In some cases, this may reflect a certain comfort level associated with its use of standard deviation, which is a more traditional measurement of volatility. Funds that cite their Sortino Ratio have traditionally been those with the least tolerance for risk. In these cases, the Sortino may be presented as a compliment to an investment thesis that stresses the containment of losses to a minimum.

Although both the ratios are measurements of return-to-risk, understanding the distinctions of each may provide insight into their unique drawbacks. It is important to note that thorough interpretation of their values requires attention to each of these considerations

6.6.2 Upside-Potential Ratio

An alternative for using the expected return is the so-called Upside Potential Ratio, which is the probability-weighted average of returns above the reference rate:

$$UPR = \frac{\text{Upside Potential}}{\text{Downside Risk}} \quad (6.6.2.1)$$

The upside potential ratio was developed by Sortino, Van der Meer, and Plantinga [1999] and is defined as:

$$UPR = \frac{\sum_{t=1}^T i^+ \frac{1}{T} (R_t - \tau)}{\sum_{t=1}^T i^- \frac{1}{T} (R_t - \tau)^2} \quad \text{with } i^+ = 1 \text{ if } R_t > \tau; i^+ = 0 \text{ if } R_t \leq \tau; i^- = 1 \text{ if } R_t \leq \tau; i^- = 0 \text{ if } R_t > \tau \quad (6.6.2.2)$$

where t is the number of periods in the sample, R_t is the return of an investment in period t . An important advantage of using the upside potential ratio rather than the Sortino ratio is the consistency in the use of the reference rate for evaluating both profits and losses.

An important difference between downside risk and standard deviation is the use of an exogenous reference rate versus the mean return. The investor's objective function motivates the choice of the reference rate. As a result, a part of the investor's preference function is introduced into the risk calculation. Therefore, the resulting calculation is only valid for individuals sharing the same reference rate. Investors with different minimal acceptable rates of return will have different rankings.

6.6.3 The Generalised Sortino Ratio: the Kappa and Omega RAP

The Sortino Ratio and the more recently-specified Omega statistic, as defined by Shadwick and Keating [2002], can be used as alternatives to the Sharpe ratio in measuring risk-adjusted return. Unlike Sharpe, neither assumes a normal return distribution, and each focuses on the likelihood of not meeting some target return. In contrast Sharpe measures only the sign and magnitude of the average risk premium relative to the risk incurred in achieving it. As specified, the Sortino Ratio and Omega appear distinct. **Omega** is defined as:

$$\Omega(\tau) = \frac{E(R) - \tau}{LPM_1(\tau, X)} + 1 \quad (6.6.3.1)$$

Despite their (Omega and Sortino Ratio) apparent distinctiveness, each represents a single case of a more generalized risk-return measure, defined below, called Kappa (K). K_α generates Omega when $\alpha=1$, the Sortino Ratio when $\alpha=2$, or any of an infinite number of related risk-return measures when n takes on any positive value. Kappa is undefined where $\alpha \leq 0$

Definition of Kappa

$$K_\alpha(\tau) = \frac{E(R) - \tau}{LPM_\alpha(\tau, X)} \quad (6.6.3.2)$$

Substituting equation (6.6.3.2) into equation (6.6.3.1) provides an alternative, fully equivalent definition of the Sortino ratio as

$$\Omega(\tau) = K_{\alpha}(\tau) + 1 \quad (6.6.3.3)$$

Hence the Omega statistic and the Sortino ratio have identical structures, being equal to K_1+1 and K_2 respectively. Despite the addition of a constant to K_1 in the Shadwick and Keating definition of Omega, Omega and K_1 are for all practical purposes identical. I will refer to them interchangeably below. K_{α} is defined for any value of n exceeding zero. Thus, in addition to K_1 and K_2 , any number of K_{α} statistics may be applied in evaluating competing investment alternatives or in portfolio construction. All Kappa curves represent an inverse relationship between the threshold return chosen and the value of Kappa. The steepness of the K_{α} curve at any given threshold return is inversely related to the chosen value of the n parameter.

By construction, every Kappa variant has a value of zero when the threshold return equals the average figure. Note that although K_2 – the Sortino ratio - is usually expressed as a single-point statistic relative to a single threshold return, it is more informative to plot this Kappa variant, as well as others, against a range of threshold values. All Kappa curves are monotonic. Interpretation of differences in K_{α} values at different return thresholds is complex: it can be seen that the rate of change in Kappa as a function of τ is inversely proportional to α . Kappa is insensitive to skewness for values of τ which lie close to or above the mean return; but sensitive to skewness when τ is substantially below the mean return. Kappa sensitivity to skewness, at low values of τ , is a negative function of α , and the α parameter can be interpreted as a measure of skewness risk “appetite” for threshold returns below the mean. As is the case with skewness, Kappa is insensitive to kurtosis for values of τ that lie close to or above the mean return; but sensitive to kurtosis when τ is substantially below the mean return. the α parameter of Kappa can be interpreted as a measure of kurtosis risk aversion for return thresholds below the mean. This is in contrast to skewness, for which the α parameter appears to represent a measure of risk appetite.

Detailed derivations, descriptions and suggested applications of K_1 (Omega) and K_2 (Sortino ratio) already exist. There remains the question of what Kappa “means” in these cases or, as well, when α is set to some other value such as 0.5 or 3⁷⁵.

⁷⁵ While interesting, the relationship between Kappa variants and particular moments of a return distribution is not meaningful. For instance, while K_2 depends in part on semi-variance, and K_1 depends in part on a semi-mean, there is no corresponding distribution moment for Kappa variants with non-integer parameter values, nor is the relationship between K_3 and a notional “semi-skewness” statistic, or K_4 and a “semi-kurtosis” statistic, easily interpreted. All Kappa variants are sensitive, to some degree, to the first four as well as other moments of the return distribution.

6.6.4 Remarks

Omega and the Sortino ratio are two among many potential variants of Kappa. In certain circumstances, other Kappa variants may be more appropriate or provide insights that are more powerful. The ranking of a given investment alternative can change according to the Kappa variant chosen, due in part to differences among the variants in their sensitivity to skewness and kurtosis. The choice of one Kappa variant over another will therefore materially affect the user's evaluation of competing investment alternatives, as well as the composition of any portfolio optimized to maximize the value of Kappa at some return threshold.

It does not exist, unfortunately, a generally applicable rule for choosing the "correct" Kappa variant for a given purpose. For the purposes of simple comparisons among competing investment alternatives at "ordinary" minimum return thresholds, Kappa may be estimated efficiently using a parameter-based calculation that eliminates the need to gather and manage discrete return data. However, this estimation method may lead to material discrepancies in Kappa estimates at low return thresholds, and so should be used with caution for the purposes of stress testing or portfolio construction.

7 Portfolio management and risk adjusted performance

In this paragraph four portfolio management frameworks will be illustrated:

1. Portfolio management when investor preferences can be resembled through quadratic utility function (see section 7.1)
2. Portfolio management in the traditional mean-variance framework (see section 7.1)
3. Portfolio management with risk adjusted performance built on top of downside risk measure (see section 7.2)
4. Portfolio management with risk adjusted performance measure built on top of distorted risk measure (see section 7.3)

In section 7.3 a RAP built on top of a distorted risk measure, used only in the insurance industry, is for the first time analysed within a banking and finance industry set up. This section represents one of the most important contributions of the paper.

7.1 Portfolio management when using quadratic utility theory and mean/variance principle

Von Neumann/Morgenstern (1953) assumed that rational investors behave according to some axioms of rationality⁷⁶. They showed that this assumption is consistent with maximizing expected utility function $U(\tilde{W}_{t+1})$, where \tilde{W}_{t+1} denotes the end-of-period wealth of a representative investor, if $U(\tilde{W}_{t+1})$ shows the following properties:

1. twice differentiable
2. increasing
3. strictly concave⁷⁷

Then the investors' optimization problem can be stated as follows⁷⁸:

$$\text{Max } E(U(\tilde{W}_{t+1})) \quad (7.1.1)$$

In this section, a single period case is considered and therefore the time sub-index is dropped from now on. To investigate the possible outcomes of the future end of wealth value Taylor's expansion of the expected utility around the expected wealth need to be evaluated.

$$U(\tilde{W}) = U(\bar{W}) + U^{(1)}(\bar{W})(\tilde{W} - \bar{W}) + \left[\frac{1}{2!} U^{(2)}(\bar{W}) \right] (\tilde{W} - \bar{W})^2 + \sum_{i=3}^{\infty} \left[\frac{1}{i!} U^{(i)}(\bar{W}) \right] (\tilde{W} - \bar{W})^i \quad (7.1.2)$$

where

$$U(\bar{W}) = U(E\tilde{W})$$

$$E(\tilde{W} - \bar{W})^2 = \sigma_w^2 \text{ denotes the variance of wealth.}$$

Substituting the above expression into equation (4.19), the indirect utility (preference) function takes the form

$$E[U(W)] = EU(\bar{W}) + \left[\frac{1}{2!} U''(\bar{W}) \right] \sigma_w^2 (\tilde{W}) + \sum_{i=3}^{\infty} \left[\frac{1}{i!} U^{(i)}(\bar{W}) \right] m_i (\tilde{W})^{80} \quad (7.1.3)$$

m_i represents the i -th moment of the probability distribution of $U(\tilde{W})$. Equation (7.1.3)

highlights the fact that the expected utility of wealth depends on:

- the expected end-of-period wealth
- the variance of end-of-period wealth, σ_w^2
- the higher moments of the probability distribution of end-of-period wealth

⁷⁶ See Huang and Litzenberger (1988)

⁷⁷ See Ingersoll (1987), pp. 21-22

⁷⁸ See Huang and Litzenberger (1988), pp. 60-61

⁷⁹ It is easy to show that $E[U'(\bar{W})(\tilde{W} - \bar{W})] = U'(\bar{W})E(\tilde{W} - \bar{W}) = 0$

⁸⁰ If I calculate the third and fourth derivative of the utility function U with respect to the term in brackets I get:

$$E[U(W)] = EU(\bar{W}) + \left[\frac{1}{2!} U''(\bar{W}) \right] \sigma^2(W) + \left[\frac{1}{3!} U'''(\bar{W}) \right] s^3(W) + \left[\frac{1}{4!} U^{(4)}(\bar{W}) \right] k^4 + \dots$$

If the end-of-period wealth $U(\tilde{W})$ is normally distributed, then the moments of order greater than two can be stated as functions of the expected end-of-period wealth and its variance. This allows reformulating (7.1.3) as:

$$E[U(W)] = EU(\bar{W}) + \left[\frac{1}{2!} U''(\bar{W}) \right] \sigma_w^2(\tilde{W}) \quad (7.1.4)$$

Therefore, (7.1.4) indicates that investors have to concern themselves merely with the mean and the variance of the end-of-period wealth \tilde{W} when \tilde{W} is normally distributed.

Some interesting remarks can be drawn:

- equation (7.1.3) holds for any probability distribution
- equation (7.1.4) holds for normal probability distribution
- the mean/variance principle can be deduced out of (7.1.4)
- the higher the expected utility of wealth given a constant level of σ_w^2 , the higher the expected level of end-of-period wealth will be

If $E\tilde{W}$ is held constant, σ_w^2 has to be minimized to maximize the expected utility.

Thus:

a) if \tilde{W} is assumed to be normally distributed:

- the "performance" of any portfolio according to von Neumann/Morgenstern's utility theory is measured as the expected end-of-period wealth assessed by a portfolio of assets
- the portfolio's risk is measured by the variance of the end-of-period wealth. This is called a mean/variance principle of portfolio diversification.

b) if a normal distribution for the end-of-period wealth cannot be assumed

- the mean/variance principle can be rescued for arbitrary wealth probability distributions, if quadratic utility functions of the form $U(\tilde{W}) = \tilde{W} - \frac{b}{2}\tilde{W}^2$ are imputed.

In the last case all moments of order greater than two are zero and can thus be dropped. Substituting the above utility function into equation (7.1.3) the new objective function that maximize the investor's utility may be expressed as:

$$EU(\tilde{W}) = E\tilde{W} - \frac{b}{2}E\tilde{W}^2 = E\tilde{W} - \frac{b}{2}E((E\tilde{W})^2 + \sigma^2) \quad (7.1.5)$$

Just as in equation (7.1.4) equation (7.1.5) shows that investors endeavor to minimize the variance of wealth for a given level of expected end-of-period wealth.

Summarising two alternative assumptions can be made to deduce the mean-variance principle out of utility theory:

1. The end-of-period wealth is distributed normally⁸¹
2. Investors are faced with quadratic utility functions⁸².

From a portfolio management perspective the quadratic utility function says that, after a certain point, when the investor becomes wealthy enough, he starts enjoying the incremental returns even less. Consequently, the more he earns the less satisfaction he has! This means that he would now try to give back some of the wealth he is earning. It is because the shape of the curve is actually the shape of a quadratic function. The first part goes up reasonably well within a range of behaviors that says, as the investor gets more and more wealth, he values the next unit at slightly less value than the previous unit. When he gets to the point of having enough, it is like giving it all back. Mathematically the wonderful simplifying assumption of quadratic utility tied in with mean-variance. However, this means also carry on the baggage of negative utility beyond a certain point along with the benefits of the simplicity calculating at the front-end of the utility function.

7.2 Portfolio management when using RAP downside risk measure

Another important issue for portfolio asset and risk managers concern the optimal portfolio selection problem which is solved by making use of downside risk measure in the denominator of the RAPM. As one might expect, the resulting efficient frontier contains all portfolios that own the highest level of expected return at a given level of risk and the lowest risk for a given expected return. In order to find the efficient portfolio with expected return μ from the opportunity set P of N securities, the following non-linear optimization problem must be solved:

$$\begin{aligned} & \min LPM_{\alpha}(\tau, X) \\ & E(X) = \sum_{i=1}^N \varpi_i X = \mu \\ & \sum_{i=1}^N \varpi = 1 \quad \varpi_i > 0 \end{aligned} \tag{7.2.1}$$

With the expectation (or mean) μ being defined as the average of T historical portfolio returns.

After having separated the MLP-*efficient* portfolios from the dominated portfolios, the stage of portfolio choice entails the selection of an optimal portfolio depending on the specific preferences of the investor. Though intuitively very appealing, the lower partial moment risk

⁸¹ The normal distribution assumption with risk equal to standard deviation is the basis of all the Markowitz work, the Sharpe work and the Capital Assets Pricing Model Theory. Virtually all of modern finance builds itself on this underlying concept that the world is normally distributed and risk is measured symmetrically. When start breaking those restrictions, the ability to measure many more things is created.

⁸² It is worth to point out that the concept of uncertainty incorporated in the mean-variance framework is that both upside and downside volatility are penalized the same. It measures risk only relative to the mean. It does not measure the risk relative to a target level.

measures are computationally much more complex than their full domain equivalents and can be rather problematic to use in a portfolio context. Full domain moments display nice symmetry and decomposition properties. The portfolio's return variance can be decomposed into covariances between the component returns, and the covariance of security i with security j equals the covariance of j with i . Likewise the portfolio's higher order return moments can be decomposed into co-moments between the incorporated securities and, for example, the co-skewness between securities i, j and k equals the co-skewness between securities k, i and j .

Of course, the ability to decompose a portfolio's full domain return moment into individual co-moments translates into the ability to aggregate individual co-moments to the portfolio's full domain return moment. Lower partial moments, in contrast, are constructed to be asymmetrical risk measures, so it is not surprising that co-lower partial moments possess a similar asymmetry. Consequently, the co-lower partial moments between the individual security returns cannot be aggregated to the portfolio's return lower partial moment. This aggregation to the portfolio level cannot be done, not even when one takes into account that the co-lower partial moment of security i with security j is different from the co-lower partial moment of j with i .

$$LPM_2(\tau, R_p) = E\left\{\max[0, \tau - R_p]^2\right\} = \sum_{i=1}^N \varpi_i E\left\{(\tau - R_i) \max[0, \tau - R_p]\right\} \quad (7.2.2)$$

where R_p is the return of the portfolio.

But, unfortunately

$$\sum_{i=1}^N \varpi_i E\left\{(\tau - R_i) \max[0, \tau - R_p]\right\} \neq \sum_{i=1}^N \sum_{j=1}^N \varpi_i \varpi_j E\left\{(\tau - R_i) \max[0, \tau - R_j]\right\} \quad (7.2.3)$$

Hence, copying the security analysis from the Markowitz framework is of no use: it is simply not possible to construct the portfolio's target semi-variance from the N securities' (co-) semi-variances. Since, in general, the set (co)lower partial moments does not provide a shortcut to circumvent the computational complexity of the portfolio selection problem, we must use the entire empirical joint distribution of the security returns in the opportunity set.

One clever way to circumvent this problem and construct the RAPM Efficient Frontier based on downside risk measure might be the one that makes use of Monte Carlo Simulation. In fact, due to the existence of multiple local minima in the case of downside risk measure when no assumption is made about the distribution of returns, the efficient frontier must be constructed by a complete enumeration of the feasible portfolios⁸³. This methodology although is understandably subject to higher estimation risk is a lot more efficient when dealing with highly

⁸³ This is a computationally demanding task even for a small number of assets

asymmetric and fat tailed return distributions. Other authors (Zagst and al.)⁸⁴ have solved the portfolio optimization problem under the shortfall risk by making use of stochastic optimization procedures, namely mixed integer programming. In this case, stochastic optimisation problems have to be approximated.

7.3 A new RAPM based on distorted risk measures: definition and portfolio management perspectives

In this section a RAPM built on top of a distorted risk measure is analysed within a banking and asset management end user perspective.

To explore how RAP based on downside risk measure like the TVaR and ESF, can lead to paradoxical results and, at the same time, the Wang measure can help identifying the investment alternative with the better risk return profile, let consider the following example and the input data in table 4. Consider two portfolios A and B with assigned losses (X), reported in the first column of table 4, and correspondent probability (Prob f(x)), reported in the second column. The third column represents the product of the first two (losses time probability). Moreover let assume, for the sake of simplicity, that the expected return of both portfolios are the same and equal to 1.

Portfolio A			Portfolio B		
Loss x	Prob f(x)	x*f(x)	Loss x	Prob f(x)	x*f(x)
0	60.0%	0	0	60.0%	0
1	37.5%	0.375	1	39.0%	0.39
5	2.50%	0.125	11	1.00%	0.11

Table 4 Portfolio A and B loss distribution

In this case, how is reported in table 5, results $TVaR(A)=TVaR(B)$. Also the $ESF(A)=ESF(B)$. Therefore any RAP that shows in the numerator a measure of the reward of the investment or business opportunity and at the denominator a measure of downside risk as the ones reported in the example like the TVar or ES will produce the same synthetic number, disregarding all the information in the remaining part of the distribution. As already shown in section 5, distorted risk measure will capture information also contained in the remaining part of the pdf because are compliant with the stochastic dominance principle. Only by using a distorted risk measure, like the Wang distorted risk measure, it is possible to correctly discriminate between the two

⁸⁴ Portfolio Optimization Under Credit Risk, R. Zagst, J. Kehrbaum, B. Schmid, Risk Lab Germany, working paper, 2002.

investment alternatives, because it is clear that portfolio B is dominated by portfolio A. Below follows the table summarizing all risk measures previously analysed.

Portfolio A		Portfolio B	
ESF(α)	1.250	ESF(α)	1.250
TVaR(α)	3.000	TVaR	3.000
WT(α)	2.420	WT	3.400

Table 5 Comparison between downside and distorted risk measure

From a theoretical portfolio management perspective it is clear the advantage of adopting a RAP that take in consideration all information of the portfolio value distribution. In this case, one is able to measure:

1. the likelihood of having a positive returns, information disclosed in the “central” part of the portfolio value distribution (i.e., the expected value or other similar statistics)
2. the riskiness of the considered portfolio value distribution, information disclosed by the shape of the tail of the distribution at the hand.

To better clarify the pro-active portfolio management interesting features of distorted risk measure let consider also the following example.

Let recall from section 5 that a continuous increasing function $g: [0,1] \rightarrow [0,1]$ such that $g(0) = 0$ and $g(1) = 1$ is called *distortion* function. For $X \in \mathcal{X}$ with probability distribution $F_X(x)$, the transform $g(F_X(x))$ defines a distribution function, which is called *distorted distribution* function $\rho_g[X]$ as reported in eq. 5.2. Wang et al.(1997), theorem 3, implies that the risk measure $\rho_g[X]$ is a coherent risk measure provided that $g(x)$ is a concave function⁸⁵.

Now consider eq. 5.2. Let $\rho_g[X]$ defining a coherent distortion measure through the increasing concave distortion function $g_\varepsilon(x) = \min\{\frac{x}{\varepsilon}, 1\}$, where ε is a small probability of loss, say $\varepsilon = 0.05$. To see⁸⁶ how distorted RAP measures can provide superior information, let now consider Y be a loss consisting of two scenarios with loss amounts 20\$, 2100\$ such that $P(Y = 20) = 1 - P(Y = 2100) = \frac{25}{26}$. Through active risk management, assume that the lower amount can be eliminated and that the higher loss amount can be reduced to 1700\$. By equal mean and variance, this results in a loss X such that $P(X = 0) = 1 - P(X = 1700) = \frac{16}{17}$. Suppose a risk manager is weighing the cost of risk management against the benefit of capital

⁸⁵ This implies that (7.3.1) and (7.3.2) are coherent provided that $g(x)$ is a convex ($\gamma(x)$ is a concave) function.

⁸⁶ This example is adapted from Distortion Risk Measures And Economic Capital, working paper Hurlimann (2003) pag.

relief. Then ETL does not promote risk management because $ETL_{\alpha}[X] = 1700 > ETL_{\alpha}[Y] = 20 + 2080 \cdot \left(\frac{20}{26}\right) = 1620$, which shows that there is a capital penalty instead of a capital relief for either removing or reducing the initial loss amounts. However, the Wang measure offers a capital relief because $\rho_g[X] = 1700 \cdot \sqrt{\frac{1}{17}} = 412.3 < \rho_g[Y] = 20 + 2080 \cdot \sqrt{\frac{1}{26}} = 427.9$. Since Y is evidently a higher loss than X , the ETL measure fails to recognize this feature.

Summarising, downside risk measure ignores useful information in a large part of the loss distribution. Therefore, if a risk manager uses a coherent RAP distortion measure, he can adjust more properly extreme low frequency and high severity losses. The intuition is that a RAP built on top of this risk measure should provide more useful and congruent information.

8 Conclusions

In this chapter different measures of risk used in portfolio management have been analysed and compared, distinguishing traditional mean-variance, other two-sided measures based on the entire distribution, probability of shortfall and downside measures of risk based on a part of the distribution.

None of these provides an ideal risk measure, appropriate in all circumstances.

Quantile risk measures are unsatisfactory because they do not satisfy the sub-additivity axiom of coherence. In section 4 it has been shown that, using the p-quantile risk adjusted performance measure (like the RoVaR) for determining a solvency capital is meaningful in situations where the default event should be avoided, but the size of the shortfall is not important. When addressing the solvency capital puzzle bondholders, regulators and, to a certain degree, rating agencies are interested not only in the estimation of the frequency of default, but also of the severity of default. Interestingly, as it will be shown in much more detail in the next chapter, shareholders and top management design their strategies with the main aim of maximising the value of financial institutions. In doing so, they can take a conservative view and therefore will care about designing those risk management (contingent) strategies that will allow the institution to “survive”. However, if the management is experiencing a difficult moment they will try to get the most out of their (risky) investment strategy without caring about the size of the loss in the event of default because of the limited liability assumption. Thus, not only there are different perspectives between bondholders, regulators, rating agency and shareholders and management;

but, ultimately, the behavior of the investors is contingent to the life cycle of the financial institution and the market opportunities.

It is therefore hard to think that a RAPM based only on a p-quantile risk measure can provide answers that can reconcile all these peculiar different points of view. It appears obvious that there exists a need for a more flexible and comprehensive RAPM when moving from a “provision principle” perspective towards the “premium principle” set up, where all the values of the loss (and gain) distribution play a role.

One possible way to attempt to solve this problem is to use RAPMs that show in the denominator downside risk measure. It is argued that RAPMs of this type are shown to be useful when focusing on the provision principle because they show the advantage to be, by construction, more conservative than quantile risk measure. Moreover, in section 4 it was shown that they comply with the sub-additivity axiom, thus showing the interesting feature of coherence. In this case it is possible to use these measures to avoid the problems that the quantile based RAPM is not capable to cope with as illustrated in section 3.2. Unfortunately, from a proactive portfolio risk management perspective these types of RAPM, do not allow to undertake strategies that make it possible to mitigate riskier scenarios by leveraging other possibilities of gains that can be evaluated by considering the whole distribution. RAPM built on top of downside risk measure such as the one based on measure of shortfall risk, conditional tail expectation, expected shortfall and tail VaR, because do not satisfy the second-order stochastic dominance principle cannot be safely used for pro-active portfolio management as shown in section 4.2.5.4 and 7.3

To this aim, I propose the use of distortion concave risk measures developed in the actuarial literature as the basis to build on top of this measure RAPM that can help identify and rank investment alternatives when their distribution is non-normal and fat tail are possible. In section 7.3 it has shown that, if portfolio A and B have the same return, and portfolio A dominates portfolio B in accordance with the SOSD principle, RAPM based on measure of shortfall risk such as the Tail VaR are not able to correctly rank the two portfolios. The distorted concave RAPM, such as the Wang Transform, on the other hand, ranks correctly.

These types of distorted RAPM offer portfolio managers a family of different full distribution measures of risk, which can be adjusted to reflect a range of different attitudes to outcomes in different parts of the distribution by choosing an appropriate distortion function. Examples where the distortion risk measure provides summaries of portfolio performance that are more helpful than any of the other commonly used metrics are provided in section 7.3.

Chapter 2

Summarising, in this chapter it has been shown that, when using RAPM for portfolio management it is crucial to check that the RAPM at the hand it is at the same time coherent and SD compliant. The concave distorted risk measures are a good example to help solve successfully this puzzle. Actuaries have started to use this risk measure also for capital allocation purposes and I think that the finance researchers and professionals will start to consider even more carefully than the past the need for more tailored and consistent RAPM because risk and portfolio management represent the next challenge for risk and asset managers in the near future.

Appendix 1 First Degree Stochastic Dominance Principle**Proposition 1 The FSD rule**

Given two probability distributions $f(x)$ and $g(x)$, distribution f dominates distribution g by FSD if:

1. the decision maker has positive marginal utility of wealth for all x ($u'(x) > 0$)
2. cdf of $f(x)$ is less than or equal to cdf under the $g(x)$ distribution, for all x with strict inequality for some x ⁸⁷.

If f is preferred to g then $\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx > 0$.

Conversely,

if g is preferred to f results $\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx < 0$.

