Credit Portfolio Losses

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties, in het openbaar te verdedigen
op maandag 12 oktober 2009 om 12:30 uur

door

Xinzheng HUANG
wiskundig ingenieur

geboren te Wuxi, PR China

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. C. W. Oosterlee

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.ir. C. W. Oosterlee,	Technische Universiteit Delft, promotor
Prof.dr.ir. A.W. Heemink,	Technische Universiteit Delft
Prof.dr.ir. G. Jongbloed,	Technische Universiteit Delft
Prof.dr. A. Bagchi,	Universiteit Twente
Prof.dr. L. Overbeck,	Justus-Liebig-Universität Gießen, Germany
Prof.dr. W. Schoutens,	Katholieke Universiteit Leuven, Belgium
Dr. J.A.M. van der Weide,	Technische Universiteit Delft

Credit Portfolio Losses.
Dissertation at Delft University of Technology.

ISBN 978-90-9024609-3

Acknowledgments

This dissertation concludes nearly four year's of my PH.D study at the Delft University of Technology. It is real pleasure to acknowledge those who have contributed to this dissertation in one way or another.

My most sincere gratitude goes to my supervisor prof. Kees Oosterlee. I have benefited greatly from his guidance and numerous discussion we have had. I am also thankful to him for proofreading the entire manuscript thoroughly and giving valuable criticisms and suggestions. It has been an honor and great pleasure to work with Kees.

I am grateful to Mâcé Mesters for initiating a research project that combines perfectly the academic research and industry practice. He also gave me all the freedom to work on the issues I find interesting.

I would also like to thank Dr. Hans van der Weide who introduced me to the world of financial mathematics. I learned a lot from his lectures and instruction.

I am indebted to Peter Sonneveld and Jos van Kan for their significant contributions to the chapter on nonnegative matrix factorization.

Thanks also go to the members of my doctoral examination committee for putting their efforts on this thesis and providing helpful comments and suggestions.

I would like to express my gratitude to all colleagues in the Numerical Analysis group of TU Delft and at Research division of Group Risk Management, Rabobank and also former colleagues at Modelling and Research, Rabobank International, for the pleasant working atmosphere and all scientific and social discussions and activities. In particular I thank Manicka Pijnenburg and Martin van Jole for translating the summary of this thesis into Dutch. The data used to produce the figure on the front cover are kindly provided by Fang Fang.

Finally, my greatest debts are to my family - to my parents and my wife Ting, - for their love and support over the years.

Summary

Credit Portfolio Losses

Xinzheng Huang

Credit risk is most simply defined as the risk of loss resulting from an obligor's inability to meet its obligations. Generally speaking, credit risk is the largest source of risk faced by banking institutions world-wide.

On an obligor level, the three basic components of credit risk are

- exposure at default (EAD), the amount to which the bank was exposed to the obligor at the time of default,
- loss given default (LGD), the proportion of the exposure that will be lost if a default occurs,
- probability of default (PD) within a fixed time horizon.

EAD is usually assumed to be deterministic. The LGD is random but normally replaced by its expectation for simplification. Then the uncertainty in the loss due to an obligor comes solely from the outcome whether he survives or defaults.

A major principle of sound management of credit risk is to quantify credit risk on a portfolio level. Financial institutions need to determine whether they hold adequate capital against possible extreme losses and whether they are adequately compensated for risk incurred. These are often measured by Value at Risk (VaR) and Value at Risk contributions (VaRC), which are the quantile of the portfolio loss distribution at a given confidence level and the sensitivity of the VaR to an infinitesimal fractional change in EAD for each obligor, respectively. Aggregation of credit risk from individual obligors to a portfolio level involves specification of the dependence among obligors. Common practice is to utilize the factor models, in which

the obligors are independent conditional on some common factors, e.g., state of the economy, different industries and geographical regions.

Our starting point is the Vasicek model, which is the basis of the Basel II (Basel Committee on Bank Supervision 2005) internal rating-based (IRB) approach. It is a Gaussian one-factor model such that the default events are driven by a single common factor that is assumed to follow the Gaussian distribution. Under certain homogeneity conditions, the Vasicek one-factor model leads to very simple analytic asymptotic approximations of the loss distribution, VaR and VaRC. This asymptotic approximation works extremely well if the portfolio consists of a large number of small exposures. The model may be extended to portfolios that are not homogeneous in terms of default probability and pairwise correlation. However, the analytic approximation of the Vasicek model can significantly underestimate risks in the presence of exposure concentrations, i.e., when the portfolio is dominated by a few obligors.

We show that the saddlepoint approximation is an efficient tool to estimate the portfolio credit loss distribution in the Vasicek model and it is able to handle well exposure concentration. The saddlepoint approximation can be traced back to Esscher (1932) when he considered the total claim amount in a risk process and has been recognized as a valuable tool in asymptotic analysis and statistical computing.

We further compare various numerical methods for the estimation of the VaR and the marginal VaRC in the Vasicek one-factor portfolio credit loss model. The methods we investigate are the normal approximation, the saddlepoint approximation, a simplified saddlepoint approximation and importance sampling. We investigate each method in terms of speed, accuracy and robustness and in particular explore their abilities of dealing with exposure concentration.

Then we go beyond the Vasicek model to consider two extensions.

In the Vasicek model the loss given default (LGD) is assumed to be constant. However, extensive empirical evidence shows that LGD is high when the default rate is high. To account this, we propose a new framework for modeling systematic risk in LGD. The class of models is very flexible and accommodates well skewness and heteroscedastic errors. The quantities in the models have simple economic interpretation. Inference of models in this framework can be unified. Moreover, it allows efficient numerical procedures, such as the normal approximation and the saddlepoint approximation, to calculate the portfolio loss distribution and VaR.

The single factor in the Vasicek model represents generally the state of economy. More factors are necessary if one wishes to take the effects of

different industries and geographical regions into account in credit portfolio loss modeling. Therefore we move from the one-factor model to multi-factor credit portfolio loss models. We propose algorithms of adaptive integration for the calculation of the tail probability, with either a deterministic multiple integration rule or a Monte Carlo type random rule. Both algorithms are asymptotically convergent and consistently outperform the plain Monte Carlo method. The adaptive Monte Carlo integration algorithm is able to provide reliable probabilistic error bounds.

To be able to take advantage of the adaptive integration algorithm, an assumption of nonnegative coefficients in the multi-factor model is made. To make sure that this assumption is satisfied, we propose a dedicated algorithm for the nonnegative factorization of a correlation matrix.

Rapidly growing, at least before the current credit crisis, is the practice of financial institutions transferring portfolio credit risk to third parties for capital relief. Collateralized Debt Obligations (CDOs) are probably the most popular credit instruments for this purpose. The valuation of CDOs basically reduces to calculation of the expectation $\mathbb{E}[(X-K)^+]$, where X is the sum of n independent random variables and K is a known constant. We finally derive two types of saddlepoint approximations to this expectation and establish error convergence rates of the approximations in the *i.i.d.* case. The approximations are further extended to cover the case of lattice variables.

Samenvatting

Verlies in Kredietportefeuilles

Xinzheng Huang

Kredietrisico kan het best gedefinieerd worden als het risico op verliezen als gevolg van het onvermogen van een tegenpartij om aan zijn betalingsverplichtingen te voldoen. In het algemeen is kredietrisico de grootste bron van risico die bancaire instellingen wereldwijd ondervinden.

Op het niveau van de individuele tegenpartij zijn er drie basiscomponenten van kredietrisico te onderscheiden:

- Vordering in geval van wanbetaling (EAD), het bedrag dat de bank had uitstaan bij de tegenpartij op het moment van wanbetaling,
- Verlies in geval van wanbetaling (LGD), de fractie van de vordering in geval van wanbetaling dat verloren gaat wanneer een wanbetaling optreedt,
- Kans op wanbetaling (PD) binnen een bepaalde tijdshorizon.

Het wordt in het algemeen aangenomen dat de EAD deterministisch is. De LGD is stochastisch, maar het is gebruikelijk dat deze eenvoudigweg wordt vervangen door zijn verwachting. De onzekerheid van het kredietverlies wordt dan volledig bepaald door het wel of niet betalen door de tegenpartij.

Een belangrijk principe van degelijk kredietrisicomanagement is het kwantificeren van kredietrisico op portefeuilleniveau. Financiële instellingen dienen te bepalen of zij voldoende kapitaal aanhouden voor mogelijke extreme kredietverliezen en of zij voldoende gecompenseerd worden voor de genomen risico's. Deze worden meestal bepaald aan de hand van Value at Risk (VaR) en Value at Risk contributies (VaRC), welke respectievelijk zijn het

kwantiel van de verliesverdeling in een portefeuille bij een gegeven betrouwbaarheidsniveau en de gevoeligheid van de VaR voor een oneindig kleine verandering in de EAD van elke individuele tegenpartij. Aggregatie van kredietrisico van individuele tegenpartijen naar portefeuilleniveau betekent dat de afhankelijkheid tussen de tegenpartijen gespecificeerd dient te worden. In het algemeen worden hiervoor factor modellen gebruikt, waarbij de tegenpartijen onafhankelijk zijn, gegeven een aantal gemeenschappelijke factoren, bijvoorbeeld de toestand van de economie, verschillende bedrijfstakken en geografische regio's.

Ons uitgangspunt is het Vasicek model, welke de basis vormt voor de Interne Rating Benadering (IRB) van het Basel II Akkoord Basel Committee on Bank Supervision (2005). Dit is een model waarin de wanbetalingen worden gemodelleerd door één normaal verdeelde gemeenschappelijke factor. Onder bepaalde homogeniteitsvoorwaarden geeft het één-factor model van Vasicek een zeer eenvoudige analytische asymptotische benadering van de verliesverdeling, VaR en VaRC. Deze asymptotische benadering werkt bijzonder goed als de portefeuille bestaat uit een groot aantal kleine vorderingen. Het model kan worden uitgebreid naar portefeuilles welke niet homogeen zijn in de kans op wanbetaling en paarsgewijze correlatie. Echter, de analytische benadering van het Vasicek model kan het risico substantieel onderschatten wanneer er concentraties zijn in de vorderingen. Dat wil zeggen wanneer de portefeuille gedomineerd wordt door een klein aantal tegenpartijen.

We tonen aan dat de zadelpuntmethode een efficiënte methode is om de verdeling van de kredietverliezen op portefeuilleniveau te schatten in het Vasicek model en dat het ook concentraties goed kan benaderen. De zadelpuntmethode vindt zijn oorsprong bij Esscher (1932) die het totale claimbedrag onderzocht voor een risicoproces. De zadelpuntmethode wordt gezien als een waardevol instrument voor asymptotische analyse en statistische berekeningen.

Vervolgens vergelijken we verschillende numerieke methoden voor het schatten van de VaR en de marginale VaRC in het één-factor model van Vasicek voor verliezen in kredietportefeuilles. De methoden die we onderzoeken zijn de normale benadering, de zadelpuntmethode, een vereenvoudigde zadelpuntmethode en importance sampling. We bestuderen elke methode op snelheid, precisie en robuustheid en in het bijzonder bestuderen we elke methode op het vermogen om tegenpartijconcentraties goed te benaderen.

Daarna onderzoeken we twee uitbreidingen op het Vasicek model.

In het Vasicek model wordt de LGD als constant verondersteld. Echter, uitgebreid empirisch bewijs toont aan dat de LGD hoog is wanneer het wan-

betalingspercentage hoog is. Om dit systematisch risico te meten stellen we een nieuwe opzet voor om LGD te modelleren. Dit type modellen is erg flexibel en omvat de scheefheid en heteroskedasticiteit van de verdeling van de fouten. De grootheden in de modellen hebben een eenvoudige economische interpretatie. Statistische afleiding en inferentie van verschillende modellen kunnen in deze opzet worden verenigd. Daarnaast staat deze opzet het gebruik van efficiënte numerieke procedures toe voor het berekenen van de verliesverdeling in de portefeuilles en de VaR, zoals de normale benadering en de zadelpuntmethode.

De factor in het Vasicek model stelt in het algemeen de stand van de economie voor. Meerdere factoren zijn nodig wanneer rekening gehouden dient te worden met de effecten van meerdere bedrijfstakken en geografische regio's voor het modelleren van de verliesverdeling van de kredietportefeuille. Derhalve gaan we over van het één-factor model naar een meer-factorenmodel voor de verliezen in de kredietportefeuille. We bepalen adaptieve integratie algoritmes voor het berekenen van de staartkans, namelijk een deterministisch meervoudig integratie-algoritme en een Monte Carlo algoritme. Beide algoritmes convergeren asymptotisch en presteren consistent beter dan de standaard Monte Carlo methode. Het adaptieve Monte Carlo integratie-algoritme heeft betrouwbare grenzen voor de kans op fouten.

Om het adaptieve integratie-algoritme te benutten, wordt voor het meerfactoren-model aangenomen dat de coëfficiënten niet-negatief zijn. Om er zeker van te zijn dat aan deze aanname is voldaan, introduceren we een specifiek algoritme voor het niet-negatief ontbinden in factoren van een correlatiematrix.