Let define $a=u(x)$ and b the difference between two cdf, namely $b=[F(x)-G(x)]$, where

$$F(x) = \int_{-\infty}^x f(x)dx \text{ and } G(x) = \int_{-\infty}^x g(x)dx$$

Moreover let also define

$$da = u'(x)dx \text{ and } db = [f(x) - g(x)]$$

By integrating by parts results

$$\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx = \int_{-\infty}^{\infty} adb = ab|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} bda \quad (\text{A.1.1})$$

Assume both $f(x)$ and $g(x)$ is the probability associated to each level of wealth for alternatives f and g . It is therefore possible to write the difference in the expected utility between the prospects (see eq. A.1.1) as follows:

$$\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx = [u(x)[F(x) - G(x)]]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx \quad (\text{A.1.2})$$

It is possible to observe a couple of things about this result.

Let look at the first addendum of the right hand part, i.e. $[u(x)[F(x) - G(x)]]_{-\infty}^{\infty}$.

1. First notice that when $x \rightarrow -\infty$, $F(x)=G(x)=0$ because at the far left hand tail of the probability distribution the cumulative probabilities equal zero.

Thus, for $x \rightarrow -\infty$ results $[u(x)[F(x) - G(x)]]_{-\infty}^{\infty} = 0$.

2. Second, when $x \rightarrow +\infty$ since these are cumulative probability distributions both will equal one and therefore the utility of plus infinity times a term that equals one minus one which is zero. Thus, for $x \rightarrow +\infty$ results $[u(x)[F(x) - G(x)]]_{-\infty}^{\infty} = 0$

⁸⁷ Note, the strict inequality requirement means the distribution cannot be the same.

Now let look at the second addendum, namely $-\int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx$

If the overall sign of

$$[u(x)[F(x) - G(x)]]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx$$

is positive then f dominates g.

I will restrict the sign by adding assumptions.

1. First, assume non-satiation i.e., that more is preferred to less and therefore results $u'(x) > 0$ for all x ⁸⁸. This means this term takes its sign from the $F(x) - G(x)$ term. That term gives the difference between the two cdfs.
2. Second, suppose $[F(x) - G(x)] \leq 0 \quad \forall x$ ⁸⁹. Because by assumption cdf of f is less than or equal to cdf under the g distribution, for all x with strict inequality for some x, it will result $[F(x) - G(x)] < 0$

Having shown that the term at the left hand side of eq. A.1.2 is zero and the second term is positive because $[F(x) - G(x)] < 0$ and the marginal utility is positive (with a negative sign in front of it) results

$$[u(x)[F(x) - G(x)]]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx \geq 0 \text{ for some } x$$

⁸⁸ Thus $u'(x)$ it will always be a positive multiplier

⁸⁹ This means that the cdf of f must always lie on to the right of the cdf of g

Appendix 2 Second Degree Stochastic Dominance**Proposition 2 The Second Order Stochastic Dominance rule**

If

- 1) An individual has positive marginal utility (non-satiation), $u'(x) > 0$
- 2) An individual has diminishing marginal utility of income $u''(x) < 0$
- 3) $[F_2(x) - G_2(x)] \leq 0$ for all x , with strict inequality for some x

Then

f dominates g by a SDSD.

Proof

If risk aversion matters the FOSD decision rule $\int_{-\infty}^{\infty} u(x)[f(x) - g(x)]dx$ is replaced by the SOSD

$$\int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx = adb = -[u'(x)[F_2(x) - G_2(x)]]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} u''(x)[F_2(x) - G_2(x)]dx \quad (\text{A.2.1})$$

By applying the integration by parts formula and defining for convenience

$$a = u'(x)$$

and

$$db = [F(x) - G(x)]dx$$

so that

$$da = u''(x)dx$$

and

$$b = [F_2(x) - G_2(x)]dx,$$

where $F_2(x)$ and $G_2(x)$ are the second integral of F and G with respect to x , i.e.:

$$F_2(x) = \int_{-\infty}^x \int_{-\infty}^x f(x)dx = \int_{-\infty}^x F(x)dx \quad (\text{A.2.2})$$

Integrating by parts results

$$\int_{-\infty}^{\infty} u'(x)[F(x) - G(x)]dx = -[u'(x)[F_2(x) - G_2(x)]]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} u''(x)[F_2(x) - G_2(x)]dx$$

Therefore, f dominates g under the SOSD rule if eq.(A.2.1) is always positive. To see this, consider the right hand part of (A.2.1). SDSD makes two assumptions that render this term positive:

1. First, assume that the second derivative of the utility function with respect to x is negative everywhere, i.e. $u''(x) < 0 \forall x \in \mathfrak{R}$
2. Second, assume that $[F_2(x) - G_2(x)] \leq 0$ $F_2(x)$ for all x with strict inequality for some x .

Under these circumstance results a negative number times a negative leading to a positive⁹⁰, therefore

$$\int_{-\infty}^{\infty} u''(x)[F_2(x) - G_2(x)]dx \geq 0$$

Consider now the left hand part of (A.2.1). If

3. the assumption on non-satiation is now added, i.e. $u'(x > 0), \forall x \in \mathfrak{R}$, because

$[F_2(x) - G_2(x)] \leq 0$, results, for

- $x \rightarrow +\infty$ results $[F_2(x) - G_2(x)] \leq 0$
- $x \rightarrow -\infty$ results $[F_2(x) - G_2(x)] \cong 0$, since there is no area at that stage.

Therefore, also the left hand side term is positive

$$- [u'(x)[F_2(x) - G_2(x)]]_{-\infty}^{\infty} \geq 0$$

Consequently

$$- [u'(x)[F_2(x) - G_2(x)]]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} u''(x)[F_2(x) - G_2(x)]dx \geq 0$$

Remarks

- a. If $u''(x) < 0$ and $u'(x) > 0$ then the Pratt risk aversion coefficient is positive
- b. The assumption that the integral under the cdf of f must be smaller than the integral under the cdf of g allows the cumulative distributions to cross as long as the difference in the areas before they cross is greater than the difference in their areas after they cross.

⁹⁰ For second degree dominance to apply therefore the utility function must be non-decreasing and concave.

Appendix 3 The relationship between downside and maximum potential loss based risk measure

VaR calculates under a certain confidence level k the maximum loss (or the maximum loss return τ) over a given time horizon T . Conversely, $LPM(0;\tau)$ represents the probability of failing an ex ante defined target return τ . Due to their similarity, it is easy to derive a formal relationship between these two risk measures - despite the fact that they follow different goals. Given that $LPM(0;\tau)$ corresponds exactly to the confidence probability α of the considered VaR equation, with $\alpha=0$ is taken as a starting point. Assuming the existence of a corresponding density function for the distribution function $F(x)$, results

$$LPM_0(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^0 f(X) dX = [F(X)]_{-\infty}^{\tau} = F(\tau) - F(-\infty) = F(\tau) = k$$

The transition to the loss return can be executed by using the inverse function of $F(x)$:

$$F^{-1}(k) = F^{-1}[LPM_0(X, \tau)] = \tau$$

The VaR can be equivalently expressed as a continuously compounded loss return τ . In fact, The maximum loss L can be transformed into the continuously compounded loss return

$$\tau = \ln\left(1 - \frac{L}{W}\right) \text{ by using the relation } L = W(1 - e^{\tau}), \text{ with } W \text{ the initial wealth.}$$

Therefore:

$$VaR_k = L = W(1 - e^{\tau})$$

and consequently, the relationship between VaR and $LPM_0(X, \tau)$ is:

$$VaR_k = L = W(1 - e^{\tau}) = W(1 - e^{F^{-1}(k)}) = W(1 - e^{F^{-1}(LPM_0(X, \tau))})$$

Therefore, a VaR can be calculated on the basis of $LPM_0(X, \tau)$ and vice versa. It is easy to verify that the VaR decreases with increasing shortfall-probability.

This relationship can be exemplified with a standard-normally distributed random variable x . The

negative argument of the distribution, $z \equiv \frac{\tau - 0}{\sigma}$, approaches to zero with increasing shortfall

probability⁹¹. Therefore in the limit, $z \rightarrow 0$, implies $e^z = 1$. The implication is that both the term in brackets and the VaR disappear.

An extension of the classical VaR-framework can be illustrated by considering the target shortfall for a target return τ . Recalling that

⁹¹ The possibility of a negative VaR is excluded

$$LPM_1(\tau, X) = \int_{-\infty}^{\tau} (\tau - X)^1 f(X) dX = E[\text{Max}(\tau - X, 0)]$$

It can be shown that $LPM_1(\tau, X)$ permits a quantitative judgement about the expected loss, if the effective portfolio loss exceeds the calculated VaR. Based on the target return τ , which corresponds to the pre-specified confidence probability k , the shortfall-conditioned mean $\mu_{x \leq \tau}$ can be calculated using the standardized target shortfall previously introduced

$$\mu_{x \leq \tau} = \tau - \frac{\int_{-\infty}^{\tau} (\tau - X)^1 f(X) dX}{\int_{-\infty}^{\tau} f(X) dX}$$

Through this equation it is possible to quantify the expected loss of a portfolio, if the classical VaR-forecast underestimates the effective loss realization and insufficiently captures the existing risk. This loss expectation is therefore called **Shortfall-Value at Risk (SVaR)**, since it estimates the average wealth reduction, if the effective loss exceeds the forecasted loss by VaR:

$$VaR_{K, x \leq \tau} = L_{x \leq \tau} = W(1 - e^{\mu_{x \leq \tau}}) = W \left(1 - e^{\left\{ \tau - \frac{\int_{-\infty}^{\tau} (\tau - X)^1 f(X) dX}{\int_{-\infty}^{\tau} f(X) dX} \right\}} \right) = SVaR(k, x \leq \tau)$$

Appendix 4 Example SR vs GSR⁹²

Let consider two investments A and B. Suppose that returns and related probabilities are those reported in table 7 Clearly investment A dominates investment B. By comparing SR vs GSR (see also table 8 and 9 and figure 9 and 10) it is possible to note how, when returns are non-normally distributed the GSR correctly ranks the considered alternative investments. SR instead flawed. Investors, fund and asset managers, risk managers, regulators must carefully consider the problems which arise when coherence assumptions, on one hand, and the statistical properties of estimated parameters considered in the used RAPM, on the other hand, are not carefully taken in consideration when evaluating or structuring both financial and strategic investments or financial products (especially the most illiquid ones).

Probability	R_A	Probability	R_B	$E[R_A]$	$E[R_A^2]$	$E[R_B]$	$E[R_B^2]$
0.01	-13.00	0.01	-10.00	-0.10	1.00	-0.13	1.69
0.04	-7.00	0.04	-5.00	-0.20	1.00	-0.28	1.96
0.25	-1.00	0.25	-1.00	-0.25	0.25	-0.25	0.25
0.40	5.00	0.40	5.00	2.00	10.00	2.00	10.00
0.25	11.00	0.25	15.00	3.75	56.25	2.75	30.25
0.04	17.00	0.04	28.00	1.12	31.36	0.68	11.56
0.01	23.00	0.01	32.00	0.32	10.24	0.23	5.29
Mean R_A	6.64	Mean R_B	5.00	6.64	110.10	5.00	61.00
Std dev R_A	8.12	Std dev R_B	6.00				
SR R_A	0.83	SR R_B	0.82				
GSR R_A	0.82	GSR R_B	0.95				

Table 7 Sharpe Ratio and Generalised Sharpe Ratio of Investment A and B

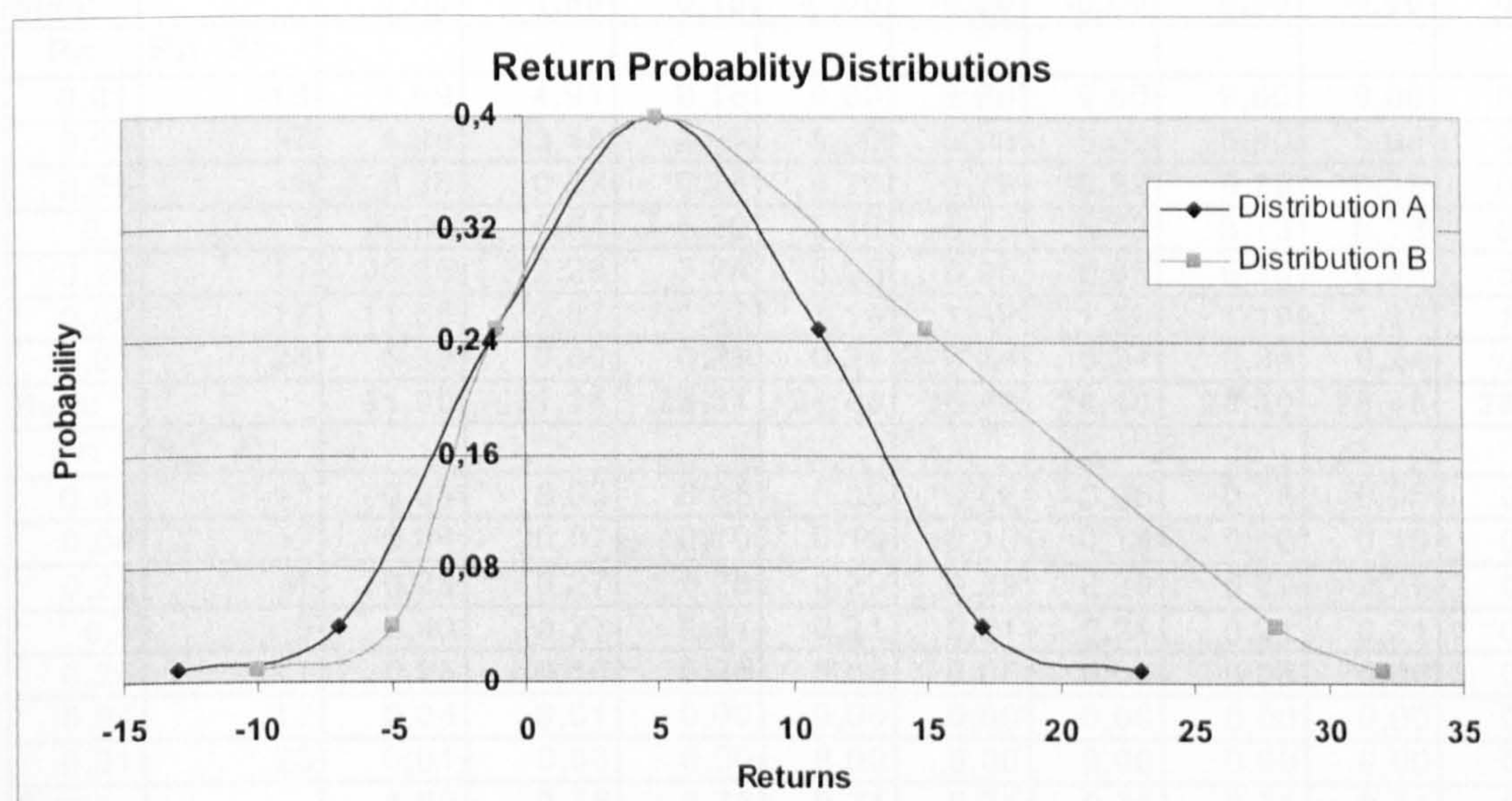


Figure 9 SR vs GSR: Return Probability Distribution Function of Investment A and B

⁹² The author wishes to thank S. Hodges for providing the worksheets illustrated in table 7,8 and 9

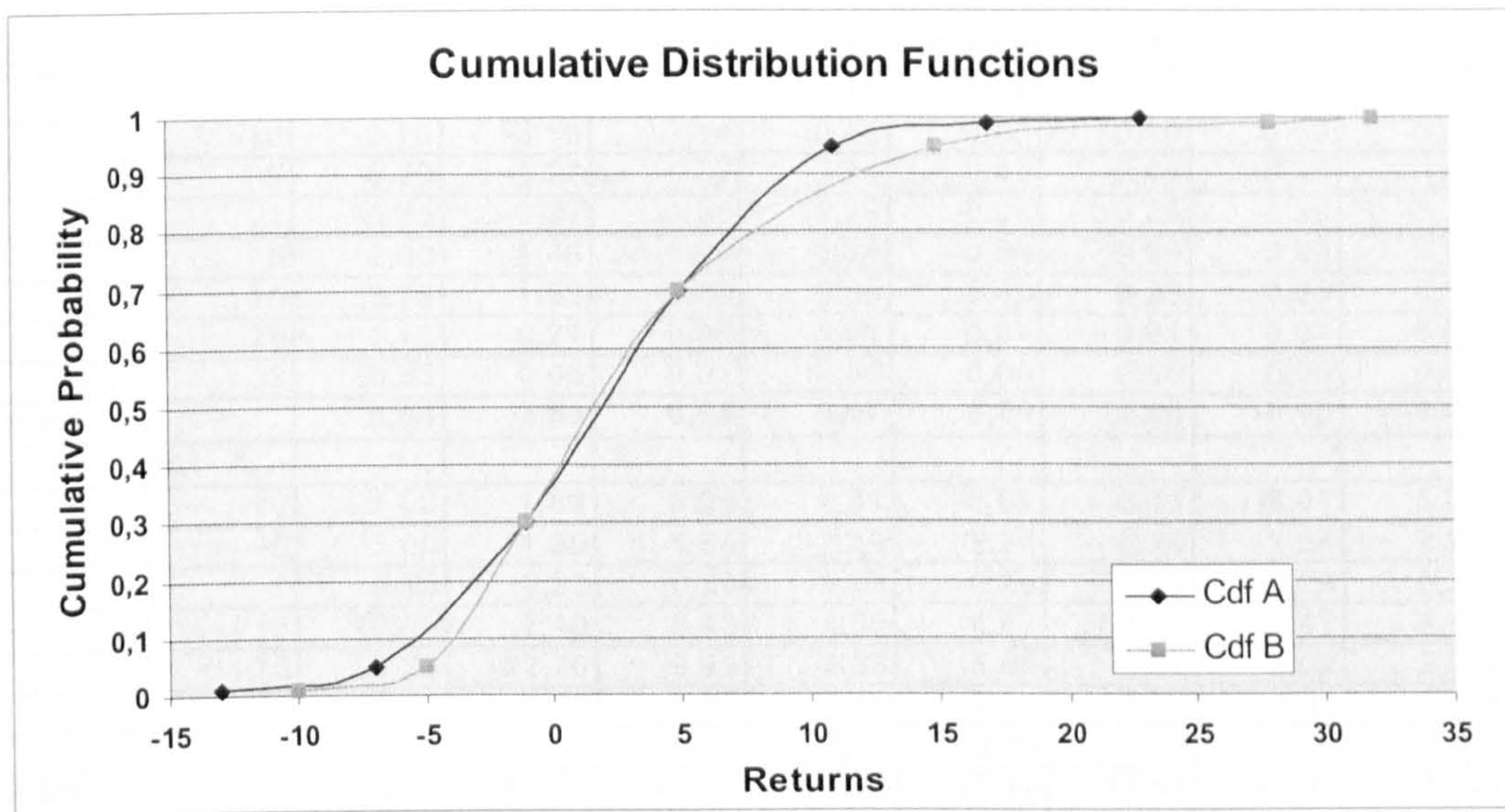


Figure 9 SR vs GSR: Cumulative Distribution Function of Investment A and B

Worksheet for calculating Generalized Sharpe Ratio R_A										
Pr:	$R_A \setminus X:$	0,00	0,08	0,13	0,13	0,13	0,13	0,13	0,13	0,13
0,01	-13	-0,13	-0,38	-0,70	-0,74	-0,74	-0,74	-0,74	-0,74	-0,74
0,04	-7	-0,28	-0,50	-0,70	-0,71	-0,71	-0,71	-0,71	-0,71	-0,71
0,25	-1	-0,25	-0,27	-0,28	-0,29	-0,29	-0,29	-0,29	-0,29	-0,29
0,4	5	2,00	1,33	1,04	1,03	1,03	1,03	1,03	1,03	1,03
0,25	11	2,75	1,12	0,66	0,63	0,63	0,63	0,63	0,63	0,63
0,04	17	0,68	0,17	0,07	0,07	0,07	0,07	0,07	0,07	0,07
0,01	23	0,23	0,03	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Sum:		5,00	1,50	0,10	-0,00	-0,00	-0,00	0,00	-0,00	0,00
Pr:	$R_A \setminus X:$									
0,01	-13	1,69	4,91	9,16	9,60	9,60	9,60	9,60	9,60	9,60
0,04	-7	1,96	3,48	4,87	5,00	5,00	5,00	5,00	5,00	5,00
0,25	-1	0,25	0,27	0,28	0,29	0,29	0,29	0,29	0,29	0,29
0,4	5	10,00	6,64	5,22	5,13	5,13	5,13	5,13	5,13	5,13
0,25	11	30,25	12,28	7,24	6,95	6,95	6,95	6,95	6,95	6,95
0,04	17	11,56	2,87	1,27	1,19	1,19	1,19	1,19	1,19	1,19
0,01	23	5,29	0,80	0,27	0,24	0,24	0,24	0,24	0,24	0,24
Sum:		61,00	31,24	28,31	28,40	28,40	28,40	28,40	28,40	28,40
Pr:	$R_A \setminus X:$									
0,01	-13	0,01	0,03	0,05	0,06	0,06	0,06	0,06	0,06	0,06
0,04	-7	0,04	0,07	0,10	0,10	0,10	0,10	0,10	0,10	0,10
0,25	-1	0,25	0,27	0,28	0,29	0,29	0,29	0,29	0,29	0,29
0,4	5	0,40	0,27	0,21	0,21	0,21	0,21	0,21	0,21	0,21
0,25	11	0,25	0,10	0,06	0,06	0,06	0,06	0,06	0,06	0,06
0,04	17	0,04	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,01	23	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Sum:		1,00	0,75	0,71	0,71	0,71	0,71	0,71	0,71	0,71
	SR	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83
	GSR	0,00	0,76	0,82	0,82	0,82	0,82	0,82	0,82	0,82

Table 8 SR and GSR of Investment A

Worksheet for calculating Generalized Sharpe Ratio R_B										
Pr:	$R_B \setminus X:$	0,00	0,06	0,12	0,16	0,16	0,16	0,16	0,16	0,16
0,01	-10	-0,10	-0,18	-0,34	-0,48	-0,50	-0,50	-0,50	-0,50	-0,50
0,04	-5	-0,20	-0,27	-0,37	-0,44	-0,45	-0,45	-0,45	-0,45	-0,45
0,25	-1	-0,25	-0,27	-0,28	-0,29	-0,29	-0,29	-0,29	-0,29	-0,29
0,4	5	2,00	1,48	1,09	0,91	0,89	0,89	0,89	0,89	0,89
0,25	15	3,75	1,52	0,60	0,36	0,33	0,33	0,33	0,33	0,33
0,04	28	1,12	0,21	0,04	0,01	0,01	0,01	0,01	0,01	0,01
0,01	32	0,32	0,05	0,01	0,00	0,00	0,00	0,00	0,00	0,00
sum:		6,64	2,53	0,74	0,07	0,00	0,00	-0,00	-0,00	-0,00
Pr:	$R_B \setminus X:$									
0,01	-10	1,00	1,83	3,39	4,81	5,01	5,01	5,01	5,01	5,01
0,04	-5	1,00	1,35	1,84	2,19	2,24	2,24	2,24	2,24	2,24
0,25	-1	0,25	0,27	0,28	0,29	0,29	0,29	0,29	0,29	0,29
0,4	5	10,00	7,40	5,43	4,56	4,47	4,47	4,47	4,47	4,47
0,25	15	56,25	22,76	8,99	5,33	5,02	5,02	5,02	5,02	5,02
0,04	28	31,36	5,79	1,02	0,39	0,34	0,34	0,34	0,34	0,34
0,01	32	10,24	1,49	0,20	0,07	0,06	0,06	0,06	0,06	0,06
Sum:		110,10	40,89	21,17	17,64	17,43	17,43	17,43	17,43	17,43
Pr:	$R_B \setminus X:$									
0,01	-10	0,01	0,02	0,03	0,05	0,05	0,05	0,05	0,05	0,05
0,04	-5	0,04	0,05	0,07	0,09	0,09	0,09	0,09	0,09	0,09
0,25	-1	0,25	0,27	0,28	0,29	0,29	0,29	0,29	0,29	0,29
0,4	5	0,40	0,30	0,22	0,18	0,18	0,18	0,18	0,18	0,18
0,25	15	0,25	0,10	0,04	0,02	0,02	0,02	0,02	0,02	0,02
0,04	28	0,04	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,01	32	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Sum:		1,00	0,74	0,65	0,63	0,63	0,63	0,63	0,63	0,63
	SR	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82
	GSR	0,00	0,77	0,93	0,95	0,95	0,95	0,95	0,95	0,95

Table 9 SR and GSR of Investment B

Chapter 3

Default probability, capital structure and the cost of economic capital¹

*There is a history in all men's lives figuring the natures of the times deceased;
the which observed, a man may prophesy, with a near aim,
of the main chance of things as not yet come to life*

William Shakspeare, *King Henry IV*, Part II, Act 3, Scene 1

¹ We thank Giovanni Barone Adesi, Anthony Saunders, Constantine Thanassoulas, participants at the Economic Capital Roundtable at Cass Business School, City University for helpful discussion. We also thank participants at Risk Capital Conference Geneva 2004; PRMIA Greece Conference, Athens 2004; Economic Capital ICBI Conference Nice 2004, and Credit Risk Summit London 2004. Please note this is a joint research paper with Alistair Milne

1- Introduction

One of the most important developments in risk management over the past few years has been the widespread implementation of risk-adjusted performance measurement (RAPM)² systems, designed to put the returns of very different risk business on a comparable, risk-adjusted basis. The main benefit of such systems is that they help financial institutions to identify businesses where they have a competitive advantage. The immediate purpose of such risk measurement system is to provide bank managements with a more reliable way to determine the amount of capital necessary to support each of their major activities, and thus, to determine the overall leverage for the bank as a whole. The ultimate goal of these risk based capital allocation systems is to provide a uniform measure of performance. Management can in turn use this measure to evaluate performance for capital budgeting, risk management, optimal capital structure and as an input to the compensation system used for senior managers.

Allocating equity capital on the basis of the risk of individual business units seems pointless in the classical theoretical paradigm of “frictionless” capital markets (one with perfect information and without taxes, bankruptcy cost or conflicts between managers and shareholders). If markets operated in this manner, the pricing of specific risks would be the same for all banks and would not depend on the characteristic of an individual bank’s portfolio. Moreover, given market prices of risk, whether a bank varied leverage on the basis of risk or varied the cost of capital with the risk of the project would result in the same capital budgeting decisions and investment activity for the bank. Of course, in real world, banks and other financial intermediaries add value precisely through their ability to reduce market frictions. This role implies that a large portion of bank assets are likely to be difficult for outside investors to value, which in turn may create information and agency problems for banks themselves when they have to raise capital externally. In this situation systems of RAPM can add value.

This chapter provides the combined first analysis of the impact of default probability, capital structure, limited liability, deposit insurance, franchise value and taxes on Risk Adjusted Return on Capital (RAROC)³ type risk measures. Previous work of Crouhy and al (1999) has addressed part of this problem but does not provide a complete analysis. The suggested model

² A generic RAPM takes the following form, $RAPM = \frac{\text{Expected Revenues} - \text{Costs} - \text{Expected Losses}}{\text{Economic Capital}}$, where: Expected

Revenues is the expected revenues assuming no losses, Expected Losses is the expected losses from default, and Economic Capital is usually defined as the capital necessary to cushion against unexpected losses, operating risks and market risks, and is often referred to as Value at Risk (VaR). For a more detailed description of this measure see Ong 1999, pag 218. The magnitude of the Economic Capital is usually determined so that the probability of unexpected losses is below some specified level. Matten (2000, pp 146-166) describes several different RAPM.

³ In this framework only the ex-ante perspective will be considered. The cost of economic capital is therefore based on expectations and risk assessment

also adopts a rigorous equilibrium perspective by enforcing the pricing of banking assets using the Capital Asset Pricing Model (CAPM). This alters significantly some of the results reported by Crouhy and al. (1999). Finally, this framework clearly establishes the relationship amongst different measures of capital, namely the market value of the equity, the risk capital and the economic capital.

This chapter is organised as follows. The next section Section 2 reviews the most relevant literature. The third Section 3 discusses the different concepts of 'capital' and how these relate to both performance measurement and value creation, illustrates the theoretical framework by considering their relationship from the perspective of both bond holders and shareholders.

The core analysis of this chapter is presented in Section 4, linking the economic capital to the performance measurement, i.e how to enhance (maximise) the shareholder value of the financial institution through the analysis of the cost of economic capital. Section 4.1 introduces the core assumptions of the suggested model. Because, in this framework, assets are measured on a mark-to-market basis market and book capital are the same by construction. In Section 4.2 the distinction between actual and target default probability is then used to show how a bank can change the capital structure to reach the pre-defined credit rating that allows the bank itself to ultimately reach the cost of capital target. It will be shown that the most important risk driver is the market correlation and not the volatility as previously reported by Crouhy and al (1999). Sections 4.3 and 4.4 introduce shareholder limited liability and the possibility that deposits are insured. When considered the put option associated with the insurance then the asset return volatility plays a role. The impact of volatility on the cost of capital is higher the riskier is the bank or, stated alternatively, the higher the existing bank leverage. The resultant cost of capital is calculated also in the case that the target credit rating is considered. Section 4.5 summarises and compares the results obtained in the various sub-sections 4.2-4.4. Section 5 generalises the analysis by adding the franchise value and the taxes. When all adjustments are made the risk capital becomes the economic capital and can be immediately compared against the market hurdle rate. Section 6 provides a further analysis of the situation when actual capital exceeds the risk capital associated with the target level of probability. Section 7 concludes.

2- Literature Review

According to Schroeck (2002) since early 1990s the financial institutions have been very focused on designing systems to measure the risk that is involved in their different lines of

business. The purpose of such measurement systems is to determine the amount of capital that is necessary for each business unit and thus determine the equity capital required by the bank as a whole. In this framework the bank can define the profitability of businesses with different capital requirements and different sources of risk. These capital-based capital allocation systems are known as Risk Adjusted Return On Capital or RAROC. Development of the RAROC methodology was started in the late 1970s by a group at Bankers Trust. Before the development of risk-adjusted profitability measurements, the banks used Return On Assets (ROA) or Return On Equity (ROE) to measure their performance and assigned capital to different businesses according to the regulatory requirements-Basel guidelines. But neither of these approaches took into account the different kind of risk that are embodied in the bank's various activities.

James (1996) argued that the capital budgeting process resembles the operation of an internal capital market in which firms allocate capital with the objective of mitigating the costs of external financing. RAROC systems allocate capital for two reasons: a) risk management and b) performance evaluation.

For risk management purposes, RAROC assigns capital to individual business units in order to determine the bank's optimal capital structure i.e. the proportion of equity to assets that minimizes the bank's cost of funding. This process involves estimating how much the volatility of each business unit contributes to the total risk of the bank. There are two factors in determining a business's risk contribution: a) the total volatility of its returns⁴, b) the extent of the correlation of its returns with the returns of the overall bank⁵. Therefore financial institutions try to increase value by reducing market frictions and choosing the optimal capital structure. Most of the banks intend to have high leverage because of the benefits they can have. One of the most important benefits is the tax shield provided by tax-deductible interest payments. Another motive to use debt financing and reduce equity is the potential of high leverage, especially in industries with excess capital, to motivate managers to operate as efficiently as possible by accepting projects that add value and rejecting projects that they will not give a desirable return. Also, the financial institutions should take into account the costs that the high leverage involves. Analytically, banks are imposed to heavy costs and liquidity constraints if they violate the minimum capital standards, which will cause major disruptions in the operating activities⁶. Another important issue is that the customers, that they are also the largest liability holders, are very concerned about the bank's credit rating. If a financial institution fails to meet the credit quality standards, it will probably lose important customers and face problems in its operational activities. To maintain a high credit rating the financial

⁴ The more volatile the unit's performance, the more capital is assigned to it.

⁵ If its returns are less than perfectly correlated with the bank's returns its economic capital is reduced.

⁶ In extreme cases, extensive use of debt financing can lead to default and a costly reorganization

institution evaluates the capital adequacy by measuring the volatility of the assets' market value and not the book value. The credit ratings are influenced by the volatility of the variables that affect bank stock prices rather than the changes in earnings and book capital⁷.

For performance evaluation purposes, the goal of allocating capital to individual business units is to determine the economic profit of each unit and its contribution to shareholder value. This process provides a basis for effective capital budgeting and incentive compensation. RAROC systems are used to measure the performance of each business unit. If RAROC of a particular business is higher than the cost of equity-shareholders' minimum required rate of return-the unit creates value for the shareholders. On the contrary, if RAROC is less than the cost of equity, the unit reduces the shareholders' value. But RAROC does not provide information about the amount of value that is being created or destroyed by an operation and thus the use of RAROC can lead to underinvestment. If managers are rewarded solely on the basis of RAROC, they might reject projects with positive Net Present Value that will lower their average returns. To avoid this problem and create the right incentives for the managers, performance should also take into account the sign of the economic profit (i.e. earnings less a charge for the cost of capital).

As Froot and Stein (1995) pointed out, faced with an increasing cost of raising external funds banks will behave in a risk-averse fashion. Specifically, a business unit's contribution to the overall cash flow volatility of the bank will be an important factor in the capital budgeting decision. Moreover, in this environment capital structure, risk management and capital budgeting are inextricably linked together. Finance theory suggests that, in designing a capital allocation system, the first step is to identify the cost and benefits of holding equity capital in the context of these market frictions. In banking, as in most industries, the tax shield provided by tax-deductible interest payments (as opposed to non-deductible dividends) creates an incentive to make extensive use of debt financing. Banks' access to fixed-rate deposit insurance also makes debt in the form of deposits a low-cost source of funding⁸. The advantages of debt financing for commercial banks suggest, that holding a large capital buffer will be costly. The most serious deterrent to high leverage in banking is the possibility for liquidity constraints to cause major disruptions on a bank's operating activities.

⁷ Although, many of the bank's businesses do not mark their portfolios to market on a regular basis and as a result the volatility of market values cannot be defined. For these businesses, the volatility in economic earnings is used instead.

⁸ When combined with this federal insurance subsidy, depositors' further reduction of their required interest rates for the liquidity and convenience of demand and time deposits is an added incentive for banks to use this form of leverage

Froot and Stein (1998) stressed the relation between risk management, capital budgeting, and capital structure policies for banks⁹. In their model all three of these policies are shaped by two related primitive frictions. First, it is costly for banks to hold a buffer stock of equity capital on the balance sheet, even if this equity is accumulated over time through retained earnings. Given these frictions, bank value-maximisation implies the following conclusions. Banks should: hedge any risks that can be offloaded on fair-market terms; hold some capital as a device for absorbing those illiquid risks which cannot be hedged, but the optimal amount of capital is limited; value, given limited capital, illiquid risks much as an individual investor would – according to their impact on overall portfolio risk and return- with the degree of risk aversion being a decreasing function of the amount of capital held.

Merton and Perold (1993) argued that, banks and other financial firms can be distinguished from industrial companies by the fact that their customers are often also their largest liability holders. And because these customers place a premium value on assurances of performance on their contracts, they show a strong preference for banks with a high credit quality. As a consequence, a high credit rating is generally held to be essential for bank to be major swaps dealer, to underwrite securities or to compete effectively in the corporate banking market.

In sum, the benefits of debts financing for banks suggests that there are cost associated with holding a lot of capital. This implies that risk-management concerns will enter into capital budgeting decisions¹⁰. This has two important implication for the design of a capital allocation system. First, when evaluating the risk of a new project or business unit the project's contribution on the overall variability will affect the project's hurdle rate or cost of capital. A second implication of these market frictions is that the bank should hedge all tradable risk— risk that can be hedged at little cost in the capital market. This implication follows directly from the fact that the bank's required price for bearing tradable risk will exceed the market price for the risk by the contribution of a hedgeable risk to the overall variability of the bank's portfolio. The only risk the bank should assume are illiquid or non-tradable risk in which it has a comparative advantage in bearing.

But what would happen if the RAPM systems did not correctly adjust for risks? What if its results were fundamentally biased relative to the way that the market actually compensated risks through higher returns? Then the RAPM evaluations would also be biased, leading to

⁹ Because capital budgeting in the Froot and Stein model is driven by an exposure's impact on portfolio risk and return, decentralised implementation becomes a problem if the pre-existing portfolio cannot be treated as fixed for all time. Thus, for example, if several new investment opportunities arrive at once, the optimal allocation to each must be jointly determined. This requires a well informed central planner operating out of the headquarters. The whole issue of centralisation versus decentralisation should also be investigated with respect of full versus asymmetric information and in presence of capital adequacy restriction

undercapitalise business in which it truly had a competitive advantage and overcapitalise those units with mediocre or even negative performance.

Crouhy, Turnbull and Wakeman (1999) shown that the market's standard RAPM system, namely RAROC, is fundamentally biased. In the following a biased RAPM system is the one that does not make returns comparable even in efficient market. Wilson (1992) showed that for most risk portfolios there is a trading strategy which will generate an infinite RAROC. If both the Capital at Risk (CAR) and the portfolio return are correctly calculated, Wilson (1992) showed that the bias comes from the RAROC calculation itself. In fact, RAROC calculation rule implicitly assumes that only the portfolio risk capital is compensated, rather than the economic capital invested in a portfolio¹¹. Froot and Stein (1998) showed that, in the common RAROC framework, at each investment under consideration is allocated a certain amount of capital. Multiplying the allocated capital by the cost of capital yield a capital charge. The hurdle rate for the investment is then a risk-less rate plus the capital charge. As also shown by Wilson (1992) and James (1996), traditional RAROC can be thought of as a one factor pricing model¹², the factor being the market¹³. Froot and Stein argued that the hurdle rate for small investments in new project¹⁴ can be more efficiently decomposed in a two factor hurdle rate, namely: the covariance of the return on the tradable component with the market R_m and the correlation on the non tradable component of the new risk (μ_i^N) with the non-tradable risks of the existing portfolio¹⁵. In their work Froot and Stein highlighted the ways in which one can go wrong, at least relative to their model's implications for value maximising behaviour, in applying a

¹⁰ Stoughton and Zechner (1999) argued that discussions with practitioners support the view that corporates CFOs as well as their counterparts at financial institutions view capital budgeting, risk management and capital structure as being inextricably linked

¹¹ Consider a generic portfolio that has a non-zero, strictly positive, mark-to-market value, and the risk manager has access to a risk-free asset investment opportunity. It is easy to show that, under these weak conditions, an infinite positive RAROC can be achieved by liquidating the entire portfolio immediately and investing the proceeds in the risk-free asset. This strategy would generate positive realised returns with no capital at risk, implying an infinite RAROC

¹² In this model the risk driver factor is the market: $\mu_i = \gamma \text{cov}(\mu_i, R_m) + \lambda \text{cov}(\mu_i, R_P)$, where

γ = market unit price of risk for the (market) priced factor

R_m = return of (market) price factor

λ = unit cost for volatility of the bank's cash flows

R_P = return on the bank's existing portfolio

μ_i = hurdle rate or required return for project i

¹³ The market is by definition a tradable component and this is the reason of the absence of the super-script T

¹⁴ The two factor model can be represented as: $\mu_i = \gamma \text{cov}(\mu_i^T, R_m) + G \text{cov}(\mu_i^N, R_P^N)$, $\mu_i = (1) + (2)$, where: component (1) represents the first factor and, like the standard CAPM, describes the covariance of the return on the tradable component (μ_i^T) with the market R_m ; component (2) represents the second factor and, like in the modified version of CAPM which considers the non marketable asset, is based on the correlation on the non tradable component of the new risk (μ_i^N) with the non-tradable risks of the existing portfolio (R_P^N) , multiplied by G (the bank's risk aversion factor that depends on the Economic Capital).

“single factor” RAROC to capital budgeting. In particular they argued that there are three main possible source of errors: one factor model are inadequate to separate priced and non-priced risks¹⁶, capital allocation is based on measures of variance rather than covariance¹⁷ and it is likely that an incorrect cost of capital is estimated and used¹⁸.

Crouhy, Turnbull and Wakeman (1999) argued that the definition of risk inside the financial institutions has moved away from a market-driven definition of risk to a measure of risk that is purely firm specific. In particular, one of the most appealing features of the traditional RAROC methodology is that of avoiding to calculate the beta for the business because it is able to measure the inherent risk by using only the current probability of default of the bank. Consequently, if the RAROC number is greater than the firm’s cost of equity capital the business will add value to the firm. Zaik et al. (1996) state the Bank of America’s policy is to capitalise each of the business units in a manner consistent with bank’s desired credit rating on the unit’s stand alone risk¹⁹. This assumes that the risk of economic capital of the stand-alone business is the same as that of the bank’s equity. Consequently, the implicit assumption is that the RAROC measure adjusts the risk of a business to that of a firm’s equity²⁰. Crouhy, Turnbull and Wakeman (1999) showed that maintaining the probability of default constant is inconsistent with a constant expected return on equity and vice versa thus showing that RAROC cannot be directly compared with the firm’s cost of equity capital. According to Zaik and al. and James (1996), state that Bank of America’s policy is to use a common hurdle rate (cost of equity capital) for all the business units. The main reason is that the portion of the overall risk of the single business units is diversified away by the structure of the bank as a whole, which means that the risk contribution of each unit is based on an internal beta. As a

¹⁵ Extending to many new products means we have to incorporate the covariance of each new non-tradable exposure with other exposure, the optimal mix of new and old investments will thus depend on this variance covariance matrix

¹⁶ One of the key implications of the Froot and Stein two factor model is that a bank should evaluate investments according to both their correlation with any priced market factors and their correlation with the bank existing portfolios.

¹⁷ Under traditional RAROC the amount of capital allocated is related to a measure of the investment’s risk such as its total volatility, but from prospect of a bank value maximization it makes more sense to be driven by the covariance with the rest of the bank’s portfolio. Many financial institutions have now recognised the usefulness of this approach. For example James (1996), in discussing the implementation of RAROC at Bank of America writes “the amount of capital allocated varies with the contribution of the project to the overall volatility of earnings at Bank of America (the projects so called internal beta)”

¹⁸ Even if the RAROC capital allocation is based on the appropriate covariance measure, one need to come up, first, with the right measure of bank tolerated default probability (For example according to Zaik and al. (1996) at Bank of America a default probability of 0,03% is tolerated), and, second, with the appropriate cost of capital. Following Zaik and al. (1996) the hurdle rate is typically set to equal the required return on equity of the bank’s shareholders, which could be calculated from the standard CAPM. This latter calculation is inconsistent within the Froot and Stein framework. In fact, in the polar case where the bank hedges all risks, the bank’s shareholders are left only holding non-priced risks, so their required return on equity is simply the risk-less rate. But, if this is true, the traditional one factor RAROC method says that the capital charge should be zero for any target default probability and cost of equity capital value.

¹⁹ If a Business Unit’s RAROC is higher than the cost of capital (or the cost of bank’s equity) –the minimum rate of return required by shareholders- then the unit is deemed to be adding value to shareholders.

²⁰ If a firm is considering investing in a business, it can compute the RAROC for the business and compare it with the firm’s cost of equity capital.

result the application of a single hurdle rate will produce results very similar to those generated by CAPM system with multiple discount rates²¹. Especially if the bank holds a well diversified portfolio of businesses, the risk of various activities on a levered basis will be the same and the internal betas will not be different from the market-based betas. Another reason is that as the cost of equity should reflect the systematic risk of every activity, which is measured by its CAPM beta, it is very difficult and costly to estimate betas for individual businesses. This process implies that investments in riskier projects will require less leverage than investments in less risky projects. They argued that this framework produces two types of errors: accepting “high-risk” projects that will decrease the value of the firm and rejecting “low-risk” projects that will increase the value of the firm. In order to overcome this distortion they suggested the use of the Adjusted RAROC which is defined as: $Adj.RAROC = \frac{RAROC - R_f}{\beta_E}$, where β_E is the

systematic risk of equity and R_f is the risk-free interest rate²². They declare that if the Adjusted RAROC of a project is greater than the excess rate of return on the market ($R_M - R_f$), it will increase the shareholder value. On the other hand if it is less than the excess rate of return on the market, it will decrease the shareholder value. They also observed that Adjusted RAROC is insensitive to changes in volatility and correlation. But in their analysis they didn't take into consideration the distinction between the tradable and non-tradable risk factor component introduced by Froot and Stein. This means that even the Adjusted RAROC is an inefficient Risk Adjusted Performance Measurement if the tradable component of the undertaken project is not diversified away. Zaik and al. (1996), while recognising the errors that may occur, argue that the costs of estimating individual betas for different businesses outweigh the benefits. This line of reasoning raise two fundamental questions. How is the expected rate of return to existing shareholders related to RAROC? Second, how sensitive is the RAROC methodology to changes in the risk of the underlying business?

Crouhy, Turnbull and Wakeman (1999) showed that, assuming that the risky asset is totally financed by debt and that the market value of debt is less than the market value of the risk asset, results: for a positive NPV project the RAROC of existing shareholders is greater than the RAROC to equity shareholders and this last one is greater than the RAROC of new

²¹ The primary objective of Bank of America system is to assign equity capital to business units (and ultimately to individual credits) so each BU has the same cost of equity capital. This process implies that investments in riskier projects or BUs (measured by the projects contribution to the overall volatility of the market value of the bank) will be required to use less leverage than investments in less risky BUs

²² It is easy to show that the adjusted RAROC is just equal to the expected excess rate of return of the market portfolio.

shareholders; for a zero NPV these three RAROCs are the same²³ and for a negative RAROC the opposite relationship of case 1 holds²⁴. It is also interesting to note that this result is independent of used valuation model and is simply a consequence of the RAROC definition. Crouhy, Turnbull and Wakeman (1999) also founded that: the Adjusted RAROC is insensitive to changes in volatility and correlation and RAROC is sensitive to both²⁵; RAROC is an increasing function of NPV²⁶; RAROC is contingent on both the sign of NPV and the volatility of the underlying risky asset, in particular: a) if the project has positive NPV, the variation of RAROC with volatility is U-shaped; b) if the project has zero or negative NPV, RAROC increases as volatility increases; c) For low volatility, RAROC is relatively sensitive as to whether the project as positive or negative NPV; the level of sensitive decreases as the level of volatility increases²⁷.

A crucial result is that by following the traditional RAROC methodology, because the RAROC depends upon the level of the volatility, it could happen that, for large volatility project, RAROC may become sufficiently large that it would be result greater than the required hurdle rate. Hence a project with a negative NPV may will be accepted. Summarising, the Adjusted RAROC provides the correct investment decision, but this RAPM requires to measure the beta of the equity. Practitioners have developed RAROC mainly because it adjusts risk without requiring to measure the beta of the equity. However, Crouhy, Turnbull and Wakeman (1999) demonstrated that the traditional RAROC does not completely adjust for risk and consequently financial institution are still faced with the problem of estimating the equity beta. On the other hand, this last approach does not take in consideration the distinction between the tradable (hedgeable) and non-tradeable risk factor component introduced by Froot and Stein. This means that, even the Adjusted RAROC is an inefficient RAPM if the implicit tradable component of the generic project i is not diversified away. In the Froot and Stein framework, capital budgeting is driven by an exposure's impact on portfolio risk and return and consequently, the optimal allocation to new investment opportunities should be jointly

²³ When making a new investment project, only for zero net present value project the RAROC of the existing shareholders equals the RAROC of new shareholders and the RAROC of sum of the two classes of shareholders. In appendix 2 a numerical example of this limiting case is shown. The zero NPV relationship provides some credence for comparing RAROC to the firm's cost of equity capital for stand-alone projects (see Zaik and al. 1996). Bank of America's policy is to capitalise each BU in a manner consistent with an AA credit rating, based on the unit's stand alone risk, but also including an adjustment for any internal diversification benefits provided by the unit. Each of these individual capital allocations are then aggregated to arrive at the optimal level of equity capital for the entire bank.

²⁴ Crouhy, Turnbull and Wakeman (1999) in the first appendix of their paper, pag 32, showed this result.

²⁵ If, as described by Zaik and al. (1996) a generic financial institution follows Bank of America's use of a fixed hurdle rate, than it would pick up high volatility and high correlation projects.

²⁶ Suppose the NPV decreases; this implies that the numerator in RAROC decreases and then denominator increases, and hence RAROC decreases.

determined. Stoughton and Zechner (1999) showed that optimal equity capital allocation is based on a business unit's contribution to the institution's total capital requirement and can be decomposed into two parts: the economic capital and the risk adjustment term. The economic capital term is measured by the divisional Incremental Value at Risk (IVaR). The sum of the IVaRs is equal to the institution's overall VaR. In the case of multiple business units, the Economic Capital is equal to a price of risk multiplied by the division's own standard deviation. The risk adjustment term is a constant related to the economic value added (EVA) at the optimal investment level. They also declared that the right definition of RAROC involves in the numerator the business unit's expected return minus an adjustment factor and in the denominator an economic capital amount equal to the unit's IVaR. This adjustment reflects both the economic value added (EVA) at the optimal risk level and risk externalities imposed by one business unit to the IVaR of other business units. According to Stewart (1994) Economic Value Added or EVA, which is an estimate of a business true economic profit, acts as an incentive compensation plan that can be used to motivate everyone in the organization from senior management to the ordinary employees. Stewart argued that EVA can be considered the best metric of wealth creation as it embodies in one measure capital budgeting, financial planning, goal setting, performance measurement and shareholders' wealth and also eliminates the accounting and financial distortions to the extent it is practical to do so. It is the only measure that takes into account the cost of capital and the amount of capital invested in the company and thus the most appropriate management tool of financial management and incentive compensation that creates value for shareholders and management.

3- Economic capital and performance measurement²⁸

This section discusses the meaning of 'economic capital' and contrasts it with other commonly used measures of banking capital. It also compares the different perspectives of bond holders and shareholders, considering how these can both be incorporated within the contingent claims analysis of Merton.

²⁷ Crouhy, Turnbull and Wakeman (1999) showed also that the RAROC of a loan, when assuming the cost of borrowing for the bank is the risk-free of interest, varies as the volatility of the firm's risky assets change, even though the probability of default for the firm and the bank are held constant. It is also an increasing or decreasing function of the volatility of the firm's risky assets depending upon its credit rating; for low credits, RAROC increases as volatility increases.

²⁸ This section is based on the following papers: Wilson, T.C. (2002) Overcoming the Hurdle: Integrating the Shareholder/ Debtholder-Perspectives, Mercer Oliver Wyman, New York; Wilson, T.C. (2002) Cost of capital for financial institutions, Working paper, Mercer Oliver Wyman; Wilson, T.C. (2003) Valuation, earnings volatility and cost of capital, Working paper, Mercer Oliver Wyman

3.1 Alternative definitions of capital

In a recent Survey it was found that there is no one consistent definition of Economic Capital in use in the marketplace²⁹. Main themes of the various practical alternatives currently in use can be summarised in the following three definitions:

1. Economic Capital is defined as sufficient surplus to meet potential negative cash flows and reductions in value of assets or increases in value of liabilities at a given level of risk tolerance, over a specified time horizon.
2. Economic Capital is defined as the excess of the market value of the assets over the fair value of liabilities required to ensure that obligations can be satisfied at a given level of risk tolerance, over a specified time horizon.
3. Economic Capital is defined as sufficient surplus to maintain solvency at a given level of risk tolerance, over a specified time horizon³⁰.

While Definitions 1 and 3 refer to “sufficient surplus”, Definition 2 instead focuses on the characteristics of the assets (market value) and the liabilities (fair value) that define this surplus. Each definition presents a different expression for the adverse outcome that the Economic Capital is intended to protect against. Definition 1 refers to “potential cash flows and reductions in value of assets or increases in value of liabilities.” Definition 2 is concerned only that “obligations can be satisfied.” The goal of Definition 3 is to “maintain solvency³¹.” These broad definitions seem to imply that all risks³² are to be taken into account. Moreover we are aware of

²⁹.Capital Market Risk Advisor, Economic Capital Survey Overview, 2002. Participants include some of the largest global banks and investment banks in the world as well as representative domestic financial institutions from US, UK, Japan, Italy, Canada, Norway, South Africa, Brazil, Mexico and Estonia.

³⁰ All definitions above refer to a “specified time horizon.”

³¹ Several methods are commonly used to set the risk tolerance levels, including: A specified percentile (e.g., 98th percentile), often related to the financial strength rating of the company; Conditional Tail Expectation (CTE), CTE(n) represents the average of the (100-n) worst scenarios (for example, CTE(90) is the average of the worst 10% of scenarios). The first approach is the one commonly used by the rating agencies, while the latter is being used in Canada, and will likely be used for setting regulatory capital requirements for variable products with guarantees (“RBC C-3 Phase II”), starting in 2004, in the insurance industry.

³² The following main risk categories were identified as relevant for calculating Economic Capital: interest rate risk, pricing risk, credit risk, equity market risk, liquidity, operational (business) risk. In determining the appropriate level of EC, the interaction among these risks should be considered. This could be accomplished by the use of either multivariate distributions or correlation factors. Some other financial institutions mentioned other liability risks that would need to be considered, including: separate account risk, i.e., the risk of adverse market performance, which can lead to lower negative profit margins on equity-based products, as well as payouts under the death and living guarantees typically offered with such products; transfer risk, i.e. the risk of policyholders exercising their transfer rights under equity-based products to the detriment of the insurance company; the minority think that all operational risks should be included, and others referred to subsets of this category.

some companies (particularly those which are owned by banks) using earnings³³ oriented approaches for calculating Economic Capital.

Summarising, risk is related to the amount of capital the firms requires to achieve a sufficient level of protection against adverse circumstance and, at the same time, risk is used to adjust the returns from business activities to determine whether activities are value adding or value destroying. The first part should reflect a debtholder's perspective on risk (i.e. is there sufficient capital to cover "worst case" risk?). The second part should reflect a shareholder's perspective on risk (i.e. are we getting a sufficient return for the systematic risk begin taken?). The debtholder and shareholder views of risk differ, but as it will be shown in the following sections, they are related. In classical corporate financial theory capital carries out two particular roles: transference of ownership³⁴ and funding the business³⁵.