De laatste jaren, althans voor de huidige kredietcrisis, is het zeer gebruikelijk geworden dat financiële instellingen het vereiste kapitaal verlagen door het transfereren van portefeuillekredietrisico's naar derde partijen. Collateral Debt Obligations (CDO's) zijn waarschijnlijk de meest populaire kredietinstrumenten hiervoor. De waardering van CDO's kan in principe worden gereduceerd tot het bepalen van de verwachting $\mathbb{E}[(X-K)^+]$, waarbij X de som is van n onafhankelijke stochastische variabelen en K een bekende constante. Tenslotte leiden we twee types van zadelpuntbenaderingen af voor het bepalen van deze verwachting en bepalen de snelheid van de convergentie van de benaderingsfouten in het geval van identiek en onafhankelijk verdeelde stochastische variabelen. De benaderingen worden uitgebreid tot het geval van roosterstructuur-variabelen.

Contents

Acknowledgments **iii**

Summary **iv**

Samenvatting **viii**

1 Introduction **1**
1.1 Risk measures and risk contributions 5
1.2 Factor models for credit portfolios 7
1.2.1 The Vasicek one-factor model 7
1.2.2 Multi-factor models 10
1.3 Saddlepoint approximations 12
1.4 Setup of the thesis 13

2 Higher Order Saddlepoint Approximations in the Vasicek Portfolio Credit Loss Model **17**
2.1 Introduction 17
2.2 Saddlepoint approximations in the Vasicek model: a conditional approach 18
2.3 Numerical results 21
2.4 Further discussion 26
2.4.1 Adaptive integration in a one-factor model 26
2.4.2 Multi-factor models 28
2.4.3 Non-Gaussian factor models 30
2.4.4 Random LGD 31
2.5 Conclusions 31
2.A Binomial expansion method 32

3 Computation of VaR and VaR Contribution in the Vasicek Portfolio Credit Loss Model: A Comparative Study 35
3.1 Introduction . . . 35
3.2 Numerical methods . . . 37
3.2.1 Normal approximation . . . 37
3.2.2 Saddlepoint approximation . . . 38
3.2.3 Simplified saddlepoint approximation . . . 39
3.2.4 Importance sampling . . . 39
3.3 A stylized portfolio . . . 41
3.4 Analysis of robustness . . . 44
3.5 Conclusions and discussions . . . 52

4 Generalized Beta Regression Models for Random Loss Given Default 55
4.1 Introduction . . . 55
4.2 Existing LGD models . . . 57
4.3 Generalized Beta Regression Models . . . 58
4.3.1 Parameterization of a beta distribution . . . 58
4.3.2 Beta Regression Model . . . 59
4.3.3 Extensions . . . 61
4.4 Estimation . . . 62
4.4.1 Least squares . . . 63
4.4.2 Maximum likelihood estimation . . . 64
4.4.3 A simulation study . . . 65
4.5 Loss distribution approximations . . . 71
4.5.1 Normal approximation . . . 71
4.5.2 Saddlepoint approximation . . . 72
4.5.3 Numerical results . . . 73
4.6 Conclusions . . . 74
4.A Score function and Fisher information matrix . . . 75
4.B LGD Statistics by Year . . . 77

5 Adaptive Integration for Multi-factor Portfolio Credit Loss Models 79
5.1 Introduction . . . 79
5.2 Multi-factor portfolio credit loss model . . . 80
5.3 Globally adaptive algorithms for numerical integration . . . 82
5.3.1 Globally adaptive algorithms . . . 82
5.3.2 The Genz-Malik rule . . . 84
5.3.3 A tailor-made adaptive Genz-Malik rule . . . 86

5.3.4 Numerical results 88
5.4 Adaptive Monte Carlo integration 91
5.5 Conclusions 96

6 Nonnegative Matrix Factorization of a Correlation Matrix 99
6.1 Introduction 99
6.2 An algorithm for nonnegative factorization 100
6.2.1 The non-negativity conditions 101
6.2.2 Unit norm condition 102
6.2.3 Algorithm 107
6.3 Numerical results 108
6.4 Conclusions 109
6.A Correlation matrix: example one 113

7 Saddlepoint Approximations for Expectations 115
7.1 Introduction 115
7.2 Densities and tail probabilities 116
7.3 Measure change approaches 118
7.4 Classical saddlepoint approximations 120
7.5 The Lugannani-Rice type formulas 124
7.6 Lattice variables 130
7.7 Numerical results 132
7.8 Conclusions 133
7.A Alternative derivation of the Lugannani-Rice type formulas . 135

8 Conclusions and Outlook 139
8.1 Conclusions 139
8.2 Outlook 140

Bibliography 143

Curriculum vitae 151

List of publications 153

Chapter 1

Introduction

On September 15, 2008, Lehman Brothers Holdings Inc., the fourth-largest US investment bank that dated back to 1850, announced it would file for bankruptcy. The fall of Lehman marks the biggest bankruptcy filing ever. The filing cited total debt of $613 billion, including $155 billion bond debt and assets worth $639 billion as of May 31, 2008 and said "In the judgment of the Board (of Directors), it is desirable and in the best interests of the Company (Lehman), its creditors, employees, and other interested parties that a petition be filed by the Company seeking relief under the provisions of chapter 11 of the (bankruptcy) code".

Bondholders face a loss at the bankruptcy of Lehman as Lehman is not able to fully repay its debt. The magnitude of loss can be inferred from the price of the credit default swap (CDS), an instrument used to insure against losses due to the default on bonds. A CDS is a credit derivative contract between two counterparties such that the buyer makes periodic payments to the seller, and in return receives a payoff if an underlying financial instrument (typically a loan and bond) defaults. A bond holder who enters a CDS transaction can thus transfer his credit risk to the protection seller. The auction to settle Lehman's CDS on October 10, 2008 put the value of those bonds at merely 8.625 cents on the dollar. This roughly means that a holder of bonds issued by Lehman had to take a horrific loss of 91.375 cents on the dollar by expectation.

Leading to Lehman's collapse are losses of billions of dollars through the year 2008 in the US mortgage market, in particular in subprime and other lower-rated mortgage-backed securities. Subprime mortgage is the type of mortgage loan offered to borrowers who do not meet the usual criteria for borrowing. Subprime borrowers are those who have a history of not paying loans back, those with a recorded bankruptcy, or those with a limited in-

come, thus having a reasonable chance of defaulting on the debt repayment. According to Mortgage Bankers Association's National Delinquency Survey, subprime mortgages represented 13.1% of the mortgages outstanding in the US and 55% of the foreclosures started during the third quarter of 2007.