There are at least six different definition of capital that a financial institution should consider and each of them depends on the views and the interest of the group that it represents:

1. **Book Capital.** This is an accounting measure of the capital that the bank holds and can be observed on the balance sheet. The book capital measures mainly on-balance-sheet assets and liabilities on an historical cost basis, that may deviate substantially form the actual market values and does not include most off-balance-sheet items. Also, the book capital cannot be withdrawn in times of crisis³⁶.
2. **Market Capital.** The market value of capital is its current value. Although it is supposed to be better measure, there are difficulties in its application. First, most of the banks assets are non-tradable in the open market because of the bank has specific knowledge or informational advantages and many investors are not willing to buy those assets because of the asymmetric information. Consequently, it is very difficult to determine the market

³³ Risk is often modelled as volatility of accounting earnings. Hence, the distribution of asset value is approximated by a distribution of accounting earnings. This can seriously misstate the true economics risk to the firm. For example, the balance sheet interest rate risk many US thrifts during the 1980s, if measured by Δ NII (net interest income), would have looked manageable, but if measured on Δ NPV basis, would have looked catastrophic. The latter view was correct, but was obscured by the one-year accounting earnings-based view of the Δ NII analysis. We believe strongly that capital markets investors, both debtholders and shareholders, rely on an economic view, not an accounting view, of risk-albeit an imperfect one as they only have financial statements on which to form their view. Hence, we believe that risk must be modelled as the value volatility to conform to an investor's view of risk.

³⁴ By shelling shares to third parties a company transfers ownership of assets, profits and risks to the holders of those shares

³⁵ Companies use capital as a source to raise funds for their investments. However, since interest payments to bondholders are tax deductible the companies prefer to increase their leverage (debt/equity), as it is less expensive for them to finance their projects, rather than use equity. But if leverage is too high, there is a risk of bankruptcy and other associated financial distress costs. On the other hand a very low leverage might lead to hostile takeover bids which will bring leverage close to the optimal level.

³⁶ Therefore, according to Cordell and King (1995) accounting-based measures of capital may overstate the actual value of capital that is available to absorb losses

value of those assets. Second, the market capitalization of the bank's traded shares is very volatile as markets are very sensitive to changing conditions

3. **Regulatory Capital.** Regulators require banks to hold capital to ensure the safety and soundness of the banking system and protect the stakeholders from the consequences of a financial distress³⁷. The required capital from a regulatory point of view is divided into Tier-1 and Tier-2. Both are book capital amounts. Tier-1 capital (Basic) is defined as: Equity capital plus Disclosed reserves. Tier-2 capital components are (Supplementary): Undisclosed reserves, Revaluation reserves General provisions/general loan-loss reserves, Hybrid debt capital instruments and Subordinated term debt. Tier-2 must amount to less than Tier-1 capital. In other words Tier-1 must at least amount to 50% of the total capital required. It should be noted that a sufficiently high capital ratio is necessary but it is not sufficient to ensure that a bank operates as a sound financial institution. Since Regulatory capital varies by political jurisdiction, if an institution works across many jurisdictions and have different strategies for each region decides to base its decisions on Regulatory capital requirements, it will not take any advantage on the economies of scale³⁸
4. **Rating Agency Capital.** This type of capital is defined by the rating agencies and it is very similar in spirit to the Regulatory capital. Rating agencies are concerned with the level of financial strength and general creditworthiness of a firm. Capital requirements are assigned by such organizations such as Moody's, Standard & Poor and Fitch
5. **Risk Capital.** This specific feature of capital can be considered as the amount of capital required by the debtholders³⁹ to cover a potential decrease of the assets' value over a given time period and at a given statistical confidence level⁴⁰. There are two dimensions of risk: expected loss and unexpected loss. Expected loss is the average rate of loss expected from a portfolio. Unexpected loss is the volatility of returns or losses around their expected levels and it is the main reason that creates the need for economic capital. The unexpected

³⁷ In the early 1970s after the collapse of the Bretton Woods agreement there was a significant increase in volatility of interest rates and exchange rates, which raised concerns about the financial institutions. In order to avoid unpleasant surprises a committee was set up to see what could be done in an international level. The committee consisted of representatives of the central banks of the G10 countries and is known as Basel Committee on Banking Regulation and Supervisory Practices (BIS). The Committee published its proposals in July 1988 and is known as the "Basel Accord". The Basel Accord succeeded in ensuring a security in the banking system by requiring all financial institutions to hold capital ratios of at least 8% of risk-weighted assets and encouraged the international competition by subjecting all the banks to the same rules. In January 2001 released the Basel II which represents a major change in the approach to capital regulation. Under the new Basel, capital will be assessed according to the risks that a bank undertakes: market risk, credit risk and operational risk. The new Accord consists of three pillars. Pillar I encourages banks to collect and analyze data. Pillar II will be the supervisory review of data and risk-management systems. Finally Pillar III will ensure that the banks that don't comply with the regulations will be penalized by the financial markets.

³⁸ Specialty Guide on Economic Capital, Version 1.4, June 2004.

³⁹ In other words, debtholders want to know if there is sufficient capital to cover "worst case" risks

losses are estimated by calculating the deviation of losses from their mean i.e. the volatility of losses. The unexpected loss contribution determines the level of capital that is assigned to each business unit as it takes into account the correlation of the credit with the overall portfolio⁴¹.

6. Economic Capital. It seems reasonable to define Economic Capital as that capital required by the shareholders to optimize their returns and compensate them for the systematic risk that they bear. Apart from the compensation the shareholders seek dividend payments and share price appreciation. Therefore the only difference between risk and economic capital should be the goodwill: Economic Capital = Risk Capital + Goodwill.⁴² As it will be illustrated in the next section, the shareholders are mainly concerned for the relationship between risk, return and value. Particularly, the shareholders are interested only in the systematic risk and not the total risk, because they can diversify away the unsystematic risk. Additionally, they are interested in the “fat part” of the net asset value distribution because they are exposed to the entity of the distribution. Hence, the best way to measure the risk that they bear is to use the standard deviation of the assets.⁴³

3.2 A conceptual framework for attributing capital

3.2.1 The Merton model of default.

Many financial institutions can be analysed using a framework based on Merton (1974)'s model of default and its extension to the analysis of bank risk taking (Merton (1977)). This model effectively says the following:

- Shareholders own the right to default on debtholders, and will do so if the value of the firm's equity (or “net asset”) drops to zero;
- Debtholders charge shareholders for default risk by demanding a spread over the risk-free rate on the funds they provide;

⁴⁰ The choice for time horizon for measuring risk is arbitrary. One could use volatility measured over long periods, such as five or ten year intervals in order to capture “full cycles” in risk. On the other hand, it is hard to get reliable data for long periods and therefore some prefer to use volatility over much shorter periods of time

⁴¹ For example, an investment can be characterized as very risky on a stand-alone basis but if the predicted losses are not perfectly correlated with the rest of the bank's projects its addition might reduce the overall risk.

⁴² Chris Matten, *Managing Bank Capital* p. 34.

⁴³ A survey that was conducted over a six-week period from July to September 2002 in Boston in which they participated 491 members of the financial, investment and international section in America provides useful results about Economic Capital. Particularly, most responders agreed that Economic Capital should cover various types of risks (market risk, credit risk, operational risk, liquidity risk). Almost half of them have used EC in their work to measure risk and performance and they consider that it will be excessively used in the future. Economic Capital Survey, July-September 2002, Boston.

- The probability of default is a function of the firm's net asset value distribution and its current net asset value.

A theoretically robust estimate for the level of capital required on an economic basis can be computed from the Merton model if the net asset value distribution can be estimated reliably and a solvency standard (probability of default) is selected carefully.

This model can be used to bring the various measures of capital – economic capital, market capital, risk capital, etc – into a common analytical framework. The purpose of this sub-section is to provide an overview of how this is done, using the Merton model to analyse economic capital and value from both the shareholder and bondholder perspectives. Whereas bondholders are concerned only with the downward tail of risk, shareholders, protected by limited liability, may care only about the outcomes where the bank remains solvent. If the bank has a continuing franchise value, then shareholders and bond-holder interests are better aligned; shareholders will then seek to avoid insolvency but they will still not be concerned about the extent of losses in the event that insolvency occurs. This discussion provides a setting for the subsequent theoretical analysis of Section 4 in which a formal model is proposed for conducting this analysis of decision-making within financial institutions, combining the perspectives of both bondholders and shareholders outlined here.

3.2.2 Economic capital and value measurement in financial institutions: the bondholders perspective

In this framework, from the debtholder's perspective, risk can be defined as the possibility that the value of the equity of the firm drops to zero (or less than zero). The debtholder cares about:

- The total firms risk, not just the systematic component.
- The tail of the net asset value distribution. Creating this distribution requires that we define “net asset value”, and define the time horizon over which we will measure its volatility
- Value in the event of financial distress. What really matters for bondholders it is the value of the financial institution that would exist to protect debtholders in such event
- In practice, net asset value for our debtholder risk capital calculations can be the algebraic sum of the first three of the following four elements:
 - Current book equity;
 - Mark-to-market adjustments on tangible asset

- Hard intangibles⁴⁴, i.e. those net assets that have no accounting book value, but which would still have marketable value in the event of distress;
- Soft intangibles (value of a customer franchise, value of a brand) are that net asset whose value would not be realizable in distress scenario.

It is necessary that these judgments be made consistently. The definition used to generate the value distribution needed for economic capital calculations must be identical to that used to define the current value of firm's equity (a quantity we will call "adjusted book equity")⁴⁵. This will enable an accurate comparison between required resources (economic capital) and available resources (adjusted book capital).

The next step is aimed at generating value distribution. In generating value distributions, we group risks in a manner consistent with the identified bank risk measurement and management framework, based on the drivers of risk and nature of the risk distribution. In theory, the risk contribution from each of these risk sources should be measured by using the definition of value outlined above. In practice, measuring the volatility of a key asset value driver is sometimes sufficient to approximate full value volatility.

Having constructed the aggregate value distribution, we must now determine the probability of default (solvency standard) that is acceptable to the institution. Capitalization standards should be expressed as a probability of insolvency over a specified time period, e.g., a 0.1% probability of insolvency over a one-year time period i.e. economic capital corresponds to an aggregate value-at-risk. It is usual to relate this desired solvency standard to an intuitive external benchmark such as, for example, the bond ratings⁴⁶. Firms should select a target solvency standard based on an evaluation of which rating level optimises their firm's value for firm's shareholders⁴⁷. The time horizon (most often one year) for measuring the value distribution for economic capital calculation for banks is approximately the length of their capital planning cycles⁴⁸. Ultimately, the selection of a specific time horizon is a judgment call. However, once a time horizon is selected, the confidence interval to be used in computing

⁴⁴ E.g. a separately-branded mutual fund business, mortgage servicing rights

⁴⁵ There is obviously a grey area between the "hard" and "soft" intangibles on with judgment calls need to be made.

⁴⁶ For example, firm rated AA by S&P has historically defaulted with a 0.03% frequency over a one-year horizon. If a firm has an "AA" target solvency standard, then economic capital can be determined as the level of cushion required to keep the firm solvent over a one-year period with 99.97% confidence .if the solvency standard shifted to BBB, the confidence interval would shift to 99.80%.

⁴⁷ Note that a firm capitalizes to an "AA" solvency standard many not actually be rated "AA" by the agencies, as a rating agency's assessment of credit quality incorporates factors excluded from this analytical framework (eg management quality).

⁴⁸ Banks typically reassess their capital adequacy at least annually, changing dividend policy or undertaking capital markets transactions if necessary (share issuance or buyback). It is also roughly the average maturity of their illiquid credit portfolios, so it fits with the horizon over with a bank could materially shift its risk profile

economic capital should then be explicitly linked to this choice, and the earlier choice of a solvency standard⁴⁹.

3.2.3 Economic capital and value measurement in financial institutions: the shareholders perspective

As described earlier, the debtholder and shareholder views of risk are different, so some adjustments need to be made. As described in the previous paragraph, while the debtholder's primary concern is the link between risk and capital, shareholders are most concerned with the link between risk, return and value. The relationship among these three measures requires both linking required return to risk (i.e. deriving an appropriate discount rate for uncertain returns), and deriving intrinsic (market) value.

A further difference between the perspective of bondholders and shareholders is that while the former are concerned with total firm risk, shareholder value depends primarily upon the systematic component of risk. Relating required return to risk has been one of the most important problems in finance. The most broadly accepted theoretical approaches for addressing this issue such as is the Capital Asset Pricing Model (CAPM), or its more generalized version, the Arbitrage Pricing Theory (APT). The CAPM states that shareholders will require a return in excess of the risk-free rate to compensate them for un-diversifiable risk. CAPM formulates a simple linear relationship between required return and beta, where beta is a function of the volatility of the firm's equity value (σ) and its correlation with the market (ρ). Once the required return (or discount rate) for a given activity is determined based on its risk profile, its intrinsic value can be calculated by discounting expected cash-flows using for example the dividend discount (or discounted cash flow) model. Summarising, risk calculations of this kind are central to value calculations for financial services businesses at all levels of measurement.

To be effective, value measures must be fully integrated with the underlying foundation of risk analytics described in the previous sub-section, i.e. when dealing with the bondholders' view of risk capital. In fact, frequently, value is calculated by requiring a uniform return on economic capital across activities. This implicitly assumes that debtholder risk (measured by economic capital, or EC) and shareholder risk (measurement by CAPM's β) are proportional across all activities. This assumption fails to recognize that the debtholder and shareholder views differ in two fundamental ways:

⁴⁹ It is generally understood that an "AA" / one year horizon leads to the selection of a 99.97% confidence interval for capitalization, while on "AA" /25-year horizon leads to a selection of a 95.5% confidence interval.

1. Systematic versus total risk: shareholders can diversify away specific risk but not systematic risk. So from a shareholder's perspective only systematic risk matters. However, debtholders care about total risk (systematic plus specific) because all risk impacts the probability of default.
2. Standard deviation versus tail risk: shareholders are exposed to the entirety of the value distribution, so they are concerned more with the "fat part" of the distribution than with the tails. Hence standard deviation (σ) is the more useful measure of value volatility, i.e. of risk⁵⁰.

For debtholder and shareholder risk to be proportional across activities, two conditions must hold: systematic risk and total risk must be proportional and σ and EC must be proportional. These considerations suggest that the adoption of uniform hurdle rates should be considered carefully, as there is an inherent accuracy vs simplicity trade-off. As we will show later, for some firms the simplicity gained from a uniform hurdle rate will outweigh the potential increase in accuracy from a more involved analysis; in others, differentiated hurdle rates are clearly necessary to avoid misleading value measures

4- Capital structure and cost of equity with constant default probability

This Section and the following Section 5 and Section 6 provide the first combined analysis of the impact of default probability, capital structure, limited liability, deposit insurance, franchise value and taxes on Risk Adjusted Return on Capital (RAROC)⁵¹ type measures. Previous work of Crouhy and al (1999) has addressed part of this problem but this does not provide a complete analysis. The suggested model also adopts a rigorous equilibrium perspective by enforcing the pricing of banking assets using the Capital Asset Pricing Model (CAPM). This alters significantly some of the results reported by Crouhy and al. (1999). It will be shown that the crucial risk driver is the correlation of the bank asset return with the market and not the volatility of asset returns.

The analysis will consider first the case of fixed default probability. This means that the target credit rating is equal to the actual one. In section six this assumption will be relaxed and the

⁵⁰ Debtholders are exposed to the tail of the distribution – so an extremely high (eg 99.97%) confidence interval is the risk measure used to determine economic capital. Very simply, the tail event is not nearly as "special" for the shareholder as the debtholder

⁵¹ In this framework only the ex-ante perspective will be considered. The cost of economic capital is therefore based on expectations and risk assessment

analysis will be carried out allowing for the target credit rating being different from the actual one. By following this framework it is possible to clearly establish the relationship amongst different measures of capital, namely the market value of the equity, the actual and target risk capital, the actual and target economic capital.

4.1 Modelling assumptions and notations

In this section the main assumptions and model notations are introduced. In the following sections some of these assumptions will be relaxed

Assumption 1 A bank holds a single asset over a one period time

Let consider a bank that holds a single asset $A(t)$ over the period $t=0$ until $t=T$. We will always take the time period for analysis T to be one year, i.e. $T=1$.

Assumption 2 Asset is partly financed by debt

The bank finances this asset by issuing an amount of debt with a market value of $D(t)$.

Assumption 3 Balance Sheet condition evaluated on a mark to market basis

The balance sheet condition, in market value term, can therefore be written as:

$$A(t) = E(t) + D(t), \text{ with } t \in [0,1].^{52} \quad (4.1.1)$$

where

$E(t)$ is the market value of the equity at time t . Therefore, by assumption, the book value of the equity is equal to the market value of the equity⁵³.

Assumption 4 Asset prices are determined in an equilibrium framework in which all returns are normally distributed.

Specifically it will be assumed that end period asset prices are given by:

$A(1) = A(0)[1 + R_A + \sigma_A z]$ where R_A is the expected return, σ_A is the standard deviation of returns, and z is a stochastic variable with a standard normal distribution. For pricing purposes, we will also utilise the correlation of z and hence $A(1)$ with the market portfolio ρ_A .

One key implication of this assumption is that the Capital Asset Pricing Model (CAPM) provides a framework for determining the net present value of a business. Expected future cash flows are discounted using a risk-adjusted expected rate of return. In this set up the risk of a business can be measured through the covariance of changes of the return of the business with

⁵² Without losing generality and for the sake of simplicity we will consider through the analysis that at time T book and market value are equal.

⁵³ This assumption is increasingly becoming a mild one because of the new IFRS (International Financial Reporting Standard) that will be enforced at the beginning of 2005 for all the holdings quoted in the Stock Exchange. This new set of rules, especially the one related with the IAS 39, is introducing the fair value evaluation for all the book items. In 2006, when this set of rules will become in due force also for the other companies (the non-holdings one) this assumption will reflect best practice and therefore, empirically it will be the most correct way to model the bank balance sheet.

changes in the values of market portfolio returns. The following general equilibrium relationship applies

$$R_A(t) - r_f(t) = \beta_{R_A, M}(t)(R_M(t) - r_f(t)) \quad (4.1.2)$$

where, accordingly:

$R_A(t)$ is the return on the bank asset at time t

$r_f(t)$ is the (spot) risk-free rate at time t ⁵⁴

$\beta_{R_A, M}(t)$ is the beta of the return on asset A with the market M

$R_M(t)$ is the market return at time t .

Simplifying the notation

$$r_f(t) = r_f; R_A(t) = R_A; \beta_{R_A}(t) = \beta_{R_A}; R_M(t) = R_M; E(t) = E; A(t) = A; D(t) = D; T - t = 1$$

The equilibrium model can now be rewritten as

$$R_A - r_f = \beta_{R_A, M}(R_M - r_f) \quad (4.1.3)$$

4.1.1 Two key measures of risk: Sharpe Ratio and the option value of default

In the subsequent analysis we will use two key measures of risk, the Sharpe ratio and the option value of default. We now derive both these measures, in our simple model.

For a portfolio of assets one measure of the relative attractiveness of different portfolios is the Sharpe Ratio. It maps the expected return and the risk into a single measure. It is argued that the larger the Sharpe Ratio, the better the portfolio, and hence the objective of the portfolio manager's is to maximise the Sharpe Ratio. We can define the Sharpe Ratio Φ_A on the return on bank asset A as

$$\Phi_A = \frac{R_A - r_f}{\sigma_A} \quad (4.1.1.1)$$

where

σ_A is the volatility on the asset return R_A

Moreover in equilibrium results

$$\Phi_A = \frac{R_A - r_f}{\sigma_A} = \rho_{A, M} \frac{R_M - r_f}{\sigma_M} = \rho_{A, M} \Phi_M \quad (4.1.1.2)$$

where

Φ_M is the Sharpe Ratio defined over the market return R_M

$\rho_{A, M}$ is the correlation between R_M and R_A

σ_M is the volatility of market return R_M

⁵⁴ The risk free rate throughout the analysis is considered deterministic.

We now turn to the option value of default. By using the previous simplifying notations, the balance sheet condition at time 0 can now be rewritten as:

$$E = A - D \quad (4.1.1.3)$$

Within the Merton framework, the market value of the equity can be considered a call option of the bondholders on the asset of the bank (or equivalently the equityholders put on the debt of the bank)⁵⁵:

$$E = AN(d_{1,rm}) - DN(d_{2,rm})^{56} \quad (4.1.1.4)$$

Where

$N(\cdot)$ is the cdf for a variable that is normally distributed with a mean of 0 and a standard deviation of 1.

$$d_{1,rm} = \frac{\left[\left(\ln \frac{A}{De^{\mu_f}} \right) + (\mu_f + 0,5\sigma_A^2) \right]}{\sigma_A} \quad (4.1.1.5)$$

$$d_{2,rm} = d_{1,rm} - \sigma_A \quad (4.1.1.6)$$

μ_f is the instantaneous risk free rate.

It should be acknowledged that in applying these equations we are assuming a correspondence between the instantaneous risk free rate and standard deviations, and their one-period equivalents⁵⁷. The equivalence of standard deviations is not exact because with geometric Brownian motion the period T returns are log-normally distributed (the assumption required for applying the Black-Scholes equation) rather than normally distributed (the assumption of the CAPM). However assuming that the two standard deviations are equivalent introduces only a minor approximation error.

In equilibrium, the following relationship, which represents how the bank cost of (market) equity⁵⁸ is expected to move with respect of the market return, also holds

$$R_E = \frac{\sigma_E}{\sigma_M} \rho_{E,M} (R_M - r_f) + r_f \quad (4.1.1.7)$$

⁵⁵ The strike of the call option is equal to the book value of the firm's liabilities at time 0 $e^{-r_f T} D(T) = D$

⁵⁶ Moreover, if A becomes very large the equity then becomes very similar to a forward contract with delivery price D(T). Consequently d_1 and d_2 becomes very large and $N(d_1)$ and $N(d_2)$ are both close to 1 and, ultimately, results $E=A-D$

⁵⁷ r_f is the annualised discretely compounded rate of return on the risk free rate and is related to the continuously compounded rate of return via the relationship $\mu_f = \ln(1+r_f)$. Moreover, in the remainder we will make the simplifying assumption that the instantaneous volatility of the return on asset A is approximately equal at the annual volatility.

⁵⁸ In the following sections, when adding franchise value at the equilibrium relationship, the cost of equity will then become the cost of economic capital or RAROC

The no arbitrage balance sheet condition at time $T=1$ (4.1), by using discrete compounding, becomes⁵⁹

$$E(T) = A(T) - D(T) = A(1 + R_A) - D(1 + r_f) \quad (4.1.1.8)$$

Note that the actual default probability $N(-d_2)$ can be different from both the target default probability $N(-d_2^*)$ and will always be lower than the risk neutral default probability $N(-d_{2,rm})$ ⁶⁰

To avoid default and to maintain the actual credit rating⁶¹ the bank needs to generate a return $R_A > 1 - d_2 \sigma_A$. If the target credit rating $N(-d_2^*)$ is different from the natural (actual) one $N(-d_2)$, the no default constraint in this case become $R_A > 1 - d_2^* \sigma_A$. This case will be analysed in the last section when the assumption of fixed default probability will be relaxed.

4.2 The determination of promised debt repayments

Suppose the Bank is financed by an amount E and the rest by Debt. Assume now that the target capital structure at time 0 is chosen to keep the target probability of default at time $t=T=1$ to be equal to $p^* = N(-d_2^*)$. In general, because $N(-d_{2,rm}) \neq N(-d_2^*)$ the bank has to identify the correspondent level of debt which allows the bank to safely maintain the desired level of risk, represented here by the target $N(-d_2^*)$ probability of default. This implies that the promised debt repayments at time T are at the appropriate point on the tail of the distribution of asset returns. In the standard economic capital modelling the bank risk appetite resembles the target bank credit rating. The lower the p^* the higher the bank competitive advantage in the interbank market due to credit comparative advantage. In this section, because of the fixed default probability assumption, by construction, the target default probability will be equal to the actual one, i.e. $p = N(-d_2) = p^* = N(-d_2^*)$

Proposition 1.

The promised debt repayments are given by

$$D = \frac{A(1 + R_A - d_2 \sigma_A)}{(1 + r_f)} \quad (4.2.1)$$

Proof

Knowing that

$$A(T) = A(1 + R_A) \quad (4.2.2)$$

⁵⁹ When the risk neutral default probability is equal to the target one

⁶⁰ If d_2^* increases then $N(-d_2^*)$ decreases

⁶¹ Which corresponds to the bank risk neutral default probability observed at time 0

A bank is in default, if at time T $A(T) < D(T)$ or, equivalently, $E < 0$. To remain solvent, at time T, the bank must comply with the following no default constraint:

$$A(T) - Ad_2\sigma_A \geq D(T) \quad (4.2.3)$$

Remarks

1. $Ad_2\sigma_A$ resembles the definition of the Capital at Risk, as introduced in section 2. Therefore the book value and market value of the equity are equal by assumption. Moreover through inspection of (4.2.3), they are also equal, by construction, to the risk capital. In this framework book value, market value and risk capital, from now on labelled E, are equal.
2. Because, by construction, the bank wish to maintain unchanged the actual probability of default we observe that:

- knowing that the return on asset A can be regarded as the sum of two components, namely the risk free rate and the spread S_A associated at the bank rating, the following relationship applies

$$R_A = r_f + S_A \quad (4.2.4)$$

Therefore it is possible to restate the no default constraint also in spread terms

$$\frac{D}{A} \leq \frac{(1 + R_A - d_2\sigma_A)}{(1 + r_f)} \Leftrightarrow \frac{D}{A} \leq \frac{(1 + r_f + S_A - d_2\sigma_A)}{(1 + r_f)} \Leftrightarrow \frac{D}{A} \leq 1 - \frac{(d_2\sigma_A - S_A)}{(1 + r_f)}$$

3. The Debt to Asset ratio can take values between 0 and 1. The closest to 0 the safer the bank. If both actual and credit rating are the same, this polar position is reached when

$$d_2\sigma_A - S_A = 1 + r_f \quad (4.2.5)$$

This means that if there is no debt outstanding, $E=A$.

The closest to 1 the riskier the bank. This polar position is reached when

$$d_2\sigma_A = S_A \quad (4.2.6)$$

Therefore, if there is no equity outstanding, $D=A$.

4. The difference between the market value of the asset and debt at time T changes with the volatility of the asset returns. Therefore the ratio $\frac{E}{A}$ will increase if σ_A increases and vice versa. This is because the bank wants to maintain the risk appetite fixed over time and consequently the riskier the business the higher the amount of equity the bank needs to have to compete in the market with both the same comparative advantage and level of protection.

5. Because in our framework the franchise value will be taken explicitly in consideration, the definition of economic capital will be Capital at Risk plus Franchise Value⁶²
6. The book value and market value of the equity are equal to the risk capital, from now on labelled E (which is associated with the risk neutral default probability), but are different from the economic capital E_{ec} .
7. Hence we have not yet determined the Risk Adjusted Return on Economic Capital (RAROC) but proposition 1 will play a key role in this calculation.

Corollary 1

The cost of equity R_E is equal to both the cost of actual and target risk capital, namely R_E^* but it is different from the cost of economic capital R_{ec} ⁶³.

$$R_E^* = R_E \neq R_{ec}. \quad (4.2.7)$$

The main difference being that the franchise value is not included in the cost of capital.