Data from Harvard's "2008 State of the Nation's Housing" study showed a rapid expansion of subprime mortgage lending. Between 2004 - 2006 the share of subprime mortgages relative to total originations ranged from 18% - 21%, versus less than 10% in 2001-2003. A major factor that accounted for this increase was the securitization of mortgage pools. These mortgage-linked securities are known as the Collateralized Debt Obligations (CDOs). With their values and payments derived from the underlying mortgage portfolios, CDOs typically are divided into several tranches that are associated with differing levels of risk and return. Interest and principal payments are made in order of seniority. Super senior designates the highest tranche of a CDO, bearing a low interest rate, and is considered the safest set of securities. Junior tranches offer higher interest rates or lower prices to compensate for the additional default risk. For the investors, CDOs were regarded as attractive fixed income instruments generating significantly higher returns than bonds of the same quality according to credit rating agencies. In the meantime, ideally the originators of these securities are able to transfer the credit risk to investors completely while collecting substantial fees. In the pursuit of profit the focus of the mortgage lending practice shifted from loan quality to loan volume, accompanied by lax lending standards and deterioration of the credit worthiness of the mortgage portfolios.

However, the major financial institutions became both the originators and the investors of the mortgage-linked CDOs, against their will to "originate to distribute" the credit risk associated with the mortgage portfolios, thus retaining a great portion of the credit risks. This was probably because either they saw the CDOs as great investment opportunities, or they were unable to sell all the securities. With the collapse of the US subprime mortgage market, the CDOs quickly lost their values and became toxic assets with rare demand. On July 29, 2008, Merrill Lynch announced the sale of $31 billion in CDOs to Lone Star Funds for merely $6.7 billion. According to a Bloomberg report on August 12, 2008, the worldwide banks' subprime losses topped $500 billion.

Across the Atlantic Ocean, many European financial institutions were also hit hard by the subprime crisis, among which UBS AG suffered the highest write-downs, totaling above $40 billion as of August 2008. A shareholder report issued by UBS on 18 August 2008 detailed where and how the subprime losses accumulated within the bank. UBS's subprime losses con-

centrated in three distinct businesses: (1) Dillon Read Capital Management (DRCM), (2) CDO desk within investment bank's fixed income business, (3) investment bank's Foreign Exchange / Cash Collateral Trading (FX/CCT) business. The primary contributor to UBS's write-downs was the CDO desk, which represented approximately two thirds of the banks's subprime losses as of December 31, 2007. DRCM and FX/CCT contributed 16% and 10% respectively.

For most deals of securitization, the CDO desk retained the super senior tranches of the CDOs it structured on its own book as a long term investment strategy. In addition to the retained super senior tranches, the CDO desk further purchased a large volume of super senior tranches from third parties. By September 2007 the CDO desk held approximately $50 billion super senior positions, among which $20.8 billion were purchased from third parties. A large proportion of these super senior positions were hedged. However even though certain positions were only offered a first-loss protection of 2%, they were assumed to be fully hedged and so the hedged positions were assumed to carry no risk at all. Under normal market conditions, these hedge strategies worked fairly well and the CDO desk was highly profitable from 2005 to 2006. As market conditions deteriorated, the zero risk assumption proved to be incorrect. Write-downs mounted as either losses exceeded the extent of the purchased protection or the counterparty risk to the protection seller rose significantly. At December 31, 2007, write-downs on hedged positions contributed approximately 73% of the total super senior losses, whereas three quarters of the CDO desk's total losses (or 50% of the bank's total losses) came from these super senior tranches. Failing to recognize the concentration risk to which the bank was exposed, the super senior positions emerged as the greatest source of risks for UBS.

To restore confidence and trust in the financial services sector, governments of European countries poured billions into major banks. In the Netherlands, the ING Group received €10 billion capital injection from the Netherlands authorities in October 2008. The cash injection allowed the bank to boost its core Tier-1 capital ratio, a key measure of a bank's financial strength, from 6.5% to around 8%.

The capital ratio is the percentage of a bank's capital to its risk-weighted assets. Weights are defined by risk-sensitivity ratios whose calculation is coined in the relevant Basel Accord, issued by the Basel Committee on Banking Supervision (BCBS).

The 1988 Basel accord, also known as the Basel I, classified assets of banks in five categories which carry different risk weights, e.g., 0% for cash and short term government bonds, 50% for residential mortgage loans and

100% weighting on commercial loans. Internationally active banks are required to hold Tier-1 capital, consisting primarily of equity capital and cash reserves, equal to 4% and total capital (Tier-1 capital plus supplementary capital) ratio of 8% of the risk-weighted assets.

The Basel I Accord is greatly expanded in the Basel II Accord, initially published in June 2004. Basel II creates a more sensitive measurement of a bank's risk-weighted assets. Moreover, banks are allowed to use internal ratings-based (IRB) approach to evaluate the riskiness of their credit exposure. Rather than using the risk weights set out in the standardized approach, banks may rely on their own internal estimates of risk components, such as the probability of default (PD), loss given default (LGD), the exposure at default (EAD), in determining the capital requirement for a given exposure.

The aim of the Basel accords is to safeguard the banking institutions' solvency against potentially extreme losses. Besides credit risk, the Basel II Accord also addresses the measurement of operational risk and market risk. Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Market risk is the risk of loss due to changes in market prices (such as equity and commodity prices) or market rates (such as interest or exchange rates). A survey conducted by International Financial Risk Institute in 2006 showed that, on average, the percentages of reserved capital at a bank which can be attributed to market, operational and credit risks are 10%, 14% and 61% respectively. Credit risk is obviously the largest source of risk face banking institutions.

A high level of credit risk management from the perspective of a bank is more than solely meeting regulatory requirements. Rather the aim is to enhance the risk / return performance of credit assets. This can be achieved by active credit portfolio management with a number of strategies such as capital allocation, concentration reduction, risk mitigation, risk transfer, securitization, etc. This dissertation is concerned with quantifying portfolio credit risk with an eye to active credit portfolio management. We shall start with a selected introduction to credit portfolio loss including risk measures and credit portfolio risk models. We also briefly cover the saddlepoint approximation, a computational technique that we find very useful in the context of portfolio credit risk. The material given in this chapter is mostly expository.

All the random variables in this dissertation are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{P}$ being the physical measure, unless specified otherwise. Throughout this dissertation we denote by $f_X(\cdot)$ and $F_X(\cdot)$ the

probability density function ($p.d.f.$) and cumulative density function ($c.d.f.$) of random variable X, respectively. We write $i.i.d.$ for independently and identically distributed. ϕ and Φ represent the $p.d.f.$ and $c.d.f.$ of a standard normal distribution. Expectation and variance are denoted by $\mathbb{E}$ and $\mathbb{V}ar$, respectively.