4.3 Capital structure and cost of equity under unlimited liability

In this section we will show that, when unlimited shareholder liability is assumed, the capital structure is influenced by both the volatility on bank return on asset A and the correlation of the asset with market return while the cost of equity is independent from the volatility of the asset σ_A but does depend on the correlation between asset and market return. This result is in sharp contrast with what Crohuy and al claim (1999). In fact in their study they found that is the volatility the driving factor when evaluating the business unit/project performance. The reason being the fact that they didn't include in their work the impact of no default constraint in the RAROC formulation as it will be shown and proved in the next section.

Proposition 2

If shareholder unlimited liability and fixed default probability is assumed the bank capital structure depends on σ_A and $\rho_{A,M}$

Proof

The proof is straightforward. Due to the balance sheet condition at time T will result $E(T) = A(T) - D(T)$. If we assume constant $\rho_{A,M}$, for a given level of σ_A , a unique $\frac{D}{E}$ is found. For a given level of volatility σ_A , market correlation $\rho_{A,M}$ and level of default probability (or pre-specified target rating bank) a unique capital structure is defined. Moreover

⁶² In practice, the term "economic capital" it is very often not clearly identified and, consequently, misunderstandings and misconceptions can induce top management towards wrong or biased decisions

⁶³ The cost of actual risk capital, for notation easy, is labelled also R_E . Moreover, in the following, the term cost of economic capital and RAROC will be used as synonyms.

it can be shown also that to avoid default and to comply with the pre-specified target rating the following relationship

$$\frac{D}{A} = \frac{(1 + R_A - d_2 \sigma_A)}{(1 + r_f)} \quad (4.3.1)$$

should be always true over time.

Given that by assumption R_M , σ_M , r_f are constant, it will result

$$R_A = f(\sigma_A \rho_{A,M}) \quad (4.3.2)$$

Consequently, because d_2 is chosen by the top management and the amount of asset A is fixed by construction, will result

$$\frac{D}{A} = f(\sigma_A, \rho_{A,M}). \quad (4.3.3)$$

Alternatively we can also write, due to assumption 3 and proposition 1 that the debt to equity ratio $\frac{D}{A}$ is also determined by σ_A and $\rho_{A,M}$. In particular:

$$\frac{D}{A} = \frac{(1 + R_A - d_2 \sigma_A)}{(1 + r_f)} = \frac{1 + \rho_{A,M} \frac{\sigma_A}{\sigma_M} (R_M - r_f) + r_f - d_2 \sigma_A}{(1 + r_f)} = 1 + \frac{\sigma_A (\rho_{A,M} \Phi_M - d_2)}{(1 + r_f)} \quad (4.3.4)$$

Corollary 2

$\frac{D}{A}$ is a decreasing function of the volatility of the Asset A

Proof

Let consider the first derivative of $\frac{D}{A}$ with respect of σ_A

$$\frac{\partial \frac{D}{A}}{\partial \sigma_A} = \frac{\rho_{A,M} \frac{1}{\sigma_M} (R_M - r_f) - d_2}{(1 + r_f)} = \frac{S'_A - d_2}{(1 + r_f)} \leq 0 \quad \forall -1 \leq \rho_{A,M} \leq 1, \quad (4.3.5)$$

where S'_A is the value over the risk free rate with respect of asset A when $\sigma_A = 1$. In equilibrium $S'_A > 0$.

Moreover due to no default constraint we can rewrite equation

$$\frac{D}{A} = \frac{(1 + R_A - d_2 \sigma_A)}{(1 + r_f)} = \frac{(1 + r_f + S_A - d_2 \sigma_A)}{(1 + r_f)} = 1 + \frac{(S_A - d_2 \sigma_A)}{(1 + r_f)} \quad (4.3.6)$$

therefore the bank will not default iff $D \leq A$ and, consequently, if $S'_A - d_2 \leq 0$. In fact, if σ_A increases and $\rho_{A,M}$ and d_2 are held constant, also E increases but D decreases. The $\frac{D}{A}$ ratio decreases because the positive impact of σ_A on E is lower than the negative impact σ_A on D due to the “No Default” constraint.

Corollary 3

$\frac{D}{A}$ is an increasing function of $\rho_{A,M}$

Proof

Let us take the first derivative of $\frac{D}{A}$ with respect of $\rho_{A,M}$

$$\frac{\partial \frac{D}{A}}{\partial \rho_{A,M}} = \frac{\sigma_A}{\sigma_M} (R_M - r_f) > 0 \quad \forall \sigma_A \tag{4.3.7}$$

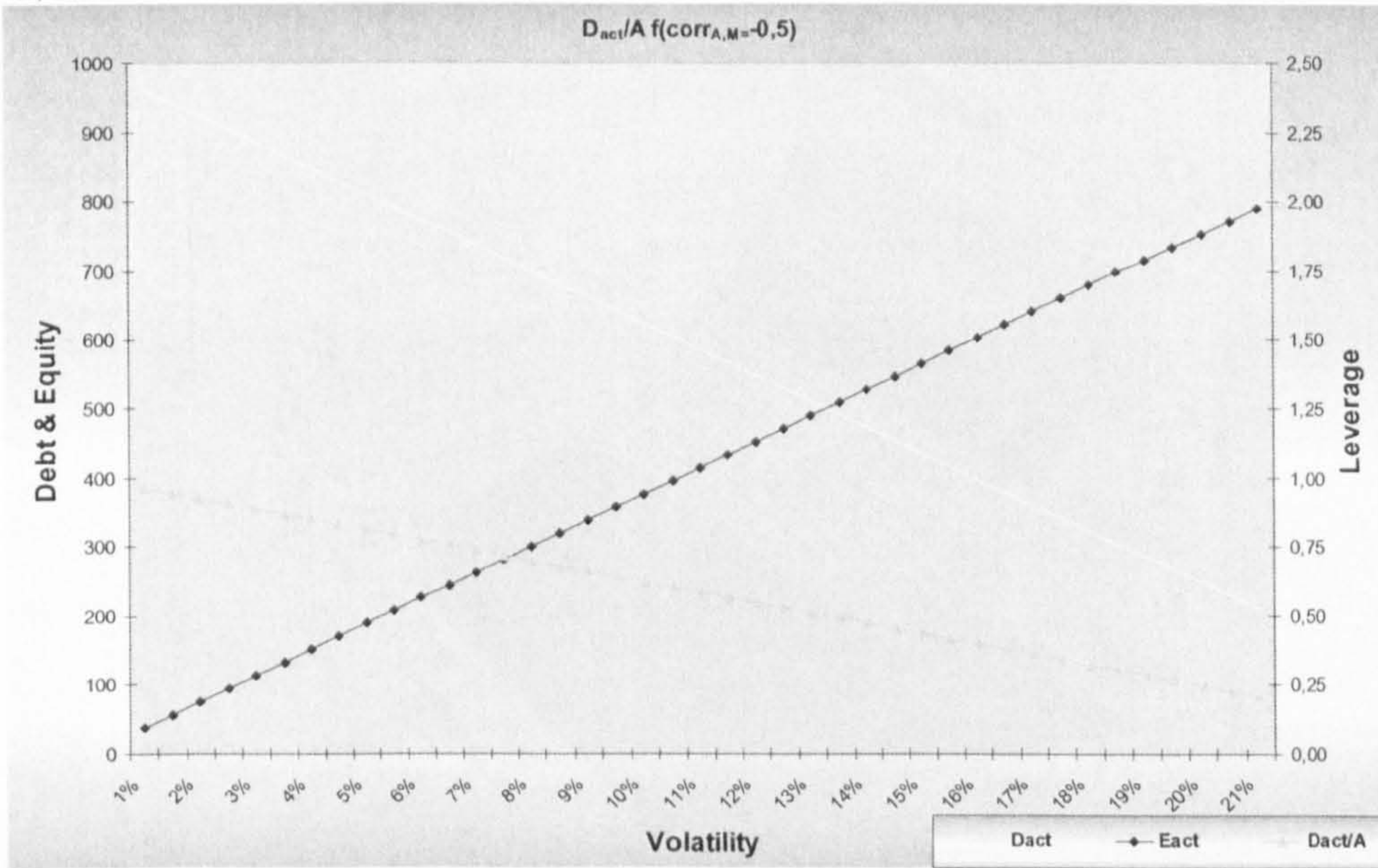


Figure 1 Debt, equity and leverage behaviour when changing volatility and fixed negative correlation (-0.5) is assumed

If $\rho_{A,M}$ increases and σ_A and d_2 are held constant, $\frac{D}{A}$ increases because the risk premium and the volatilities are always greater than zero

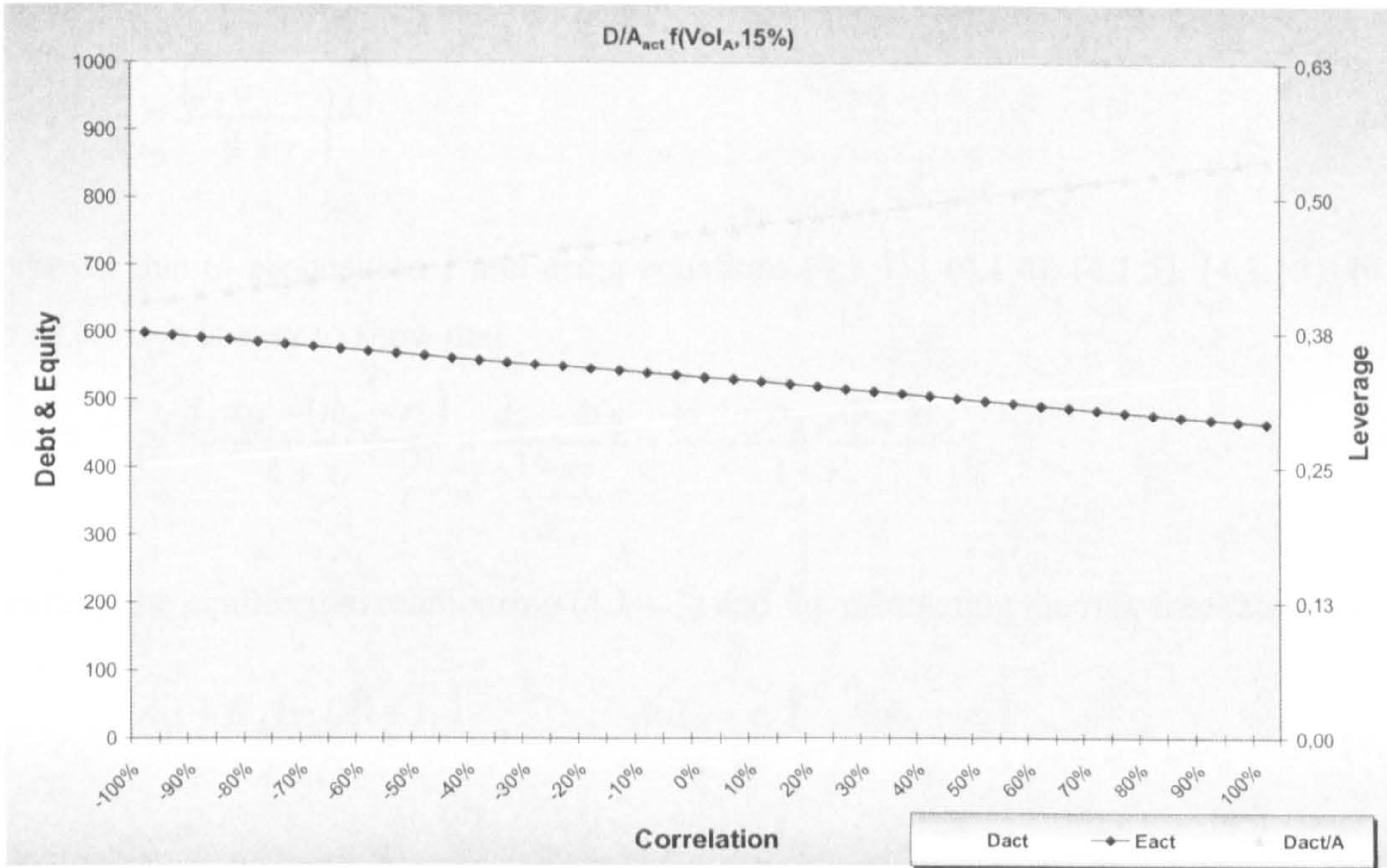


Figure 2 Debt, equity and leverage behaviour when changing correlation and fixed volatility (0.15) is assumed

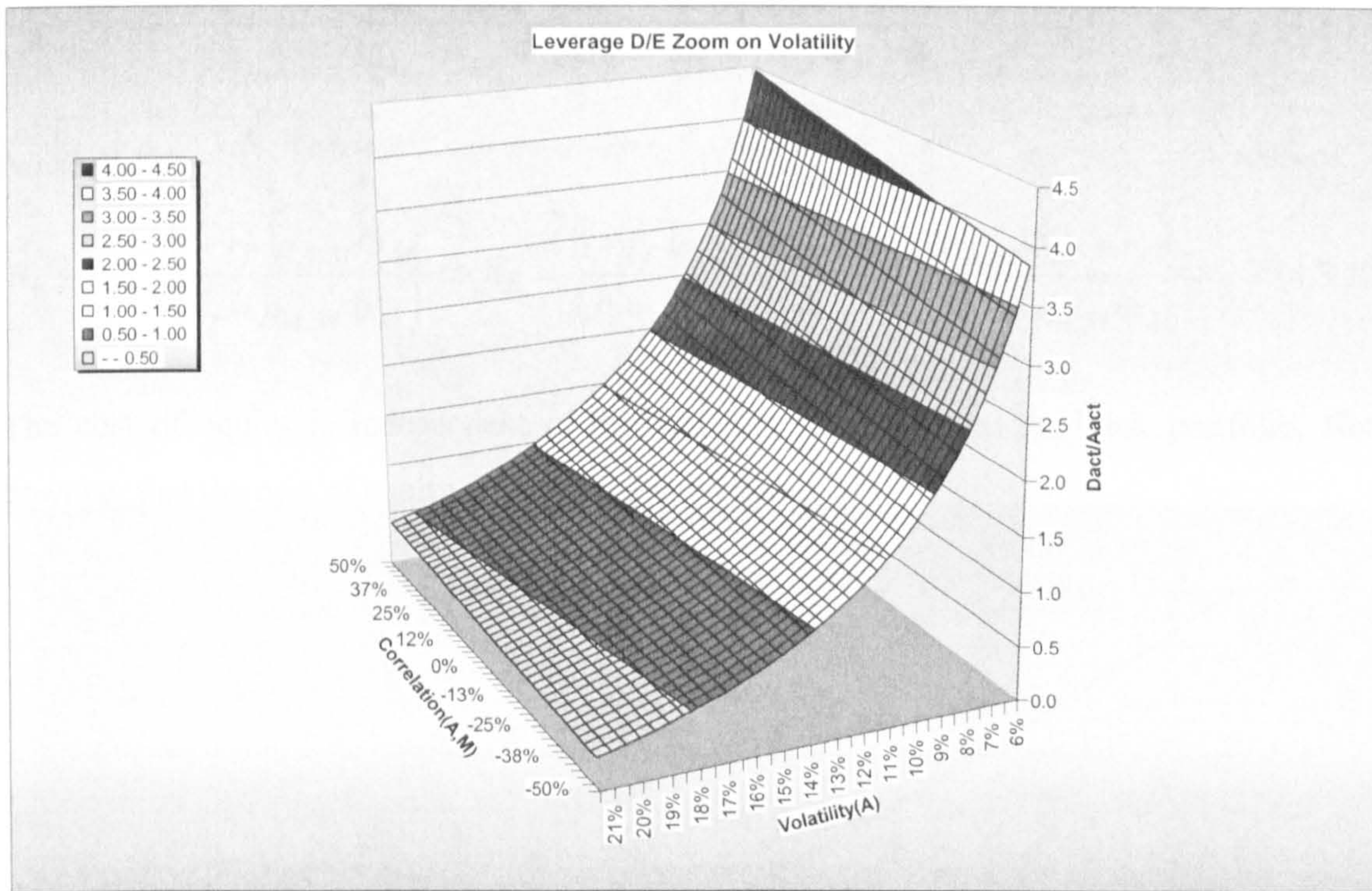


Figure 3 Debt, equity and leverage behaviour when changing both volatility and correlation

Proposition 3

If shareholder unlimited liability and fixed default probability is assumed, the cost of equity risk capital R_E depends only on $\rho_{A,M}$, i.e $R_E = g(\rho_{A,M})$

Proof

In equilibrium, relationship (4.1.14) holds. Consider the balance sheet condition at time t (4.1.6) and T (4.1.16), rearranging

$$\frac{E}{A} = 1 - \frac{D}{A} = \frac{(d_2 \sigma_A - S_A)}{(1 + r_f)} \quad (4.3.8)$$

Moreover due to proposition 1 and using equations (4.1.3) , (4.1.4), (4.1.5), (4.1.13), (4.1.16) and (4.3.6) it is easy to show that

$$\frac{E}{A} = 1 - \frac{D}{A} = \frac{d_2 \sigma_A - (R_A - r_f)}{1 + r_f} = \frac{d_2 - \Phi_A}{\frac{1 + r_f}{\sigma_A}} = \frac{(d_2 - \rho_{A,M} \Phi_M) \sigma_A}{1 + r_f} \quad (4.3.9)$$

Recalling the equilibrium relationship (4.1.1.3) and by subtracting the risk-free rate we get:

$$R_E - r_f = \frac{A(1 + R_A) - D(1 + r_f)}{A - D} - 1 - r_f = \frac{A(R_A - r_f)}{A - D} = \frac{A(R_A - r_f)}{E} = \rho_{A,M} \Phi_M \sigma_A \quad (4.3.10)$$

This increase in the expected rate of return on equity (risk capital), relative to that on the bank assets, reflects the magnifying impact of leverage on the standard deviation of equity returns.

Substituting in eq.(4.3.10) eq. (4.3.9), we obtain:

$$R_E - r_f = \frac{A(R_A - r_f)}{E} = \frac{(1 + r_f)(R_A - r_f)}{(d_2 - \rho_{A,M} \Phi_M) \sigma_A} = \frac{(1 + r_f) \rho_{A,M} \Phi_M \sigma_A}{(d_2 - \rho_{A,M} \Phi_M) \sigma_A} \quad (4.3.11)$$

consequently

$$R_E - r_f = \frac{(1 + r_f) \rho_{A,M} \Phi_M}{(d_2 - \rho_{A,M} \Phi_M)} \Leftrightarrow R_E = \frac{(1 + r_f) \rho_{A,M} \Phi_M}{(d_2 - \rho_{A,M} \Phi_M)} + r_f = \frac{\rho_{A,M} \Phi_M + r_f d_2}{(d_2 - \rho_{A,M} \Phi_M)} \quad (4.3.12)$$

The cost of equity is independent of the volatility of returns on the bank portfolio. Note however that the cost of equity does depend upon the $\rho_{A,M}$

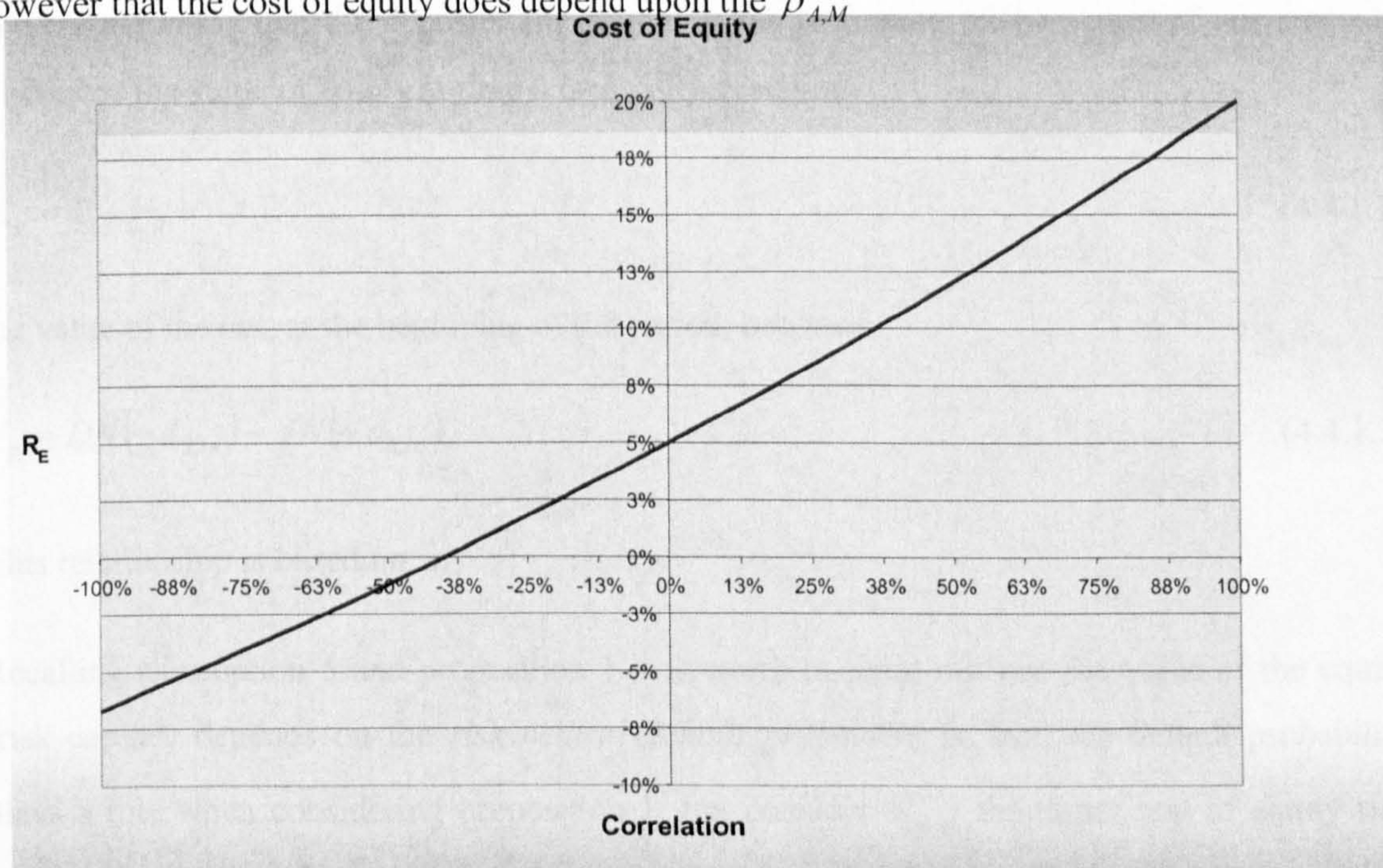


Figure 4 Cost of equity (risk capital) and correlation relationship

4.4 Capital structure and cost of equity (risk capital) under limited liability

Let consider the previous framework but when no loss of value in the event of default is possible because deposits are insured and, as such, shareholders have limited liability. In this case the equityholders are protected from downside risk because of the value of the deposit insurance put. The required rate of return is shown to be higher than the case illustrated in the previous paragraph.

4.4.1 Limited shareholders liability and deposits are insured

Let define:

E_p the value of the equity (risk capital) when the deposits are insured through the value of the put

D the value of the deposits when they are insured through the value of the put

V_p the value of the put option on the deposits of the bank

Proposition 4

If shareholder unlimited liability, fixed default probability and insured deposits are assumed, the correspondent cost of equity risk capital $R_{E,p}$ depends on $\rho_{A,M}$ and σ_A

The following balance sheet relationship, in a risk neutral world, must hold

$$E_p = A - D + V_p \quad (4.4.1.1)$$

We are assuming that the deposits are insured at the beginning of the period of our analysis. Therefore the value of equity at time 0 is E_p . Rearranging

$$E_p = E + V_p \quad (4.4.1.2)$$

the value of the put, at the beginning of the period, becomes

$$V_p = DN(-d_{2,rn}) - AN(-d_{1,rn}) \quad (4.4.1.3)$$

This relationship is based on σ_A

Recalling assumption 5 and proposition 1 it is worth to point out that the value of the equity (risk capital) depends on the *risk neutral default probability*. In fact, the default probability plays a role when considering proposition 1. Let consider $R_{E,p}$ the target cost of equity risk

capital when deposits are insured through the put. By proposition 1 and equation (4.1.16) we have

$$E_p(1 + R_{E,p}) = A(1 + R_A) - D(1 + r_f) \quad (4.4.1.4)$$

Summarising

E_p depends on σ_A and $R_{E,p}$ depends on σ_A

Subtracting the risk free rate and rearranging we get

$$R_{E,p} - r_f = \frac{A(1 + R_A) - D(1 + r_f)}{A - D + V_p} - 1 - r_f \quad (4.4.1.5)$$

$$R_{E,p} = \frac{\rho_{A,M} \Phi_M \sigma_A - \frac{V_p}{A} (1 + r_f)}{\frac{E}{A} - \frac{V_p}{A}} + r_f \quad (4.4.1.6)$$

From the previous equations it is possible to observe that σ_A has a “direct” positive impact on the $R_{E,p}$ and σ_A has a “direct” positive impact on V_p . But because V_p has both a positive impact (numerator) and a negative impact (denominator) on $R_{E,p}$ the total effect is not trivial. In particular it is possible to observe that this trade-off is not linear and it is also dependent from the probability of default. In the next two graphs, two scenarios are considered: the first one when $PD=0,01\%$ and the second one when $PD=1\%$. From a geometric inspection it is possible to observe that when the PD is low the volatility does not play a significant role. Moreover, the value of the put is almost linearly related with the level of PD . In fact by increasing the level of PD it is possible to observe the magnifying impact of volatility on the value of the put, the

value of the leverage ratio and the cost of equity (risk capital)⁶⁴.