1.1 Risk measures and risk contributions

Consider a credit portfolio consisting of n obligors. Any obligor i can be characterized by three quantities: the probability of default PD_i, the exposure at default EAD_i and the loss given default LGD_i. Obligor i is subject to default in a fixed time horizon and the default can be modeled as a Bernoulli random variable D_i such that

$$\mathbb{P}(D_i = 1) = 1 - \mathbb{P}(D_i = 0) = \mathrm{PD}_i.$$

EAD measures the amount to which the bank was exposed to the obligor at the time of default and LGD is the the proportion of the EAD that will be lost if a default occurs. For simplification of notation, we write the EAD and LGD of obligor i by ω_i and Λ_i respectively. Then the loss incurred by the obligor i is given by

$$L_i = \mathrm{EAD}_i \times \mathrm{LGD}_i \times D_i = \omega_i \Lambda_i D_i.$$

It follows that the portfolio loss is given by

$$L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} \omega_i \Lambda_i D_i. \tag{1.1}$$

For financial institutions, a major principle of sound management of credit risk is essentially to measure the credit risk at a portfolio level so as to determine the amount of capital that they need to hold as a cushion against extreme losses.

The portfolio risk is often measured by *Value at Risk* (VaR) in practice. From a mathematical perspective, the VaR for a given confidence level α is simply the α-quantile of the loss distribution of L. Thus,

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

Usually the α of interest is close to 1. While VaR is arguably the most popular risk measure, it is known to be not *coherent* in the sense of Artzner

et al. (1999). In particular VaR is not subadditive, i.e., the VaR of a portfolio can be larger than the sum of the VaRs of its subportfolios. This implies that VaR might discourage diversification. A coherent alternative to the VaR is the *Expected Shortfall* (ES). It is defined as the conditional expectation of the loss given that the loss exceeds the VaR,

$$\mathrm{ES}_\alpha = \mathbb{E}[L|L \geq \mathrm{VaR}_\alpha].$$

An equally important task as to quantify the portfolio level credit risk is to measure how much each obligor in a portfolio contributes to the total risk, i.e., the risk contributions of single exposures. Risk contribution plays an integral role in the determination of limits on large credit exposures, risk-sensitive loan pricing and eventually portfolio optimization.

A desirable property of the risk contributions is that they sum up to the corresponding risk measure. For example, we want the VaR contributions (VaRC) to add up to the total VaR, i.e., $\sum_{i=1}^n \mathrm{VaRC}_i = \mathrm{VaR}$. A common measure of risk contribution that satisfies this property is the sensitivity of the risk to an infinitesimal fractional change in exposure, as given in Gourieroux et al. (2000). Under some continuity conditions, the VaR contribution coincides with the conditional expectation of L_i given that the portfolio loss L takes value $\mathrm{VaR}_\alpha(L)$, i.e.,

$$\mathrm{VaRC}_{i,\alpha} = \omega_i \frac{\partial \mathrm{VaR}_\alpha}{\partial \omega_i}(L) = \mathbb{E}[L_i|L = \mathrm{VaR}_\alpha(L)], \tag{1.2}$$

The sum of the VaR contributions indeed equals the total VaR, i.e.,

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}\left[L_i|L = \mathrm{VaR}_\alpha(L)\right] &= \mathbb{E}\left[\sum_{i=1}^n L_i|L = \mathrm{VaR}_\alpha(L)\right] \\ &= \mathbb{E}[L|L = \mathrm{VaR}_\alpha(L)] \\ &= \mathrm{VaR}_\alpha(L).\end{aligned}$$

Similarly, the ES contribution (ESC) is given by

$$\omega_i \frac{\partial \mathrm{ES}_\alpha}{\partial \omega_i}(L) = \mathbb{E}[L_i|L \geq \mathrm{VaR}_\alpha(L)]. \tag{1.3}$$

We also have

$$\sum_{i=1}^n \mathbb{E}[L_i|L \geq \mathrm{VaR}_\alpha(L)] = \mathrm{ES}_\alpha(L). \tag{1.4}$$

For more discussions on risk measures and risk contributions we refer to Bluhm et al. (2002), Denault (2001), Tasche (1999), Kalkbrener (2005).

1.2 Factor models for credit portfolios

To evaluate the portfolio loss distribution, a key issue is to model the various dependence effects, including the dependence between defaults, the dependence between LGDs and the dependence between PD and LGD.

Direct modeling of the pairwise correlations is impractical since a bank's credit portfolio can easily contain tens of thousands of obligors. Common practice to reduce the computational complexity is therefore to utilize a so-called *factor model* of asset correlations. In a factor model, L_i and L_j are independent conditional on some common factors $\mathbf{\Psi}$. The factors $\mathbf{\Psi}$ can represent the state of the economy, different industries and geographical regions, etc. A broad class of models in the portfolio credit loss modeling, including all popular industrial models like KMV's Portfolio Manager (Kealhofer 2001), CreditRisk$^+$ (Credit Suisse Financial Products 1997) and CreditPortfolioView (Wilson 1997*a*,*b*), are factor models that take advantage of conditional independence. For a summary of the models see Crouhy et al. (2000).

1.2.1 The Vasicek one-factor model

We concentrate on the Vasicek one-factor Gaussian copula model in this section. Although the Vasicek model is often criticized to be oversimplistic by relying on the Gaussian distribution, the extension of this model to the generic one-factor Lévy model, as outlined in Albrecher et al. (2007), is straightforward. The Lévy models are able to produce more heavy-tailed loss distributions and provide a better fit to the financial market data.

For now we make a simplifying assumption that the LGDs are constant. Then the relevant dependence among obligors reduces to only default dependence. Without loss of generality, we set LGD$=1$ for all obligors. Equivalently it is also possible to interpret ω to be the effective exposure, which is the product of EAD and LGD.

The Vasicek model is a one period default-mode model, i.e., loss only occurs when an obligor defaults in a fixed time horizon $\mathtt{T}$. Based on the firm value model of Merton (1974), the Vasicek model evaluates the default of an obligor in terms of the evolution of its asset value. Suppose the asset value follows a geometric Brownian motion with initial value A_0, drift μ and volatility σ, so that $\mathrm{d}A_t = \mu A_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t$. The asset value at horizon $\mathtt{T}$ can be represented as

$$A_{\mathtt{T}} = A_0 \exp\left((\mu - \sigma^2/2)\mathtt{T} + \sigma\sqrt{\mathtt{T}}X\right),$$

where X is a standard normal variable and can be interpreted as the standardized asset log-return. Default occurs when $A_{\mathtt{T}} < B$ with B being the debt level, or equivalently, X is less than $c = (\log B - \log A_0 - \mu\mathtt{T} + \frac{\sigma^2}{2}\mathtt{T})/\sigma\sqrt{\mathtt{T}}$. The probability of default is thus given by $\mathrm{PD} = \mathbb{P}(X < c)$. X is decomposed into a systematic part Y, representing the state of the economy, and an idiosyncratic part ε, such that for obligor i we have

$$X_i = \sqrt{\rho}Y + \sqrt{1-\rho}\varepsilon_i, \tag{1.5}$$

where Y and all ε_i are i.i.d standard normal random variables and $\rho > 0$ is the common pairwise correlation. It is now easily deduced that X_i and X_j are conditionally independent given the realization of Y. This implies that L_i and L_j are also conditionally independent given Y. Further assumptions of the original Vasicek model are that all obligors have the same characteristics, such that $\mathrm{PD}_i = p$, $\mathrm{EAD}_i = 1$.