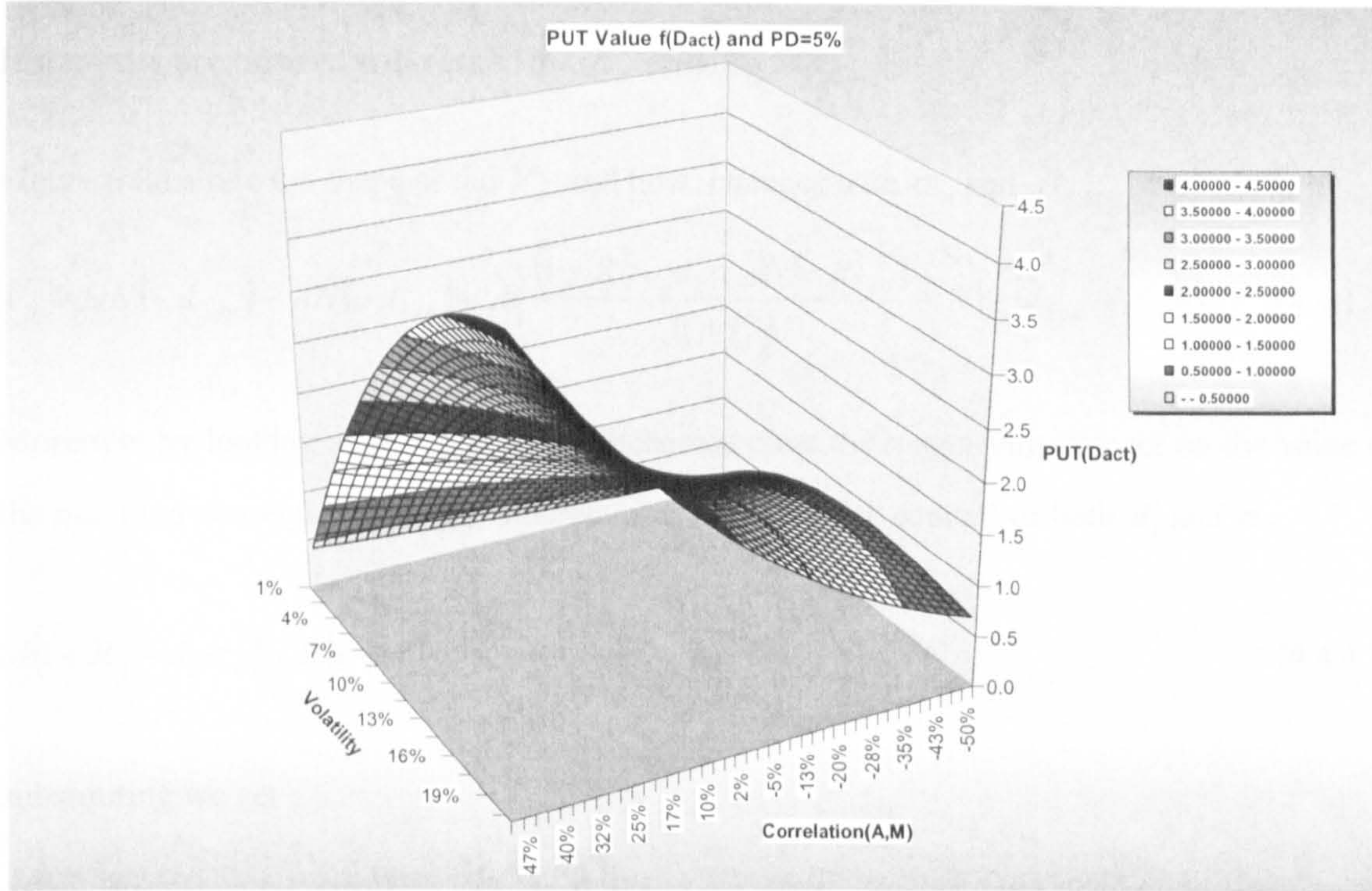


Figure 5 The put value as a function of both volatility and correlation with fixed high PD (0.5%)

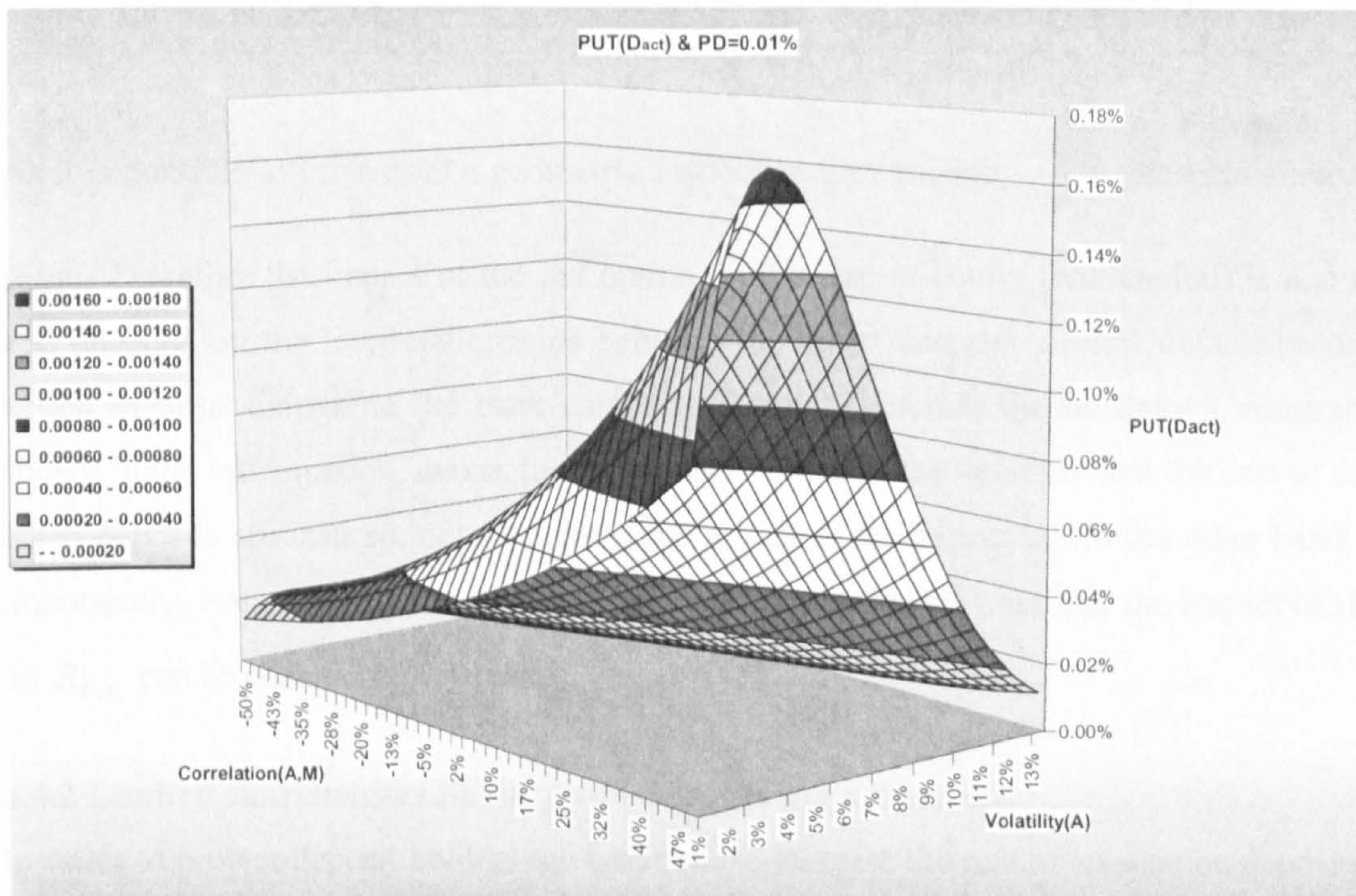


Figure 6 The put value as a function of both volatility and correlation with fixed low PD (0.01%)

⁶⁴ The magnifying impact of leverage on $R_{E,p}$ is partially offset by the impact of the put option.

Corollary 4

If deposits are insured will result $E < E_p$ and $R_E > R_{E,p}$

Figure 6 illustrate the shape of the V_p and how changes with σ_A and $\rho_{A,M}$

$$V_p = DN(-d_{2, rn}) - AN(-d_{1, rn}) = A \left\{ \frac{(1 + R_A - d_2 \sigma_A) N(-d_{2, rn})}{(1 + r_f)} - N(-(d_{2, rn} + \sigma_A)) \right\} \quad (4.4.1.7)$$

Moreover by looking at (4.4.1.7) it now becomes clear the magnifying impact on the value of the put (and consequently on the inherent cost of equity risk capital) of both d_2 and σ_A

$$A(1 + R_A - d_2 \sigma_A) = D(1 + r_f) \Leftrightarrow \frac{D}{A} = \frac{(1 + R_A - d_2 \sigma_A)}{(1 + r_f)} \quad (4.4.1.8)$$

substituting we get

$$V_p = A \left\{ \frac{(1 + R_A - d_2 \sigma_A) N(-d_{2, rn}) - [N(-(d_{2, rn} + \sigma_A))](1 + r_f)}{(1 + r_f)} \right\} \quad (4.4.1.9)$$

$$R_{E,p} = \frac{\rho_{A,M} \Phi_M \sigma_A - \left\{ (1 + R_A - d_2 \sigma_A) N(-d_{2, rn}) - [N(-(d_{2, rn} + \sigma_A))](1 + r_f) \right\}}{(d_2 - \rho_{A,M} \Phi_M) \sigma_A - \left\{ \frac{(1 + R_A - d_2 \sigma_A) N(-d_{2, rn}) - [N(-(d_{2, rn} + \sigma_A))](1 + r_f)}{(1 + r_f)} \right\}} + r_f \quad (4.4.1.10)$$

As it is possible to infer from a geometric inspection the function $\frac{\partial R_{E,p}}{\sigma_A}$ presents a maximum value. Therefore the impact of the put option on the cost of equity (risk capital) is non trivial and depends on the interrelationships between the target and risk neutral default probability which in turns determine the bank capital structure. Therefore the no default constraint, as shown in the last equation, makes the relationship between the volatility and the cost of capital, when deposits are insured, not a positive one as one might expected. On the other hand, more importantly, because the bank wants to keep the PD the lowest possible, the impact of the put on $R_{E,p}$ can be considered very poor

4.4.2 Limited shareholders liability and deposits are uninsured

In order to protect deposit holders the bank has to increase the rate of interest on deposits. The new rate of interest will be the risk free rate plus a spread s . The bank in this case is less leveraged for any given level of probability of default. The level of the deposit is given by

$$D^s + V_p = D \quad (4.4.2.1)$$

where

D^s represents the level of deposit without the insurance premium given by the put. Remember that:

$V_p = f(p)$ where p is associated to the risk neutral probability

moreover, in our one period model, results

$$D = D(T)(1 + r_f)^{-1}$$

and

$$\frac{D(T)}{D^s} = (1 + r_f + s) \quad (4.4.2.1)$$

Proposition 5

The value of the equity (risk capital) E_s and the cost of equity (risk capital) $R_E(D^s)$ if deposit are uninsured are equal, respectively, at the value of the equity (risk capital) E_p and at the cost of equity (risk capital) $R_{E,p}$ when deposits are insured

Proof

Recalling that

$$E_s = A - D + V_p = E_p \quad (4.4.2.2)$$

because

$$E_s = A - D_s \quad (4.4.2.3)$$

and

$$D^s = D - V_p \quad (4.4.2.4)$$

we have

$$E_p = E_s > E \quad (4.4.2.5)$$

Moreover, we want show that

$$R_E(D^S) = R_{E,p} \quad (4.4.2.6)$$

By applying the balance sheet condition at time T, we have

$$E_S(1 + R_E(D^S)) = A(1 + R_A) - D^S(1 + r_f + s) \quad (4.4.2.7)$$

The cost of equity risk capital if deposits are uninsured is

$$R_E(D^S) = \frac{A(1 + R_A) - D^S(1 + r_f + s)}{A - D + V_p} - 1 \quad (4.4.2.8)$$

recalling that

$$R_{E,p} = \frac{A(1 + R_A) - D(1 + r_f)}{A - D + V_p} - 1 \quad (4.4.2.9)$$

$$R_E(D^S) = R_{E,p} \quad (4.4.2.10)$$

if

$$\frac{A(1 + R_A) - D^S(1 + r_f + s)}{A - D + V_p} = \frac{A(1 + R_A) - D(1 + r_f)}{A - D + V_p} \quad (4.4.2.11)$$

rearranging, we get (4.4.2.5) because

$$D^S(1 + r_f + s) = D(1 + r_f) = D(T) \quad (4.4.2.12)$$

Recalling (4.4.2.1) we have also that

$$R_E > R_E(D^S) = R_{E,p} \quad (4.4.2.13)$$

4.5 A comparative analysis of leverage and cost of risk capital with fixed default probability, limited and unlimited liability.

By comparing the leverage effect in these previously described three cases we have:

$$\frac{D}{E} > \frac{D}{E_p} > \frac{D^S}{E_s} \quad (4.5.1)$$

because

$$D > D^S \text{ and } E_p = E_s > E \quad (4.5.2)$$

The impact of σ_A and $\rho_{A,M}$ on $R_E(D^s)$ is the same when deposits are insured due to effect of increment of the risk free rate measured by the spread.

Through the inspection of (4.4.2.13) it is possible to conclude that the impact of the limited and unlimited liability on the cost of risk capital is the same and it is negative (it lowers the value because $R_E > R_E(D^s) = R_{E,p}$).

5- From risk capital to economic capital: introducing franchise value and taxes

If a bank will default the classical banking theory tell us that what is recovered is the franchise value of the bank. This case can be described as a variation of the previous three cases through the introduction of a digital call option. Suppose there is a loss to shareholders of γ in legal and other fees, in the event of bankruptcy. This event can be measured through a digital call option which adds value to the equityholders if at time T no default will occur.

5.1 Cost of Risk Capital with Franchise Value

As discussed in section 4 when the franchise value is added to the equity risk capital the economic capital is obtained. This value of economic capital is net of the tax effect that will be introduced in the next section.

Proposition 6

If shareholder unlimited liability and fixed default probability is assumed the bank franchise value will add value to the equityholders if no bankruptcy occurs at time T

Let now define with E'_{ec} the value of the market value of the equity plus the franchise value.

Let us also define with V_{DC} the value of the inherent digital call⁶⁵.

$$E'_{ec} = A - D + V_{DC} \quad (5.1.1)$$

⁶⁵ E'_{ec} is the value of economic capital without the premium for insuring deposit and tax effect

$$V_{DC} = \begin{cases} \alpha A & A(T) \geq D(T) \\ 0 & A(T) < D(T) \end{cases} \quad (5.1.2)$$

which analytically is equal to

$$V_{DC} = \alpha A e^{-r_f} N(d_{2,m}) \quad (5.1.3)$$

V_{DC} represents the franchise value of the bank. This value will increase the value of the equity (risk capital) if no default will occur at T and α is the percentage of the assets that the bank will not lose if no default occur at time T. Clearly

$$\gamma = \alpha A \quad (5.1.4)$$

Remark

In this framework the risk capital does not reflect also the franchise value and consequently

$$E'_{ec} > E. \quad (5.1.5)$$

Corollary 5

If deposits are insured the value of the equity, $E'_{ec,p}$, increases due to the positive impact of the put and the inherent cost of economic capital (net of tax effect) decreases

In fact, because

$$E'_{ec,p} = E_p + V_{DC} \quad (5.1.6)$$

consequently

$$E'_{ec,p} > E_p = E_s > E \quad (5.1.7)$$

The cost of economic capital net of tax effect becomes

$$R'_{ec,p} = \frac{A(1 + R_A) - D(1 + r_f)}{A - D + V_p + V_{DC}} - 1 \quad (5.1.8)$$

Because (4.4.2.8), consequently

$$R_{E,p} \geq R'_{ec,p} \quad (5.1.9)$$

Corollary 6

If deposits are uninsured and limit liability applies the value of the equity $E_{ec}^{\prime s}$ and the cost of capital (net of tax effect) are the same of the insured case and therefore

$$E_{ec}^{\prime s} = E_{ec,p} \text{ and } R'_{ec,p}(D^s) = R'_{ec,p} \quad (5.1.11)$$

Prrof

Given that

$$E_{ec}^{\prime s} = A - D^s + V_{DC} = A - D + V_p + V_{DC} \quad (5.1.12)$$

therefore, as expected,

$$E_{ec}^{\prime s} = E'_{ec,p} \quad (5.1.13)$$

Moreover, results

$$R'_{ec,p}(D^s) = \frac{A(1 + R_A) - D^s(1 + r_f + s)}{A - D + V_p + V_{DC}} - 1 \quad (5.1.14)$$

because we have already shown in (4.4.2.11), results

$$R'_{ec,p}(D^s) = R'_{ec,p} \quad (5.1.15)$$

5.2 Tax and franchise value impact on cost of risk capital when deposits are uninsured

Let us first consider the tax impact. Obviously tax impact τ lower the equity value by an amount of capital $E\tau$ but increases the correspondent cost of economic capital.

$$E_\tau = A - D - E\tau = E - E\tau = E(1 - \tau) \quad \text{with } 0 < \tau < 1 \quad (5.2.1)$$

$$E_\tau = \tau \{AN(d_{1,rn}) - DN(d_{2,rn})\} \quad (5.2.2)$$

Remark

For the sake of simplicity, without loss of generality, we are considering that all equity is represented by “new” profit and consequently is taxed.

Given that the franchise value has a positive impact on the equity of the bank, we can rewrite the balance sheet condition as follows

$$E'_{ec,\tau} = E(1 - \tau) + V_{DC} \quad (5.2.3)$$

The cost of equity risk capital when tax effect and franchise value is include becomes

$$R'_{ec,\tau} = \frac{A(1 + R_A) - D(1 + r_f)}{E(1 - \tau) + V_{DC}} - 1 \quad (5.2.4)$$

5.3 Tax and franchise value impact on cost of risk capital when deposits are insured

If deposits are insured, knowing that, $E'_{ec,p} = A - D + V_{DC} + V_p$ the new equity values becomes

$$E'_{ec,\tau,p} = E'_{ec,p}(1 - \tau) = (A - D + V_{DC} + V_p)(1 - \tau) = E_{ec} \quad (5.2.5)$$

and the cost of economic capital becomes

$$R_{ec} = \frac{A(1 + R_A) - D(1 + r_f)}{(A - D + V_{DC} + V_p)(1 - \tau)} - 1 = \frac{\rho_{A,M} \Phi_M \sigma_A - \frac{V_p + V_{DC}}{A} (1 + r_f)(1 - \tau)}{\left(\frac{E}{A} - \frac{V_p + V_{DC}}{A}\right)(1 - \tau)} + r_f \quad (5.2.6)$$

Corollary 7

If deposits are uninsured the same equity and cost of economic capital of the insured case apply

If deposits are uninsured, recalling (4.4.2.3) and (4.4.2.4) we have

$$E_{s,\tau} = E_s(1 - \tau) \quad (5.2.8)$$

by adding the franchise value, we get

$$E'_{ec,s,\tau} = E_s(1 - \tau) + V_{DC} \quad (5.2.9)$$

knowing that

$$E'_{ec,s,\tau} = (A - D_s)(1 - \tau) + V_{DC} = (A - D + V_p)(1 - \tau) + V_{DC} = E_p(1 - \tau) + V_{DC} \quad (5.2.10)$$

then we get

$$E'_{ec,s,\tau} = E_p(1 - \tau) + V_{DC} \quad (5.2.11)$$

The same reasoning can be followed for the cost of equity and therefore the same results of the uninsured case apply

6- Capital structure and cost of target risk capital under limited and unlimited liability, insured deposits, franchise value and taxes

Let us now relax the assumption introduced in section 3 and consider the case when the target default probability (target credit rating) is not held constant: $p \neq p^*$ and $N(-d_2) \neq N(-d_2^*)$. As in the previous framework, r_f , σ_M and R_M are held constant and $R_M > r_f$. By following the same line of reasoning used to illustrate the fixed default probability case, in this section the cost of target economic capital will be analysed by considering the impact of the limited and unlimited liability, the deposit insurance, the franchise value and taxes on the cost of target risk capital.

6.1 The determination of promised debt repayments under unlimited liability and target credit rating

Recall (4.2.1), (4.2.2), (4.2.3) and (4.3.1), when the target credit rating $N(-d_2^*)$ is different from the actual one $N(-d_2)$

Proposition 7

The promised debt repayments are given by

$$D = \frac{A(1 + R_A^* - d_2^* \sigma_A^*)}{(1 + r_f)} \quad (6.1.1)$$

Proof

Knowing that

$$A^*(T) = A(1 + R_A^*) \quad (6.1.2)$$

A bank is in default, if at time T $A^*(T) < D(T)$ or, equivalently, $E < 0$. To remain solvent and reach the target credit rating $N(-d_2^*)$ at time T, the bank must comply with the following constraint:

$$A^*(T) - Ad_2^* \sigma_A^* \geq D(T) \quad (6.1.3)$$

Remarks

1. $Ad_2^* \sigma_A^*$ resembles the definition of the Capital at Risk, as introduced in section 2.

Therefore the book value and market value of the equity are equal by assumption.

Moreover through inspection of both (4.2.3) and (6.1.3) result that the book and market value of the equity are also equal, by construction, to the actual risk capital but different to the target risk capital. In this more general framework book value, market value and actual risk capital, from now on labelled E, are equal. Moreover the target risk capital will be labelled $E^*(T)$

2. Because in our framework the franchise value will be taken explicitly in consideration, the definition of economic capital will be target Capital at Risk, $E^*(T)$, plus Franchise Value V_{DC} ⁶⁶
3. The target risk capital will be labelled $E^*(T)$ is different from the (target) economic capital E_{ec}^* .
4. Hence we have not yet determined the Risk Adjusted Return on Economic Capital (RAROC) but proposition 7 will play a key role in this calculation.

Corollary 8

The cost of equity R_E is equal to the cost of actual risk capital but different from both the cost of target risk capital, namely R_E^* and the cost of (target) economic capital R_{ec}^*

$$R_E \neq R_E^* \neq R_{ec}^* . \quad (6.1.4)$$

The main difference being that the target default probability is different from the actual one and that the franchise value is not included in the cost of capital.

Proposition 8

If shareholder unlimited liability and target default probability is assumed, the cost of target risk capital R_E^* depends only on $\rho_{A.M}^*$, i.e $R_E^* = g(\rho_{A.M}^*)$ and is equal to

$$R_E^* = \frac{(1+r_f)\rho_{A.M}^*\Phi_M}{(d_2^* - \rho_{A.M}^*\Phi_M)} + r_f \Leftrightarrow R_E^* = \frac{(1+r_f)\rho_{A.M}^*\Phi_M}{(d_2^* - \rho_{A.M}^*\Phi_M)} + r_f = \frac{\rho_{A.M}^*\Phi_M + r_f d_2^*}{(d_2^* - \rho_{A.M}^*\Phi_M)} \quad (6.1.5)$$

Corollary

R_E^* is an increasing function of $\rho_{A.M}^*$.

The proofs are reported in the appendix

6.2 Capital Structure under unlimited liability and target credit rating

Proposition 9

⁶⁶ In practice, the term "economic capital" it is very often not clearly identified and, consequently, misunderstandings and misconceptions can induce top management towards wrong or biased decisions

Leverage, as represented by the $\frac{D}{A}$ ratio, is an increasing function of target default probability for any value of volatility and independent from correlation values and signs

Proof

Recall (6.1.1); let us consider the impact on leverage by taking the first derivative with respect of d_2^*

$$\frac{\partial \frac{D}{A}}{\partial d_2^*} = -\frac{\sigma_A^*}{(1+r_f)} < 0 \quad (6.2.2)$$

Through inspection of the first derivative function of leverage with respect of d_2^* we can see that:

- is independent of correlation,
- it results always negative because volatility and risk free-rate are always positive by construction.

Therefore, as expected, $\frac{D}{A}$ ratio is an increasing function of default probability for any value of volatility and independently on correlation values and signs

6.3 Cost of target risk capital under unlimited liability, limited liability and insured deposits

Proposition 10

The cost of target risk capital is positively (negatively) related with the target default probability $N(-d_2^*)$ if correlation $\rho_{A.M}^*$ is positive (negative) and it is independent from volatility σ_A^*

If the target default probability is not constant we have $R_E^* = h'(\sigma_A, d_2^*, \rho_{A.M}^*)$

We know from section 3 that, if unlimited liability and default probability is constant, the cost of equity risk capital is an increasing function of correlation and is independent from volatility. By relaxing the default probability constraint, recalling (4.3.13) we can see that:

$$\frac{\partial R_E^*}{\partial d_2^*} = -\frac{(1+r_f)\Phi_M\rho_{A,M}^*}{(d_2^* - \rho_{A,M}\Phi_M)^2} > 0 \quad \forall \rho_{A,M} < 0 \quad (6.3.1)$$

$$\frac{\partial R_E^*}{\partial d_2^*} = -\frac{(1+r_f)\Phi_M\rho_{A,M}^*}{(d_2^* - \rho_{A,M}\Phi_M)^2} < 0 \quad \forall \rho_{A,M} > 0 \quad (6.3.2)$$

therefore we have $\frac{\partial R_E^*}{\partial d_2^*} > 0$ if $\rho_{A,M}^* < 0$ and $\frac{\partial R_E^*}{\partial d_2^*} \leq 0$ if $\rho_{A,M}^* \geq 0$ therefore the target default

probability $N(-d_2^*)$ is positively related with the cost of equity if correlation is positive and vice-versa

Corollary 9 The cost of equity target risk capital is an increasing function of $\rho_{A,M}^*$ independently of default probability and σ_A^*

Recalling that in section 3 we have shown that

$$R_E^* = \frac{\Phi_M\rho_{A,M}^* + d_2^*r_f}{(d_2^* - \rho_{A,M}\Phi_M)} \quad (6.3.3)$$

by taking the partial derivative with respect of $\rho_{A,M}^*$ results

$$\frac{\partial R_E^*}{\partial \rho_{A,M}^*} = \frac{d_2^*}{(d_2^* - \rho_{A,M}\Phi_M)^2} > 0 \quad \forall -1 \leq \rho_{A,M}^* \leq 1 \text{ and } d_2^* \in \mathfrak{R}^+ \quad (6.3.4)$$

Therefore given that the numerator is always positive, the cost of equity is a decreasing function of correlation no matter the level of default probability.

The relationship of cost of target risk capital when deposits are insured, $R_{E,p}^*$, with the target default probability depends on the sign of the following relationship

$$\rho_{A,M}^*\Phi_M\sigma_A^* - \frac{V_p}{A}(1+r_f) \quad (6.3.5)$$

In fact, recalling (4.4.1.6) and rearranging we have

$$R_{E,p}^* = \frac{A(1+r_f)\left\{\rho_{A,M}^*\Phi_M\sigma_A^* - \frac{V_p}{A}(1+r_f)\right\}}{A\{d_2^*\sigma_A^* - (R_A^* - r_f)\} - V_p(1+r_f)} + r_f \quad (6.3.6)$$

By calculating the first derivative with respect of d_2^*

$$\frac{\partial R_{E,p}^*}{\partial d_2^*} = - \frac{A^2 \sigma_A^* (1+r_f) \left\{ \rho_{A,M} \Phi_M \sigma_A^* - \frac{V_p}{A} (1+r_f) \right\}}{\left\{ A \left\{ d_2^* \sigma_A^* - (R_A^* - r_f) \right\} - V_p (1+r_f) \right\}^2} \quad (6.3.6)$$

Therefore if (6.2.5) is positive, the derivative is negative and, consequently, the target default probability is an increasing function of the cost of equity risk capital. If, for example, the default probability is relatively high, say 5%, and the volatility of the asset is also high, about 20%, the put value becomes approximately equal to 4.5 and its future value, one year from now, at an interest rate of 5%, becomes almost 4.8.

Assuming, for example, that the value of the asset is fixed and equal to 1.000 this means that even in this “worst case” a minimum spread in one year time of 0,5% will be sufficient to generate a positive relationship between the target default probability and the cost of equity. This spread increases proportionally with the level of σ_A^* . As expected the riskier the investment, the higher the expected cost of equity risk capital and the higher the PD. Moreover, in this framework, another crucial parameter is the sign of correlation. If the correlation is negative, the derivative is negative and the default probability is positively related with the cost of equity. If the correlation is positive the default probability is negatively related with the cost of equity (as was the case when default was constant).