Denote by $p(y) = p_i(y) = \mathbb{P}[D_i = 1|Y = y]$, i.e., the probability of default conditional on the common factor $Y = y$. Then

$$p(y) = \mathbb{P}[D_i = 1|Y = y] = \mathbb{P}[X_i < c_i|Y = y] = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\rho}y}{\sqrt{1-\rho}}\right). \tag{1.6}$$

As a consequence of the strong law of large numbers, one obtains for $n \to \infty$,

$$\mathbb{P}\left[\lim_{n\to\infty} L/n = p(y)\middle|Y = y\right] = 1.$$

Equivalently, if we denote by $L(Y)$ the portfolio loss L conditional on Y, we have

$$\lim_{n\to\infty} L(Y)/n = p(Y) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\rho}Y}{\sqrt{1-\rho}}\right) \quad a.s. \tag{1.7}$$

For all pairs of x^* and y^* such that $p(y^*) = x^*$, we have

$$\begin{aligned}\mathbb{P}\left(\lim_{n\to\infty} L/n \le x^*\right) &= \int \mathbf{1}_{\{\lim_{n\to\infty} L(y)/n \le x^*\}}\phi(y)dy \\ &= \int \mathbf{1}_{\{p(y)\le x^*\}}\phi(y)dy = \int \mathbf{1}_{\{y\ge y^*\}}\phi(y)dy = 1 - \Phi(y^*).\end{aligned}$$

The second last equality holds since $p(Y)$ is strictly monotonically decreasing in Y. Therefore the α quantile of L/n corresponds exactly to the $1-\alpha$ quantile of Y, i.e.,

$$\mathrm{VaR}_\alpha = np(\Phi^{-1}(1-\alpha)) = n\Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right). \tag{1.8}$$

As all obligors in the portfolio are equivalent, the VaR contribution of each obligor is simply $\text{VaR}/n = p(\Phi^{-1}(1-\alpha))$.

We note that although the assumptions of uniform pairwise correlation ρ and unconditional default probability PD are made in Vasicek (2002), they are not necessary conditions and can be relaxed. Moreover, the convergence in (1.7) also holds for a portfolio with unequal weights ω_i if

$$\frac{\sum \omega_i^2}{(\sum \omega_i)^2} \to 0, \tag{1.9}$$

in other words, the portfolio exposure is not concentrated on a few loans much larger than the rest. The left hand side of formula (1.9) is known as the Herfindahl-Hirschman Index (HHI), see e.g. Hirschmann (1964), Gordy (2003). It provides a simple heuristic approach for quantifying exposure concentration. Well-diversified portfolios with a very large number of very small obligors have an HHI value close to $1/n$, where n is the number of obligors, whereas heavily concentrated portfolios can have a considerably higher HHI value.

Throughout this dissertation, the Vasicek model should be considered as a one-factor Gaussian copula model that allows heterogeneous portfolios rather than the restrictive original model for homogeneous portfolios. Summarizing, for a portfolio which is not homogeneous in terms of effective weight, default probability and pairwise correlation, the individual loss variable L_i conditional on Y is given by

$$L_i(Y) = \begin{cases} \omega_i & \text{with probability } \Phi\left(\frac{\Phi^{-1}(p_i)-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right), \\ 0 & \text{with probability } 1-\Phi\left(\frac{\Phi^{-1}(p_i)-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right). \end{cases}$$

If (1.9) is satisfied, the fraction of loss $\bar{L}(Y)$ is given by

$$\bar{L}(Y) = \lim_{n\to+\infty} \frac{\sum_{i=1}^n L_i(Y)}{\sum_{i=1}^n \omega_i} = \frac{\sum_{i=1}^n \omega_i \Phi\left(\frac{\Phi^{-1}(p_i)-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right)}{\sum_{i=1}^n \omega_i} \quad a.s.$$

Then, the VaR and VaR contributions are given by

$$\text{VaR}_\alpha = \sum_{i=1}^n \omega_i \Phi\left(\frac{\Phi^{-1}(p_i)+\sqrt{\rho_i}\Phi^{-1}(\alpha)}{\sqrt{1-\rho_i}}\right), \tag{1.10}$$

$$\text{VaRC}_{i,\alpha} = \omega_i \Phi\left(\frac{\Phi^{-1}(p_i)+\sqrt{\rho_i}\Phi^{-1}(\alpha)}{\sqrt{1-\rho_i}}\right). \tag{1.11}$$

Note that the VaR contribution (1.11) is a portfolio-invariant linear function of ω_i, which implies that the capital contributions of individual exposures only depend on the characteristics of the particular exposure and not on the rest of the portfolio.

The Vasicek asymptotic formula is straightforward but it strongly relies on the assumptions of an infinitely large portfolio and of no exposure concentration. When the two conditions, especially the latter, are violated, which constantly occurs in practice, it tends to underestimate risk. Therefore, the analytic formulas are less suitable when a portfolio is of small size or dominated by a few loans much larger than the rest. These scenarios are studied in great detail in this thesis.

1.2.2 Multi-factor models

The systematic factor in the Vasicek model represents generally the state of economy. More factors are necessary if one wishes to take the effects of different industries and geographical regions into account. The resulting multi-factor model offers a better solution to identifying the correlations among individual obligors.

A Gaussian multi-factor model is given as follows. For obligor i, its standardized log-asset return is written to be

$$X_i = \gamma_{i1}\Psi_1 + \cdots + \gamma_{iM}\Psi_M + \epsilon_i = \boldsymbol{\gamma}_i \boldsymbol{\Psi} + \epsilon_i. \tag{1.12}$$

Here

- $\Psi_1, \ldots, \Psi_M$ represent sector (industry, geographic region) indices that are correlated with a known correlation matrix $\boldsymbol{C}$. Since $\boldsymbol{C}$ is a correlation matrix, it has the properties of positive semi-definiteness and a unit main diagonal. For credit portfolios we are generally interested in the *worst case* scenarios and the reason for a clustering of defaults, hence for large portfolio losses, is positive correlation. Matrix $\boldsymbol{C}$ is therefore assumed to have only nonnegative entries. Positive correlation is often justified by empirical evidence.

- $\gamma_{ij} \geq 0$ for all i, j. These coefficients are products of a company's participation percentage in a specific country/industry and the percentage of volatility of the company, which can be explained by the volatility of the country/industry index. So, these coefficients are always positive, as can be read in the CreditMetrics Technical document (Gupton et al. 1997). The non-negativity of γ_{ij} guarantees that

larger values of the factors Ψ_i, ceteris paribus, lead to a smaller number of defaults. The nonnegativity of the coefficients is common also in other risk applications, see, eg., Glasserman & Li (2005).

- ϵ_i denotes an idiosyncratic factor that only affects an obligor itself.
- $\boldsymbol{\Psi}$ and ϵ_i are assumed to be independent for all i.

Note that the non-negativity of correlation matrix $\boldsymbol{C}$ is a conservative argument in the perspective of risk management as it precludes negative linear relationship between the common factors. As all the factors tend to move in the same direction, extremely adverse scenarios leading to huge losses are more likely.

Under such a latent factor model (1.12), the tail probability of the portfolio loss can be formulated as

$$\mathbb{P}(L > x) = \int \mathbb{P}\left(L > x \,|\, \boldsymbol{\psi}\right) \mathrm{d}F_{\boldsymbol{\Psi}}(\boldsymbol{\psi}), \tag{1.13}$$

where $F_{\boldsymbol{\Psi}}(\cdot)$ denotes the joint *c.d.f.* of $\boldsymbol{\Psi}$.

The number of industrial and country indices in (1.12) can be quite large. KMV's correlation model, for example, according to Zeng & Zhang (2001), identifies "more than 40 countries and 61 industries", whereas CreditMetrics covers "152 country-industry indices, 28 country indices, 19 worldwide industry indices, and 6 regional indices". It is evident that the integral to be solved in (1.13) can be truly a high dimensional problem. Monte Carlo (MC) simulation and quasi-Monte Carlo (QMC) methods, which do not suffer from the curse of dimensionality, are the prevailing methods used to solve these multi-dimensional integration problems in finance. However, the event $\{L > x\}$ in (1.13) becomes a rare event for high loss levels x, that are often the most interesting ones in practice. In this regard both MC and QMC methods can be rather inefficient. Furthermore, the indices Ψ_i and Ψ_j are correlated, which also leads to more complexity for the numerical integration. A version of factor model with *orthogonal* indices is therefore preferred, which means that we are led to a *second level factor model* as follows,

$$\Psi_i = a_{i1}Y_1 + \cdots + a_{im}Y_m + \delta_i = \boldsymbol{a}_i\mathbf{Y} + \delta_i, \tag{1.14}$$

where Y_i and Y_j are independent and $m \leq M$ (preferably $m \ll M$). Such a decomposition is usually achieved by a principal components analysis (PCA). The two-level factor model, combining (1.12) and (1.14), reads

$$\mathbf{X} = \boldsymbol{\Gamma}\boldsymbol{\Psi} + \boldsymbol{\epsilon} = \boldsymbol{\Gamma}\mathbf{A}(\mathbf{Y} + \boldsymbol{\delta}) + \boldsymbol{\epsilon} = \boldsymbol{\Gamma}\mathbf{A}\mathbf{Y} + \mathbf{b}\boldsymbol{\varepsilon}. \tag{1.15}$$

1.3 Saddlepoint approximations

Dating back to Esscher (1932), the saddlepoint approximation has been recognized as a valuable tool in asymptotic analysis and statistical computing. It has found a wide range of applications in finance and insurance, reliability theory, physics and biology.

Let $X_i, i = 1 \dots n$ be n independent continuous random variables and $X = \sum_{i=1}^{n} X_i$. Suppose that for all i, the *moment generating function* (MGF) of X_i is analytic and given by $\mathcal{M}_{X_i}(t) = \mathbb{E}(e^{tX_i})$. The MGF of the sum X is then simply the product of the MGF of X_i, i.e.,

$$\mathcal{M}(t) = \prod_{i=1}^{n} \mathcal{M}_{X_i}(t),$$

for t in some open neighborhood of zero. Let $\mathcal{K}(t) = \log \mathcal{M}(t)$ be the Cumulant Generating Function(CGF) of X. The density of X can be represented by the Bromwich integral

$$f_X(x) = \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} \exp(\mathcal{K}(t) - tx) \mathrm{d}t, \tag{1.16}$$

with $\mathtt{i} = \sqrt{-1}$.

The tail probability of X is given by $\mathbb{P}(X \geq x) = \int_{\{h \geq x\}} f_X(h) \mathrm{d}h$. Replace $f_X(\cdot)$ by (1.16) and change the order of integration, we are led to the following inversion formula for the tail probability,

$$\mathbb{P}(X \geq x) = \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} \frac{\exp(\mathcal{K}(t) - tx)}{t} \mathrm{d}t \quad (\tau > 0). \tag{1.17}$$

Saddlepoint approximation arises in this setting to give an accurate analytic approximation to densities and tail probabilities. A detailed exposition of saddlepoint approximations can be found in Jensen (1995). The saddlepoint approximation can be thought of as the Edgeworth expansion at the center of an Esscher transformed density. Only using the approximation at the center of the distribution, the saddlepoint approximation usually leads to a small relative error.

The saddle point, i.e., the point at which $\mathcal{K}_X(t) - tx$ is stationary, is a $t = T$ such that

$$\mathcal{K}'_X(T) = x. \tag{1.18}$$

The density $f_X(x)$ and the tail probability $\mathbb{P}(X > x)$ can be approximated by $\mathcal{K}_X(t)$ and its derivative up to second order at T.

The Taylor expansion of $\mathcal{K}(t) - tx$ (function of t) around T gives

$$\mathcal{K}(t) - tx = \mathcal{K}(T) - Tx + \frac{1}{2}(t-T)^2\mathcal{K}''(T) + ... \tag{1.19}$$

Substitute (1.19) into (1.16), and change the integration contour from the imaginary axis along τ to that along T, we get

$$\begin{aligned} f_X(x) &\approx \frac{\exp(\mathcal{K}(T) - Tx)}{2\pi\mathtt{i}} \int_{T-\mathtt{i}\infty}^{T+\mathtt{i}\infty} \exp\left(\frac{1}{2}(t-T)^2\mathcal{K}''(T)\right) \mathrm{d}t \\ &= \frac{\exp(\mathcal{K}(T) - Tx)}{\sqrt{2\pi\mathcal{K}''(T)}}. \end{aligned} \tag{1.20}$$

The tail probability is approximated as

$$\mathbb{P}(X \geq x) \approx \begin{cases} \exp\left(-\frac{W^2}{2} + \frac{Z^2}{2}\right)[1 - \Phi(Z)] & x > \mathbb{E}(X), \\ \frac{1}{2} & x = \mathbb{E}(X), \\ 1 - \exp\left(-\frac{W^2}{2} + \frac{Z^2}{2}\right)[1 - \Phi(Z)] & x < \mathbb{E}(X), \end{cases} \tag{1.21}$$

where $Z := T\sqrt{\mathcal{K}''(T)}$ and $W := sgn(T)\sqrt{2[xT - \mathcal{K}(T)]}$ with $sgn(T)$ being the sign of T.