Corollary 10 Same relationship applies when deposits are uninsured

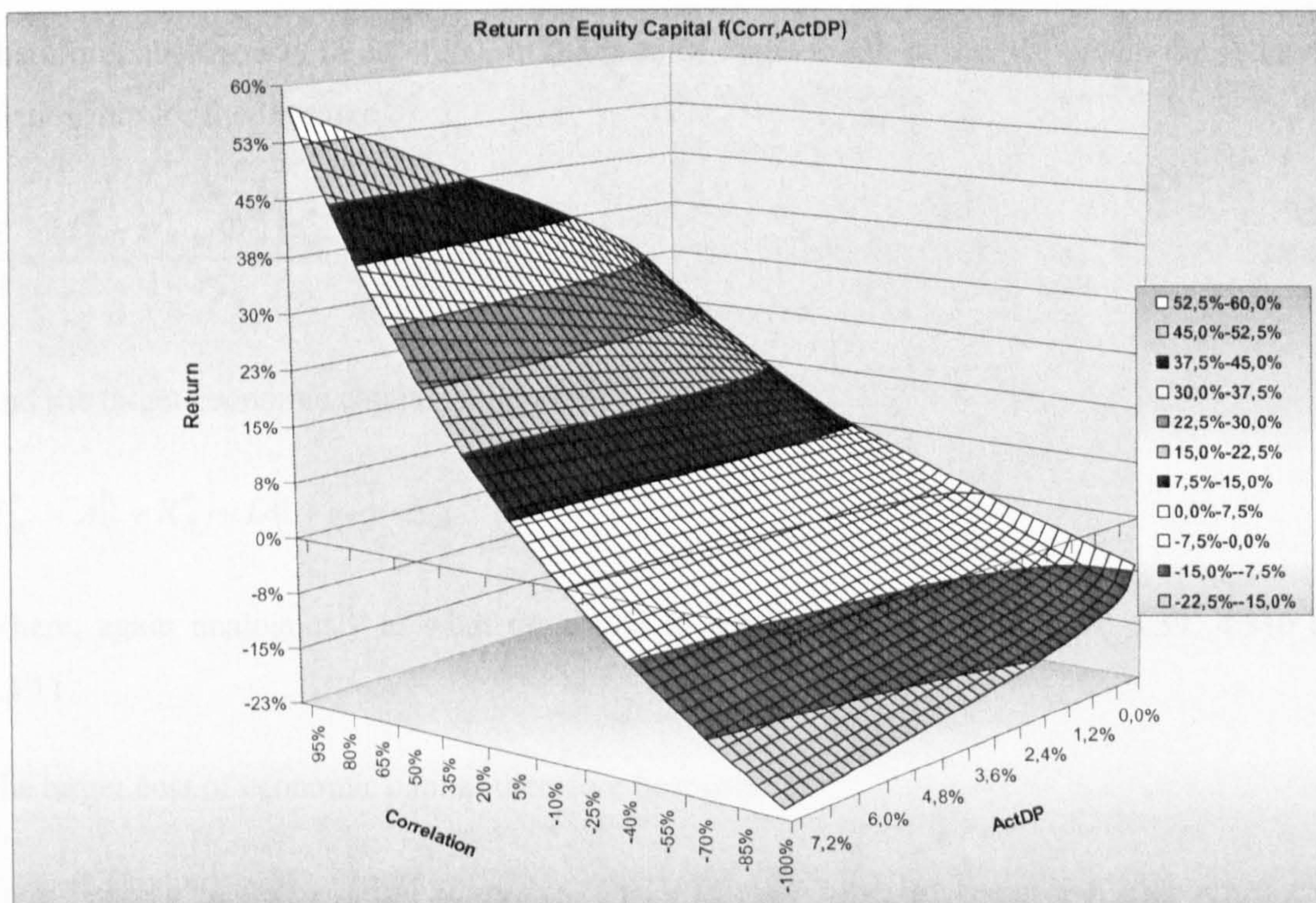


Figure 7 Cost of equity as a function of correlation and PD

6.4 Cost of target economic capital under limited liability, insured deposits, franchise value and taxes

In the general case, as illustrated in the last paragraph of section 4 we showed that

$$R_{ec} = \frac{A(1 + R_A) - D(1 + r_f)}{(A - D + V_{DC} + V_p)(1 - \tau)} - 1$$

and through 5.2.10 and (4.4.2.8) we have

$$R_{ec} = \frac{\rho_{A,M} \Phi_M \sigma_A - \frac{V_p + V_{DC}}{A} (1 + r_f)(1 - \tau)}{\left(\frac{E}{A} - \frac{V_p + V_{DC}}{A} \right) (1 - \tau)} + r_f \quad (6.4.1)$$

Moreover, knowing that, in equilibrium eq. 4.1.1.2 becomes

$$\Phi_A^* = \frac{R_A^* - r_f}{\sigma_A^*} = \rho_{A,M}^* \frac{R_M - r_f}{\sigma_M} = \rho_{A,M}^* \Phi_M$$

The following relationship must hold in equilibrium

$$\frac{R_A - r_f}{\sigma_A \rho_{A,M}} = \frac{R_A^* - r_f}{\sigma_A^* \rho_{A,M}^*} = \Phi_M$$

Therefore, analogously to eq. 4.3.9, in the case of target credit rating, we obtain the following relationship for the leverage

$$\frac{E}{A} = \frac{(d_2^* - \rho_{A,M}^* \Phi_A^*) \sigma_A^*}{1 + r_f} \quad (6.4.2)$$

And the target economic capital is therefore

$$E_{ec}^* = A(1 + R_A^*) - D(1 + r_f) - E_{ec}$$

Where, again analogously to what described in section 4.2 eq. 4.2.3, 4.34, 4.39, 4.310 and 4.3.11

The target cost of economic capital therefore becomes

$$R_{ec}^* = \frac{A(1 + R_A^*) - D(1 + r_f) - (A - D + V_p + V_{DC})(1 - \tau)}{(A - D + V_p + V_{DC})(1 - \tau)} = \frac{A(1 + R_A^*) - D(1 + r_f) - E_{ec}}{E_{ec}}$$

Consequently, we can rewrite (6.4.1) in the case that the actual default probability is different from the target one:

$$R_{ec}^* = \frac{\rho_{A,M}^* \Phi_M \sigma_A^* - \frac{V_p + V_{DC}}{A} (1 + r_f)(1 - \tau)}{\left(\frac{(d_2^* - \rho_{A,M}^* \Phi_A^*) \sigma_A^* - \frac{V_p + V_{DC}}{A}}{1 + r_f} \right) (1 - \tau)} + r_f \quad (6.4.3)$$

Therefore the relationship of cost of target economic capital R_{ec}^* with the target default probability depends on the sign of the following relationship

$$\rho_{A,M}^* \Phi_M \sigma_A^* - \frac{V_p + V_{DC}}{A} (1 + r_f)(1 - \tau) \quad (6.4.4)$$

If (6.4.4) is positive, the derivative is negative and therefore the target default probability is an increasing function of R_{ec}^* . Interestingly when the default probability is low the put value has a very low impact on R_{ec}^* . If the franchise value is added the order of magnitude of the put impact (slightly) changes because of the negative tax effect.

7- Conclusions and further research

In this chapter we have shown that the main risk drivers which allows the bank to enhance the cost of risk capital in the most simple world without frictions is the market correlation and (contrary to Crouhy et. al. (1999)) not the volatility (see eq. 4.3.12).

By introducing adjustments due to the insurance put for the bank deposits (see eq. 4.4.1.6 and 4.4.1.10.), the franchise value and the taxes the volatility plays some role especially for the riskier banks, namely the ones with the lowest credit rating, as reported in eq. 5.1.8 and 5.1.14.

In this chapter we have also shown how to estimate the capital structure by considering at the same time the target credit rating and the target cost of capital (see eq. 6.4.3)

Moreover, as illustrated in section 6, this framework clearly establishes the relationship amongst different measures of capital, namely the market value of the equity, the risk capital and the economic capital. It is shown that:

1. $Ad_2\sigma_A$ resembles the definition of the Capital at Risk, as introduced in section 2 and $Ad_2^*\sigma_A$ resembles the definition of the target Capital at Risk (here d_2 is the actual

“distance to default” in terms of standard deviations of asset returns; while d_2^* is the target “distance to default” in order to achieve the required credit rating.)

2. under our equilibrium assumption, the book value and market value of the equity are equal.
3. we show, (see (4.2.3) and (6.1.3)), that both the book and market value of the equity are equal to the actual risk capital
4. we show (again see (4.2.3) and (6.1.3)) that the book and market value of the equity are, in general, different from the target risk capital E^* .
5. this implies that the target risk capital E^* is different from the target economic capital E_{ec}^* .

These last two results hold when, in this framework, the franchise value is taken explicitly in consideration, reflecting the assumed definition of economic capital as target Capital at Risk plus Franchise Value. Consequently, another important result of the analysis, illustrated in corollary 8, is that the cost of equity R_E is equal to the cost of actual risk capital but different from both the cost of target risk capital, namely R_E^* and the cost of (target) economic capital R_{ec}^* (see eq. 6.1.4). These differences in required return reflect the findings that the target default probability is different from the actual one and that the franchise value is not included in the cost of capital.

In a further research the problem of the optimal capital structure will need to be further addressed. The hurdle rate for the bank can be viewed as the weighted average cost of capital (WACC) which will explicitly take in consideration the debt to equity ratio. At the same time the un-levered beta needs to be introduced in the capital asset pricing model. This extended framework will give the possibility to link the EVA analysis with the RAROC or cost of economic capital analysis. The optimal capital structure then will be associated with the minimum WACC which makes maximum the value of the bank. The analysis can be further extended to take account of illiquid bank assets and non-normal asset returns (see Milne and Onorato (2004)); and to the case of multiple business units within the bank. In this latter case the interaction of correlations will play a role at different level of the organisation, namely:

- among business units,
- between business unit and the bank portfolio,

- between business units and the market,
- between the bank portfolio and the market.

It will be the chief risk officer that needs to take responsibility of this capital allocation process and transform the risk management function from a pure cost centre division to a truly business function division. The risk management function will be therefore compensated for the ability to allocate consistently the economic capital across BU on the basis of the illustrated risk adjusted performance and on the basis of the ability to manage all the previously identified correlations.

Appendix 1

If shareholder unlimited liability and target default probability is assumed, the cost of equity target risk capital R_E^* depends only on $\rho_{A,M}$, i.e. $R_E^* = g(\rho_{A,M})$ and is equal to

$$R_E^* - r_f = \frac{(1+r_f)\rho_{A,M}\Phi_M}{(d_2^* - \rho_{A,M}\Phi_M)}$$

Proof

In equilibrium, as shown in (4.1.3), recalling (4.1.4) equation (4.3.11) becomes

$$\frac{E}{A} = 1 - \frac{D}{A} = \frac{d_2^*\sigma_A - (R_A^* - r_f)}{1+r_f} = \frac{d_2^* - \Phi_A^*}{\frac{1+r_f}{\sigma_A}} = \frac{(d_2^* - \rho_{A,M}^*\Phi_M)\sigma_A}{1+r_f}$$

consequently

$$R_E^* - r_f = \frac{A(1+R_A^*) - D(1+r_f)}{A-D} - 1 - r_f = \frac{A(R_A^* - r_f)}{A-D} = \frac{A(R_A^* - r_f)}{E} = \rho_{A,M}^*\Phi_M\sigma_A^*$$

rearranging

$$R_E^* - r_f = \frac{A(R_A^* - r_f)}{E} = \frac{(1+r_f)(R_A^* - r_f)}{(d_2^* - \rho_{A,M}^*\Phi_M)\sigma_A^*} = \frac{(1+r_f)\rho_{A,M}^*\Phi_M\sigma_A^*}{(d_2^* - \rho_{A,M}^*\Phi_M)\sigma_A^*} = \frac{(1+r_f)\rho_{A,M}^*\Phi_M}{(d_2^* - \rho_{A,M}^*\Phi_M)}$$

Corollary

R_E^* is an increasing function of $\rho_{A,M}^*$.

Because in this case we have assumed the risk-free rate, the market and the volatility of the market return, the probability of default constant results

$$R_E^* = \frac{(1+r_f)\Phi_M\rho_{A,M}^*}{(d_2^* - \rho_{A,M}^*\Phi_M)} + r_f = \frac{\Phi_M\rho_{A,M}^* + d_2^*r_f}{(d_2^* - \rho_{A,M}^*\Phi_M)}$$

by taking the partial derivative with respect of $\rho_{A,M}^*$ results

$$\frac{\partial R_E^*}{\partial \rho_{A,M}^*} = \frac{(1+r_f)\Phi_M(d_2^* - \rho_{A,M}^*\Phi_M) + \Phi_M(1+r_f)\Phi_M\rho_{A,M}^*}{(d_2^* - \rho_{A,M}^*\Phi_M)^2} = \frac{(d_2^*)}{(d_2^* - \rho_{A,M}^*\Phi_M)^2} > 0 \quad \forall \rho_{A,M}^*$$

Clearly, because the derivative is always positive for any value of $\rho_{A,M}^*$ R_E^* it is shown to be an increasing function of $\rho_{A,M}^*$. Summarising, in this set up, the R_E^* is independent on σ_A^* and an increasing function of $\rho_{A,M}^*$

References

Acharya, V.V., Das S.R. and Sundaram R.K. (2002) Pricing Credit Derivatives with rating transitions, *Financial Analysts Journal*, May-Jun, pp. 28-42.

Acerbi C. and Tasche D. (2002) On the coherence of Expected Shortfall. *Journal of Banking & Finance* 26(7), pp. 1487- 1503.

Aggarwal R. and L. Soenen. (1988) The nature and efficiency of the gold markets. *Journal of Portfolio Management*, 14, pp.18-21, Spring .

Aggarwal, R. and J. D. Schtzberg. (1997) Day of the week effect, information seasonality, and higher moments of security returns. *Journal of Economics and Business*, 49, pp.1-20.

Aggarwal, R. and R., Rao. (1990) Institutional ownership and distribution of equity returns. *Financial Review*, 25 pp.211-229.

Aggarwal, R., Ramesh R., and H. Takato (1989) Skeweness and kurtosis in Japanese equity returns: Empirical evidence. *Journal of Financial Research*, 12 pp.253-260.

Aguais S., Forest L. and Lawrence R. (2000) The future of Risk-Adjusted Credit Pricing in Financial Institutions, *The RMA Journal*.

Ahn, M., J., and Falloon, W.D. (1991) *Strategic risk management*, Chicago.

Alexander C. (2001) *Market models*, Ed John Wiley & Sons, LTD .

Altman E. I. (1998) *The Anatomy of the High Yield Bond Market*, Global Corporate Bond Research, Salomon Smith Barney.

Altman E.I., Onorato M., and Pastorello A., (2000) *A General Framework for Credit Risk Models*, Euro Working Group on Financial Modelling.

Altman E.I., and Kishore V.M., (1996) Almost everything you wanted to know about recoveries on defaulted bond, *Financial Analysts Journal*, Nov-Dec, pp. 57-63.

Altman E.I., Caouette J.B., and Narayan P., (1998) *Managing Credit Risk: The Next Financial challenge*, Ed. John Wiley & Sons, Inc.

Ang J.S., and Chua J.H., (1979) Composite Measures for the Evaluation of Investment Performance, *Journal of Financial and Quantitative Analysis* 14, pp. 361-384.

Arditti F., (1967) Risk and Required Return on Equity, *Journal of Finance* 22, pp. 19-36.

Arditti F., (1971) Another look at mutual fund performances. *Journal of Financial and Quantitative Analysis*, 6, pp. 909-912.

Arrow, K.J., (1971) *Essays in the Theory of Risk-Bearing*, Markham Publishing Co., Chicago.

Artzner P., F., Delbaen, J.M., Eber, and D. Heath (1997) Thinking coherently. *Risk*, 10(11), pp.68-70.

Artzner P., F., Delbaen, J.M., Eber and D., Heath (1999) Coherent Measures of Risk, *Mathematical Finance* 9/3, pp. 203-228.

Artzner Ph., (1999) Application of coherent risk measures to capital requirements in insurance, *North American Actuarial Journal* 3(2), pp.11-25.

Banfield, E. (2003) *What is Economic Capital? A quick guide to the differences between Economical Capital and Regulatory Capital*, Chicago.

Bansal A., Kauffman R.J., Mark R.M. and Peters E. (1993) Financial risk and financial risk management technology, *Information and Management*, pp. 267-281.

Barnett V. (1975) The ordering of multivariate data. *Journal of the Royal Statistics Society*, (139), pp.318-344

Basle Committee on Banking Supervision (1988) *International convergence of capital measurement and capital standards*, Basle: Switzerland.

Basle Committee on Banking Supervision (1995) *An internal model-based approach to market risk capital requirements*, Basle: Switzerland.

Basle Committee on Banking Supervision (1996) Amendment to the capital accord to incorporate market risks, Basle: Switzerland.

Basle Committee on Banking Supervision (1998) Operational risk management, Basle: Switzerland.

Basle Committee on Banking Supervision (1999) A new capital adequacy framework, Basle: Switzerland.

Basle committee on Banking Supervision (1999) Credit Risk modeling: current practice and applications, Basle: Switzerland.

Basle committee on Banking Supervision (1999) Principles for credit risk management, Basle: Switzerland.

Bassi F., P., Embrechts, and M., Kefetzaki (1993) A survival kit on quantile estimation. Talk given for Olsen & Associates.

Bauer W., (1998) Implementing and managing risk-adjusted return on risk-adjusted capital, Bank Treasurer Manual, Sheshunoff Information Services.

Bawa, V.S., (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics* 2, pp. 95-121.

Bawa, V.S., (1978). Safety-first, stochastic dominance and optimal portfolio choice, *Journal of Financial and Quantitative Analysis*, pp. 255-271.

Bawa, V.S., Lindenberg, E.B., (1977). Capital market equilibrium in a mean±lower partial moment framework. *Journal of Financial Economics* 5, pp. 189-200.

Bawa, V.S. (1982): Stochastic dominance: a research bibliography. *Manage. Sci.* 28, 698–712

Beaver, W.H., and Parker, G., (1995) Risk Management, problems and solutions, McGraw-Hill, New York.

Benninga, S. Z., 1997, 'Financial modeling', MIT Press, Cambridge, MA.

Bennett S. G. (1994) EVA: Fact and Fantasy, Stern Stewart & Co, Journal of Applied Corporate Finance, volume 7 no 2.

Berenyi Z., (2001) Accounting for illiquidity and non-normality of returns in the performance assessment, working paper.

Berger A.N., Herring R.J., and Szego G.P., (1995) The role of capital in financial institutions, Journal of Banking and Finance 19, pp. 393-430.

Berger V., (2001) The probability problem solver , Research and Education Association.

Billingsley, P. (1995): Probability and Measure. John Wiley & Sons, New York

Black F., Derman E., and Toy W., (1990) A one-factor model of interest rates and its application to treasury bond option, Financial Analysts Journal, Jan-Feb, pp. 33-39.

Bookstaber R., and R., Clarke (1983) Option portfolio strategies: Measurement and evaluation, Journal of Business, 57(4).

Bralver C. and Kuritzkes A. (1998) Risk Adjusted Performance measurement in the trading room, Oliver, Wyman & company, Journal of Applied Corporate Finance.

Brealey R.A., and Myers S.C., (2000) Principles of Corporate Finance, 6th edition, McGraw Hill, New York.

Bremaud P., (1994) An introduction to probabilistic modeling , Springer-Verlang.

Campbell J.Y., A.W., Lo, and A.C., MacKinlay (1997) The Econometrics of Financial Markets. Princeton University Press.

Campbell, R., R., Huisman and K., Koedijk (2001) Optimal Portfolio Selection in a Value-at-Risk Framework, Journal of Banking & Finance 25, pp.1789-1804.

Carty L.V., Lieberman D., (1996) Corporate Bond defaults and Default Rates 1938-1995, Moody's Investor Service, Global Credit Research, Special Report .

Carverhill. A., (1995) A Simplified Exposition of the Heath, Jarrow, and Morton Model
Stochastics 53, pp.227-240.

Cass, D. and J.E. Stiglitz (1970) The Structure of Investor Preferences and Asset Returns, and
Separability in Portfolio Allocation, Journal of Economic Theory 2/1, pp. 122.

Castle K., Keisman D., and Yang R., (2000) Suddenly structure mattered: insights into
recoveries from defaulted debt Standard & Poor's, credit weekly.

Chamberlain G., (1983) A characterization of the distributions that imply mean-variance
utility functions. Journal of Economic Theory, 29(1) pp.185-201.

Chau T. (2004) Optimal Capital Structure and Hedging, GTNews, China.

Chen Z., and Knez PJ., (1996) Portfolio Performance measurement pp. theory and
applications, The Review of Financial Studies 9, no 2, pp. 511-555.

Crouhy M., Galai D., and Mark R., (2000) Risk Management, Ed McGraw-hill

Crouhy M., Turnbull S. and Wakeman L. (1999) Measuring risk –adjusted performance
Journal of Risk Volume 2/N°1, pp. 5-35.

Cochrane J.H., and Saá-Requejo J., (2000) Beyond Arbitrage: Good-Deal Asset Pricing
Bounds in Incomplete Markets. Journal of Political Economy 108, pp. 79-119.

Cooper I.A., and A.S., Mello (1996) Default Risk and Derivative Products, Applied
Mathematical Finance 3, pp. 53-74.

Coopers I.A., Lybrand and the British Bankers' Association (1997) Operational risk
management survey.

Copeland T., Koller T., and Murrin J., (1990) Valuation: Measuring and managing the value
of companies, McKinsey & Company, Inc., 13, pp. 375-397.

Corrado C.J., and T., Su. (1996) Skewness and kurtosis in s&p 500 index returns implied by
option prices. The Journal of Financial Research, XIX(2) pp.175-192.

Credit Suisse (1997) Financial Products, CreditRisk+: A Credit Risk Management Framework.

CreditMetrics TM (1997) Technical document JP Morgan.

Crosbie P., (1998) Modelling Default Risk, KMV Corporation.

Crouhy M., and Galai D., (1986) An economic assessment of capital requirements in the banking industry, *Journal of Banking and Finance* 10, pp. 231- 241.

Crouhy M., Galai D., and Mark R. (1999) Measuring Risk-adjusted performance, *Journal of Risk*, Vol 2 no 1, Fall, pp. 5-35.

Crouhy M., Galai D., Mark R., (2000) A comparative analysis of current credit risk models, *Journal of Banking & Finance* 24, pp. 59-117.

Crouhy M., Galai D., and Mark R. (2000) *Risk Management*, McGraw-Hill.

C. L. Culp (2001) *The risk management process, Business strategy and tactics*, John Wiley & Sons Inc.

Dai Q. and K. Singleton (1998), *Specification Analysis of Affine Term Structure Models*, Research Paper, Graduate School of Business, Stanford University.

Damodaran, A. (2000) *Investment Valuation: Tools and techniques for determining the value of any asset*, 2nd edition, Wiley, New York.

Damodaran, A. (2002) *Corporate Finance*, second edition, Wiley, New York.

Das S.R., (1995) Credit risk derivatives, *Journal of derivatives*, v2 (3), pp. 7-23.

Das S.R., Sundaram R.K., (1999) A Discrete Time-Approach to Arbitrage-Free Pricing of Credit Derivatives.

Das S.R., Tufano P., (1996) Pricing credit sensitive debt when interest rates, credit ratings and credit spreads are stochastic, *Journal of Financial Engineering*, 5, No. 2, pp. 161-198.

David H.A., (1981) Order Statistics. John Wiley & Sons, New York, 2nd edition.

De Groot, J. Sebastiaan. (1998), Behavioral Aspects of Decision Models in Asset Management. Labyrinth Publication, The Netherlands,

De Haan L., Jansen D.W., Koedijk K., and C.G., de Vries (1994) Safety first portfolio selection, extreme value theory and long run asset risks. Extreme Value Theory and Applications.

Dembo R., (1993) Scenario Immunization, in Financial Optimization, Zenios S.A., Cambridge University Press.

Dhaene j., Denuit, m., Goovaerts, m.j., Kaas, r. and Vyncke, d. (2002) The concept of comonotonicity in actuarial science and finance: theory. Insurance Math. Econom. 31(1), pp.3-33.

Dhaene j., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002) The concept of comonotonicity in actuarial science and finance: application. Insurance Math. Econom. 31(2), pp. 133-161.

Dhaene J., Laeven R.J.A., Vanduffel S., Darkiewicz G., Goovaerts M.J., (2004) Can a coherent risk measure be too subadditive?, Catholic University of Leuven Belgium and Center for Risk and Insurance Studies, Leuven, and University of Amsterdam.

Dhaene J., Vanduffel S., Tang Q., Goovaerts M.J, Kaas R., and Vyncke D., (2003) Risk measures and optimal portfolio selection, working paper.

Diederich A., and J. R., Busemeyer (1998) Conflict and the Stochastic Dominance Principle of Decision Making. Indiana University Cognitive Science Program Tech Report #224.

Dittmar R.F., (1999) Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross-section of Equity Returns, Technical report, Kelley School of Business, Indiana University, Bloomington .

Doksum K., D. Peterson, and A., Samarov (1998) On bandwidth selection in local polynomial regression. submitted to the Journal of the Royal Statistical Society, series B.

Dowd, K., (1999) A Value-at-Risk Approach to Risk-Return Analysis, *The Journal of Portfolio Management* Spring, pp.60-67.

Dowd, K., (2000) Adjusting for Risk: An Improved Sharpe Ratio, *International Review of Economics and Finance* 9, pp.209-222.

Dowd, K., (1998) VaR by Increments, *Enterprise-Wide Risk Management Special Report, RISK*, November, pp. 31-32.

Dowd, K., (1999) A Value-at-Risk Approach to Risk-Return Analysis, *The Journal of Portfolio Management* Spring, pp.60-67.

Dowd, K., (2000) Adjusting for Risk: An Improved Sharpe Ratio, *International Review of Economics and Finance* 9, pp.209-222

Drzik J., Nakada P. and Schuermann T. (1998) Risk Capital Measurement in financial institutions, Part I, *ERisk*.

Drzik J., Nakada P. and Schuermann T. (1998a), 'Risk, capital, and value measurement in financial institutions—Part I: the debtholder's Perspective', *The Journal of Lending and Credit Risk Management*, September 1998, pp. 22-27

Drzik J., Nakada P. and Schuermann T. (1998b), 'Risk, capital, and value measurement in financial institutions—Part II: the shareholder's perspective', *The Journal of Lending and Credit Risk Management*, November 1998, pp. 64-69

Duffee G., (1999) Estimating the price of default risk, *Review of Financial Studies*, v12(1), pp. 197-226.

Duffee G., (1998) The relation between treasury yields and corporate bond yield spreads, *Journal of Finance*, December.

Duffie D., (1999) Credit swap valuation, *Financial Analysts Journal*, Jan-Feb, pp. 73-87.

Duffie D., and D., Lando (1997) *The Term Structure of Credit Spreads with Incomplete Accounting Information*, Graduate School of Business, Stanford University.

Duffie D., and Singleton K., (1996) Modeling term structure of defaultable bonds, *Review of Financial Studies*.

Duffie D., Pan J., and K. Singleton (1998) *Transform Analysis and Option Pricing for Affine Jump Diffusions*, Graduate School of Business, Stanford University.

Dybvig, P.H. (1988), 'Distributional Analysis of Portfolio Choice', *Journal of Business* 61, pp. 369-393.

Dybvig P.H., and Ingersoll J.E., (1982), 'Mean-Variance Theory in Complete Markets', *Journal of Business* 55, pp. 233-251.

Dybvig P.H., (1988) 'Distributional Analysis of Portfolio Choice', *Journal of Business* 61, pp. 369-393.

Dyer J.S., (1987) *Multiple Criteria Decision Making and Risk Analysis*, volume 56 of NATO ASI Series F: Computer and Systems Sciences, chapter The Effects of Risk on Decision Making. Springer-Verlag.

Dyer, J.S., and Jia J., (1996) 'A Standard Measure of Risk and Risk-Value Models', *Management Science* 42, pp. 1691-1705.

Eeckhoudt L., and Gollier C., (1995) *Risk - Evaluation, Management and Sharing*, Prentice Hall, Harvester Wheatsheaf, New York.

Elton, E. J. and M. J. Gruber, 1995, *Modern portfolio theory and investment analysis*, John Wiley & Sons.

Embrechts P., Mcneil, A.J., and Straumann, D., (2002) Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pp. 176-223, Cambridge Univ. Press, Cambridge.

Encarnacion J., Jr., (199) Portfolio choice and risk. *Journal of Economic Behaviour and Organization*, 16 pp.347-353.

Etemadi N., (1981) An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 1(55) pp.119-122.

Falkenstein E. (1997) Accounting for Economic and Regulatory Capital in RAROC Analysis, Bank Accounting and Finance, Volume 11 no 11.

Fama E., and K., French. (1992) The cross-section of expected stock returns. The Journal of Finance, XLVII(2) pp.427-465.

Fama E., (1965) The behaviour of stock markets prices. Journal of Business, 38 pp.34-105.

Fama E., (1996) 'Multifactor Portfolio Efficiency and Multifactor Asset Pricing.

Fan J., and I., (1996) Gijbels Local polynomial modelling and its applications. Monographs on statistics and applied probability. Chapman & Hall.

Fang H. and T.-Y. (1997) Lai. Co-kurtosis and capital asset pricing. The Financial Review,32(2) pp.293-307.

Federal Reserve, Bank of Boston Economic Capital Survey Results, Boston. (2003) Capital and Risk: New evidence on implications of large operational losses

Feldstein M. S. (1969) Mean-variance analysis in the theory of liquidity preference and portfolio selection. Review of Economic Studies, 36(105) pp.5-12.

Fishburn P. (1977) Models of individual preference and choice, Synthese 36, pp.287-314.

Fishburn P. (1977), Expected utility theories: a review note, in Mathematical Economics and Game Theory (ed. R. Henn & O. Moeschlin), Springer-Verlag, pp.197-207.

Fishburn P. (1977), Mean-risk analysis with risk associated with below-target returns, American Economic Review 67, pp. 116-126.

Fishbourn P. and G. A. Kochenberger (1979) Two-piece von Neumann-Morgenstern utility functions, Decision Sciences 10 , pp.503-518.

Fogiel M., (2000) The statistic problem solver, Research and Education Association.

Fortune, New York (1997) Debate: Duking It Out over EVA Gary Hamel; Al Ehrbar. August 4th.Framework', Management Science 39, pp. 568-577

Froot K., and Stein, J., (1998) Risk Management, Capital Budgeting and Capital Structure Policy for financial institutions: An Integrated Approach, *Journal of Financial Economics*, 4, pp. 55-82.

Garman M., (1996) Improving on VaR. Risk.

Gerber H.U., (1974) On additive premium calculation principles. *ASTIN Bulletin* 7, pp. 215-222.

Giles T. and Milne A. (2003) An alternative view of the Basel II, rules for calculating regulatory capital, Working paper.

Goetzmann W., Ingersoll J., Matthew Spiegel and Ivo Welch (2002) Sharpening Sharpe Ratios, Working Paper 9116, <http://www.nber.org/papers/w9116>, national bureau of economic research.

Goovaerts, M.J., de Vijlder, f. and Haezendonck, J. (1984) Insurance Premiums. North-Holland Publishing Co., Amsterdam.

Goovaerts, M.J., Kaas, R. and Dhaene, J., (2003) Economic capital allocation derived from risk measures, *North American Actuarial Journal*.

Goovaerts MJ., Rob Kaasb, Jan Dhaenea,b, and Qihe Tang (2004) Some New Classes of Consistent Risk Measures; Center for Risk and Insurance Studies (CRIS), Department of Quantitative Economics, University of Amsterdam and Katholieke Universiteit Leuven, Belgium.

Goovaerts MJ., Rob Kaas, Jan Dhaene, and Qihe Tang (2003) A unified approach to generate risk measures. Working paper

Gordy M.B., (2000) A comparative anatomy of credit risk models, *Journal of Banking and Finance* 24 119-149.

Grace A., (1995) *Optimisation Toolbox*. The MathWorks Inc.

Haan L.. de, Jansen D.W., Koedijk K. and de Vries C.G., (1994) *Extreme Value Theory and Applications*.

Haan L. de, Jansen D.W., Koedijk K. and de Vries C.G., (1995) Safety first portfolio selection, extreme value theory and long run asset risks.

Harlow W.V., (1991) Asset allocation in a downside-risk framework. *Financial Analysts Journal*, September.

Harlow W.V., and Rao (1989), Asset-pricing in a generalized mean-lower partial moment framework: theory and evidence; *Journal of financial quantitative analysis*, 24(3), pp.285-311

Harlow W.V.(1991) W. Van, Asset Allocation in a Downside Risk Framework, *Financial Analysts Journal*

Harris M., Kriebel C., and Raviv A., (1982) The Capital Budgeting Process, Incentives and Information, *Journal of Finance*.

Harrison M., and D., Kreps (1979), Martingales and Arbitrage in Multiperiod Security Markets, *Journal of Economic Theory* 20, pp. 381-408.

Grootveld H. and Hallerbach W. (1999), Variance vs downside risk: Is there really that much difference? *European Journal of Operational Research* 114, pp. 304-319.

Hillier F.S., and G.J., Lieberman (1995) *Introduction to Mathematical Programming*. McGraw-Hill, Inc.

Hodges S.D., (1997) A generalisation of the Sharpe ratio and its applications to valuation bounds and risk measures, Working paper of the Financial Options Research Centre, university of Warwick, June .

Hollifield, B., and Kraus A., (1998) Defining Bad News: Changes in Return Distribution that Induce a Decrease in Asset Demand, Technical report, GSIA Carnegie Mellon University <http://www.mcs.anl.gov/otc/Guide/CaseStudies/port/demo.html>, 1998.

Huang C.-F and R.H., Litzenberger (1988) *Foundations for Financial Economics*. Prentice-Hall.

Huge B., and Lando D., (1999) Swap Pricing with Two-Sided Default Risk in a Rating-Based Model, Working Paper, Denter for Analytical Finance, University of Copenhagen.

Hull J., and A., White (1995), The Impact of Default Risk on the Prices of Options and Other Derivative Securities, *Journal of Banking and Finance* 19, pp. 299-322.

Hull J., (2003) *Options, Futures, and Other Derivatives*, 5th edition, Pearson education international.

Hürlimann W., (2002) An economic capital allocation for lookback financial losses, Working paper, Switzerland.

Hürlimann W., (2002) Distortion Risk Measures And Economic Capital, Working paper, Switzerland.

Hurvich C.M., J.S., Simono and C.-L. Tsai. (1998) Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal R. Statist. Soc. B*, 60, pp. 271-293.

Hwang S. and Satchell S.E., (1999) 'Modeling Emerging Market Risk Premia Using Higher Moments', *International Journal of Finance and Economics* 4, pp. 271-296.

Ingersoll J. (1975), 'Multidimensional Security Pricing', *Journal of Financial and Quantitative Analysis* 10, pp. 785-798.

Ingersoll J.E., (1987) *Theory of financial decision making*. Rowman & Littlefield studies in financial economics. Rowman & Littlefield.

Jacobs B. and K. Levy. (1996) Residual risk: How much is too much. *The Journal of Portfolio Management*, Spring.

James C. (1996), RAROC Based Capital budgeting and performance evaluation: a case study of Bank capital allocation , The Wharton School, University of Pennsylvania

Jameson R. (2001) Between RAROC and a hard place, ERisk.

Heath D., Jarrow R., and Morton A., (1992) Bond pricing and the term structure of interest rates. A new methodology for contingent claim valuation, *Econometrica* 60, pp. 77-105.

Jarrow R. and S. Turnbull (1995) Pricing Options on Financial Securities Subject to Default Risk, *Journal of Finance* 50, pp. 53-86.

Jarrow R.A., Lando D., Turnbull S.M., (1997) A Markov Model for the Term Structure of Credit Risk Spreads, *The Review of Financial Studies* Vol. 10, No. 2, pp. 481-523.

Jarrow R., Turnbull S.M., (2000), The intersection of market and credit risk, *Journal of Banking and Finance*.

Jean W.H. (1971) 'The Extension of Portfolio Analysis to Three or More Parameters', *Journal of Financial and Quantitative Analysis* 6, pp. 505-515.

Jean W.H., (1973) 'More on multidimensional portfolio analysis', *Journal of Financial and Quantitative Analysis* 8, pp.475-490.

Jean, W.H., (1973) 'More on multidimensional portfolio analysis', *Journal of Financial and Quantitative Analysis* 8, pp.475-490.

Jean, W.H., (1971) 'The Extension of Portfolio Analysis to Three or More Parameters', *Journal of Financial and Quantitative Analysis* 6, pp. 505-515.

Bessis J. (2002) Risk management in banking, Ed John Wiley & Sons, LTD .

Johnson N.L., and S. Kotz. (1969) *Distributions in Statistics: Continuous Multivariate Distributions*, volume 4. John Wiley & Sons.

Jorion P., (1997) *Value at risk : the new benchmark for controlling market risk*. Irwin Professional Pub, Chicago.

Kaas R., Heerwaarden, van A.E., and M.J. Goovaerts (1994). *Ordering of Actuarial Risks.*, CAIRE Education Series 1, Brussels.

- Kaufman A. and A. Henry-Labordère (1977) Integer and Mixed Programming, Theory and Applications, volume 137 of Mathematics in Science and Engineering. Academic Press, New York.
- Kealhofer S., (1998) Portfolio Management of Default Risk, KMV Corporation,.
- Kimball M.S., (1993), 'Standard Risk Aversion', *Econometrica* 61, pp. 589-611
- Kimball R. (1997) Innovations in Performance Measurement in Banking, New England Economic Review.
- Kisgen D. (2003) Credit Ratings and Capital Structure, University of Washington, School of Business Administration.
- Klupperberg C.P., Embrechts and T. Mikosch (1997). Modelling Extremal Events, volume 33 of Applications of Mathematics. Springer.
- KMV Corporation (1993) Credit Monitor Overview, San Francisco, California.
- KMV Corporation, 1997, Portfolio Manager Overview, An Introduction to Modelling Portfolio Risk
- Koyluoglu H.U., Hickman A., Reconcilable differences, *Risk* 11(10) 1998, pp. 56-62.
- KPMG (2003) Basel II- A closer look, Managing Economic Capital.
- Kraus, A. and Litzenberger R.H., (1976) 'Skewness Preference and the Valuation of Risk Assets', *Journal of Finance* 31, pp. 1085-1100
- Ledoit O., (1995) Essays on Risk and Return in the Stock Market, MIT Sloan School of Management, PhD Dissertation.
- Leland H. (1972) Existence and diversification of optimal portfolio policies. *Journal of Economic Theory*, (4), pp.35-44
- Litterman R. and T. Iben (1991), 'Corporate Bond Valuation and the Term Structure of Credit Spreads, *Journal of Portfolio Management*, Spring, pp. 52-64.

Lo A. W., (2002) The statistics of Sharpe Ratios, *Financial Analysts Journal*, pp. 36-52.

Loistl O., (1976) 'The Erroneous Approximation of Expected Utility by Means of a Taylor's Series Expansion: Analytic and Computational Results', *American Economic Review* 66, pp. 904-910.

Longstaff F.A., Schwartz E.S., (1995) A simple approach to valuing risky fixed and floating rate debt, *Journal of Finance* 50, pp. 789-819.

Lucas A. and Dert C.L., (1998), On the Inefficiency of Portfolio Insurance and Caveats to the Mean/Downside-Risk Framework, Technical report, Department of Finance & Financial Sector Management, Vrije Universiteit Amsterdam.

Lund J. R. (2003) Probabilistic Design and Optimization working paper, University of California, Davis.

MacMinn R. D., (1987) Notes on Stochastic Dominance, working paper.

Mao J.C., (1970) Models of capital budgeting, e-v vs e-s. *Journal of Financial and Quantitative Analysis*, pp. 657-75

Markowitz, H. (1952): Portfolio Selection, *Journal of Finance*, 7, pp. 77–91.

Markowitz, H. (1959): Portfolio Selection. John Wiley & Sons, New York.

Markowitz H. (1993) The general mean-variance portfolio selection problem. *Philosophical Transactions: Physical Sciences and Engineering*.

Marmar H.S., and F.K., Louis Ng. (1993) Mean-semivariance analysis of option-based strategies: A total asset mix perspective. *Financial Analysts Journal*.

Matten, C. (2001) *Managing Bank Capital*, 2nd edition.

Mausser H., Rosen D. Efficient risk-return frontiers for credit risk, *Algo Research Quarterly*, 2 (4) pp. 35-48.

McCarl (1996) Stochastic Dominance Notes, working paper.

Merton, R.C. (1973), An Intertemporal Capital Asset Pricing Model, *Econometrica*, 41, pp. 867–887.

Merton R.C., (1974) On the pricing of corporate debt: the risk structure of interest rate, *Journal of finance*, 27(3), pp. 449-470.

Merton, R. C. (1987), A Simple Model of Capital Market Equilibrium with Incomplete Information, *The Journal of Finance*, 42, 483–510.

Merton R.C., (1990) *Continuous-time finance*. B. Blackwell.

Merton, R. and Perold, A. (1993) Theory of risk capital in financial firms, *Journal of Applied Corporate Finance*, 6 pp. 16-32.

Millbourn T. (1997) EVA's charm as a performance measure, *Financial Times*.

Milne A. and Onorato M. (2004), When does RAROC measure shareholder value? Theoretical diagnosis and practical prescription, working paper, Cass Business School, City University

Modigliani F. and Miller, H. (1958) The cost of capital, corporation finance and the theory of investment, *American Economic Review* 48, pp. 261-297.

Modigliani F. and Modigliani, L. (1997), 'Risk-Adjusted Performance', *Journal of Portfolio Management* 23, pp. 45-54.

Moody's Investors Service (1995) *Moody's credit ratings and research*, New York.

Mussman, M. (1996) In praise of RAROC, *Balance sheet*.

Myers C. S., *The Search for Optimal Capital Structure*, MIT.

Neilsen, I., J. Saa-Requejo, and P. Santa-Clara (1993) Default Risk and Interest Rate Risk: The Term Structure of Default Spreads, working paper, INSEAD, Fontainebleau, France.

Neumann Von and Morgenstern (1947) *Expected utility theory*.

Ong M. K, (1999) Internal credit risk model , Capital allocation and Performance Measurement, Ed. Risk books.

Onorato M. Altman E. An integrated pricing model for defaultable bonds and loans, European Journal of Operational Research, Volume 163, Issue 16 May 2005, pp. 65-82

Onorato M. (1999) La valutazione degli strumenti derivati, ETASLIBRI.

Onorato, M. (2004) How to use risk adjusted performance measurement to efficiently allocate economic capital, Working Paper.

Pearson N. D., Risk budgeting, Portfolio problem solving with value at risk , Jon Wiley & Sons Inc 2002.

Pedersen, C.S., and Satchell, S.E., (1999) 'Choosing the Right Measure of Risk: A Survey', in: S.B. Dahiya, ed., The Current State of Economic Science, Vol. 2 (46) of Micro Economics, Macro Economics, Monetary Economics, Spellbound Publ.

Pelsser A. (2004) On the Applicability of the Wang Transform for Pricing Financial Risks, working paper, Erasmus University Rotterdam, www.few.eur.nl/few/people/pelsser.

Peter A, (2003) Risk Measures, Working paper, University of Mannheim.

Pflug G.P., (1996) Optimisation of Stochastic Models. Kluwer Academic Publishers, Boston.

Plantinga, A. and J. de Groot S , (2001), Utility theory and value functions, In Managing downside risk in financial markets, Eds. S. Satchel and F. Sortino, pp. 169-193.

Plantinga A. and De Groot S. (2001) Risk Adjusted Performance Measures and Implied Risk Attitudes, University of Groningen.

Pliska S. (1997) Introduction to Mathematical Finance. Blackwell Publishers.

Pliska S. (1997) Risk sensitive portfolio management with underlying stochastic factors. In H. Frydman and P. Lakner, editors, Workshop on Statistical Aspects of Financial Models, New York, NYU.

Pratt, John W., 1964, Risk aversion in the small and the large, *Econometrica* Vol. 32, pp. 122-136.

Pyle D.H., and S.J., Turnovsky (1970) Safety-first and expected utility maximization in mean standard deviation portfolio analysis. *The Review of Economics and Statistics*, LII(1), February.

Reback, R. (1975), 'Risk and Return in CBOE and AMEX Option Trading', *Financial Analyst Journal*. 31, pp. 42-52.

Reichelstein S. (1997) Investment decisions and managerial performance evaluation, *Review of Accounting Studies*, 2 pp.157-180.

Reiss R.D., (1989) *Approximate distribution of order statistics*. Springer Series in Statistics. Springer-Verlag.

Riskmetrics 1995- Technical Document, JP Morgan.

Rockafellar R., (1970) *Convex Analysis*. Princeton University Press.

Rodriguez L., (1999) *Portfolio optimisation with quantile based risk measure*, MIT Ph.D dissertation.

Rosen D., (2002) *Enterprise Credit Risk Management*, Algorithmics publication.

Ross S. (1976) *Arbitrage Pricing Theory*, *Journal of Economic Theory*

Rossi G.A., and Tibiletti L., (1996) *Higher Order Polynomial Utility Functions: Advantages in Their Use*, Technical report, Department of Mathematics and Statistics, University of Torino.

Roy (1952) *Safety first and the holding of risky asset*, *Econometrica* 20, pp. 431-449

Rubinstein, M.E. (1973), 'The Fundamental Theorem of Parameter-Preference Security Valuation', *Journal of Financial and Quantitative Analysis* 8, pp. 61-69.

Rudin W. *Principles of Mathematical Analysis*. McGraw-Hill, third edition, 1976.

Salazar M.L., (2004) How useful is VaR for Asset Allocation? Cass Business School, City University, Dissertation Thesis.

Samuel C. Weaver, J. Fred Weston , A unifying theory of value based management, Working Paper.

Samuelson (1970) P. The fundamental approximation theorem on portfolio analysis in terms of means, variances and higher moments. *Review of Economic Studies*, 37, pp.537-542, October.

Saunders A. (2002), *Credit Risk Measurement Value-at-Risk and other paradigms*, Wiley

Schonbucher, P.J. (1998) Term-Structure Modelling of Defaultable Bonds, *Review of Derivatives Research* 2, pp. 161-92.

Schroeck, G. (2000) *Risk Management and Value Creation in financial institutions*, J. Wiley.

Scott R. and P. Horvath. (1980) On the direction of preference for moments of higher order than variance. *Journal of Finance*, 35, pp.915-919.

Sharpe, W.F. (1964), *Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*, *Journal of Finance*, 19, 425-442.

Sharpe W.F., (1966) *Mutual Fund Performance*, *Journal of Business*, Jan, pp. 119-138.

Sharpe, W. F. (1992) *Asset allocation: Management style and performance measurement*. *Journal of Portfolio Management*.

Sharpe W.F., (1994) *The Sharpe Ratio*, *Journal of Portfolio Management*, Fall, pp. 49-58.

Shimko D., N. Tejima, and D. Van-Deventer (1993) *The Pricing of Risky Debt when Interest Rates are Stochastic*, *The Journal of Fixed-Income* 3, pp. 58-65.

Shawn T. (1993) *The Real Key to creating wealth*, *Fortune*.

Sinan A. Cebenoyan Philip and E. Strahan (2001) *Risk Management, Capital Structure and Lending at Banks*, the Wharton School, University of Pennsylvania.

Skora R. (1998) Credit Modelling and Credit Derivatives: Rational Modelling, working paper, Skora and Co, Inc.

Society of Actuaries Risk, Management Task Force, Economic Capital (EC) Subgroup (2003) Specialty Guide on Economic Capital, Version 1.4.

Sortino, Frank A. (1999), The Price of Astuteness Pensions & Investments.

Sortino, F.A., G. Miller and J.Messina (1997) Short Term Risk-adjusted Performance: A Style Based Analysis. Journal of Investing.

Sortino, F. and R.A.H. van der Meer. Downside Risk. Journal of Portfolio Management, Summer 1991.

Sortino, F. and van der Meer R. and Plantinga A., 1999, The Dutch triangle, Journal of Portfolio Management, Vol. 18, summer, pp. 27-31.

Sortino F, Wilkinson M., Forsey H. (2001), The Upside Potential Ratio A Framework To Measure Upside Potential Relative to Downside risk

Sortino, F. A., (2001) From Alpha to Omega, in Managing Downside Risk in Financial Markets, Frank A. Sortino and Stephen E. Satchell, eds., Reed Educational and Professional Publishing Ltd.,

Stern J. M. (2000) EVA and Strategic Performance Measurement, Global Finance 2000, New York.

Stoughton M. N. and Zechner J. (1999) Optimal Capital Allocation using RAROC and EVA, Working paper.

Stoughton, N. M. and Zechner J. (2000), 'The Dynamics of Capital Allocation', SSRN, abstract number 249987

Stulz R. (1996) Rethinking risk management, Journal of Applied Corporate Finance, vol. 9 number 3, pp. 8-24

Stultz R. (1998) Derivatives, risk management and financial engineering, Southwestern Publishing.

Tasche D., (1999) Risk Contributions and Performance Measurement, working paper Zentrum Mathematik, Munich University of Technology (www.gloriamundi.org).

Telser, L., 1955. Safety-first and hedging, *Review of Economic Studies* 23, pp.1-16.

Thomson James B., (2001) PSAF, Economic Capital and the New Basel Accord, Federal Reserve Bank of Cleveland.

Treynor J., How to rate management of investment funds. *Harvard Business Review*, pp. 63-75

Tversky, A. and Kahneman D. (1979), Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 47, 263–291.

Tversky, A. and Kahneman D. (1992), Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, 5, 297–323.

Uyemura D., Kantor C., and Pettit J., (1996) EVA for Banks: Value creation, risk management and profitability measurement, *Journal of Applied Corporate Finance*.

Vasicek O., (1987) Probability of loss on loan Portfolio, KMV Corporation.

Villani E. (2004) Optimal Bank Capital Structure, Diversification and Managerial Incentive Schemes, Working Paper, Universita di Brescia.

Von Neumann, John and Oskar Morgenstern(1944), *Theory of games and economic behavior*, Princeton, NJ: Princeton University Press.

Wang, S. (1995), Insurance pricing and increased limits ratemaking by proportional hazard transforms. *Insurance: Mathematics and Economics* 17; pp. 43-54

Wang S., Young V.R., and H.H., Panjer (1997) A risk measure that goes beyond coherence. Working paper <http://www.gloriamundi.org/picsresources/sw.pdf>.

Wang, s. and Young V.R., (1998) Ordering risks: expected utility theory versus Yaari's dual theory of risk. *Insurance Math. Econom.* 22(2), 145-161.

Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67(1), 15-36.

Wang S., (2001) Equilibrium pricing transforms : new results using Bühlmann's 1980 model, Working paper.

Wilson T., (1992) RAROC remodeled, *Risk Magazine*, Vol. 5 number 8, September 1992, pp. 112-119.

Wilson T. (1996) Calculating risk capital, in Carol Alexander (ed) *The Handbook of Risk Management and Analysis: Vol. I*, Chichester: John Wiley & Sons, Inc, pp. 193-232.

Wilson T. (1997) Portfolio Credit Risk (I), *Risk* 10, No 9, September.

Wilson T. (1997) Portfolio Credit Risk (II), *Risk* 10, No 10 ,October.

Wilson T.C., (2002) Cost of capital for financial institutions, Working paper, Mercer Oliver Wyman.

Wilson, T.C., (2002) *Overcoming the Hurdle: Integrating the Shareholder/ Debtholder-Perspectives*, Mercer Oliver Wyman, New York.

Wilson, T.C. (2003) Valuation, earnings volatility and cost of capital, Working paper, Mercer Oliver Wyman.

Wirch J.L., and M.R., Hardy (1999), Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics* 21, pp. 173-183.

Wirch J.L., and M.R., Hardy (1999), Proper Ordering for Risk Measures, working paper.

Wirch J.L., and M.R., Hardy (2000), Distortion Risk Measures: Coherence and Stochastic Dominance, working paper.

Wirch J.L., and M.R., Hardy (2000), Ordering of Risk Measures for Capital Adequacy, working paper.

Wirch J.L.,and M. R., Hardy (2003) Distortion Risk Measures Coherence and Stochastic Dominance, working paper.

Winton Capital Management (2003) Assessing CTA Quality with the Omega Performance Measure, Available at www.WintonCapital.com/feature.htm

Whitmore, G. A., 1970, "Third-degree Stochastic Dominance", American Economic Review 60:3 (June), 457-459.

Wu. L. (1998), Investing in the Emerging Markets: Effects of Higher Moments and Predictability, Technical report, Stern School of Business, New York University.

Wu. L., (1998) Investing in the Emerging Markets: Effects of Higher Moments and Predictability, Technical report, Stern School of Business, New York University.

Yaari (1987), Yaari's dual theory of choice under risk.

Yaari, M.E. (1987) The dual theory of choice under risk. *Econometrica* 55(1), 95-115.

Yoshihara T., and Y., Yamai (2001) A synthesis of risk measures for capital adequacy; *Insurance : Mathematics and Economics* 25, pp. 337-47.

Yu. P.-L (1985) *Multiple-Criteria Decision Making*. Plenum Press.

Zagst R., Kerbhaum J., and Schmid B., (2002) Portfolio optimisation under credit risk, unpublished working paper.

Zagst R., Kehrbaum J., and Schmid B., (2002) Portfolio Optimization Under Credit Risk, Risk Lab Germany, working paper.

Zaik E., Walter J., Kellig, G., and James C., (1996) RAROC at bank of America: from theory to practice, *Journal of Applied corporate finance*, 9(2), pp. 83-92 .

Ziemba W.T., (1975) and R.G. Vickson. Stochastic Optimisation Models in Finance. Academic Press.

Zsolt Berényi (2001) Accounting for illiquidity and non-normality of returns in the performance assessment, working paper, University of Munich.