If all the X_i are identically distributed, the relative errors of both approximations in (1.20) and (1.21) are known to be $O(n^{-1})$ and $O(n^{-1/2})$ respectively. Higher order approximations of the density and the tail probability are given by the Daniels (Daniels 1987) formula

$$f_X(x) = \frac{\phi(Z)}{\sqrt{\mathcal{K}''(T)}} \left\{ \left[1 + \left(-\frac{5\kappa_3^2}{24} + \frac{\kappa_4}{8} \right) \right] + O\left(n^{-2}\right) \right\}, \tag{1.22}$$

with $\kappa_r := \mathcal{K}^{(r)}(T)/\mathcal{K}''(T)^{r/2}$ the standardized cumulant of order r evaluated at T and the Lugannani-Rice (Lugannani & Rice 1980) formula

$$\mathbb{P}(X \geq x) = 1 - \Phi(W) + \phi(W)\left[\frac{1}{Z} - \frac{1}{W} + O\left(n^{-3/2}\right)\right]. \tag{1.23}$$

In the context of portfolio credit risk, the above saddlepoint approximations to the density and tail probability can be employed to calculate the risk measures (VaR and ES) and risk contributions as well.

1.4 Setup of the thesis

The rest of the dissertation is organized as follows.

Chapter 2 utilizes the saddlepoint approximation as an efficient tool to estimate the portfolio credit loss distribution in the Vasicek model. Value at Risk can then be found by inverting the loss distribution. VaR contribution, Expected Shortfall (ES) and ES contribution (ESC) can all be calculated accurately. Exposure concentration, for which the Vasicek asymptotic formula fails, can also be handled well by the saddlepoint approximation. We further propose an adaptive integration algorithm to accelerate the computation of tail probability. Finally we point out that the saddlepoint approximation technique can be readily applied in multi-factor models and models with non-Gaussian factors.

In Chapter 3 we compare various numerical methods for the estimation of the VaR and the marginal VaRC in the Vasicek one-factor portfolio credit loss model. The methods we investigate are the normal approximation, the saddlepoint approximation, a simplified saddlepoint approximation and importance sampling. We investigate each method in terms of speed, accuracy and robustness and in particular explore its abilities of dealing with exposure concentration.

In the Vasicek model the loss given default (LGD) is assumed to be constant. However, extensive empirical evidence shows that LGD is high when the default rate is high. To account this, in Chapter 4 we propose a new framework for modeling the systematic risk in LGD. The class of models proposed is very flexible and accommodates well skewness and heteroscedastic errors. The quantities in the models have simple economic interpretation. Inference of models in this framework can be unified. Moreover, it allows efficient numerical procedures, such as the normal approximation and the saddlepoint approximation, to calculate the portfolio loss distribution, VaR and ES.

The systematic factor in the Vasicek model represents generally the state of economy. More factors are necessary if one wishes to take the effects of different industries and geographical regions into account. In Chapter 5 we move from the one-factor model to the multi-factor credit portfolio loss models. We propose algorithms of adaptive integration for the calculation of the tail probability. We first modify the classical Genz-Malik rule, a deterministic multiple integration rule suitable for portfolio credit models with number of factors less than approximately 8. Later on we arrive at the adaptive Monte Carlo integration, which essentially replaces the deterministic integration rule by antithetic random numbers. The latter can not only handle higher-dimensional models but is also able to provide reliable probabilistic error bounds. Both algorithms are asymptotic convergent and consistently outperform the plain Monte Carlo method.

The assumption of nonnegative coefficients in the multi-factor model is made in Chapter 5. To make sure that this assumption is satisfied, it suffices to find a nonnegative factorization for a given correlation matrix. A dedicated algorithm for this purpose is presented in chapter 6.

In the last chapter we derive two types of saddlepoint approximations to expectations in the form of $\mathbb{E}[(X-K)^+]$ and $\mathbb{E}[X|X \geq K]$, where X is the sum of n independent random variables and K is a known constant. We establish error convergence rates for both types of approximations in the *i.i.d.* case. The approximations are further extended to cover the case of lattice variables. The approximations have direct applications in finance and insurance, e.g., to the pricing of the Collateralized Debt Obligations (CDO) and the calculation of the expected shortfall of a credit or insurance portfolio.

Chapter 2

Higher Order Saddlepoint Approximations in the Vasicek Portfolio Credit Loss Model

2.1 Introduction

This chapter utilizes the saddlepoint approximation as an efficient tool to estimate portfolio credit loss distribution in the Vasicek model. The saddlepoint approximation method is well known to provide good approximations to very small tail probabilities, which makes it a very suitable technique in the context of portfolio credit loss. The use of saddlepoint approximation in portfolio credit loss is pioneered in a series of articles by Martin et al. (2001*a*,*b*). Gordy (2002) showed that saddlepoint approximation is fast and robust when applied to CreditRisk$^+$. All of them apply the saddlepoint approximation to the *unconditional* MGF of loss L, despite the fact that L_i are not independent. Annaert, Garcia, Lamoot & Lanine (2006) show that the procedure described in Gordy (2002) may give inaccurate results in case of portfolios with high skewness and kurtosis in exposure size. This chapter differs substantially from them in that we apply the saddlepoint approximation to the *conditional* MGF of L *given the common factor* Y, so that $L(Y) = \sum L_i(Y)$ is a weighted sum of independent random variables, which is exactly the situation where the saddlepoint approximation will work well. We show that this change in implementation of the saddlepoint approximation leads to very accurate results on the portfolio loss

distribution, the VaR and VaR contribution, even for small-sized portfolios and portfolios with exposure concentration. In section 2.3 we will show by a numerical example that the accuracy of our procedure is not impaired by high skewness and kurtosis in exposure size. In addition to the VaR and VaR contribution, we also give the saddlepoint approximations for the Expected Shortfall (ES) and ES contribution.

2.2 Saddlepoint approximations in the Vasicek model: a conditional approach

We should assume that the portfolio loss L, which is discrete in the Vasicek model, can be well approximated by a continuous random variable with a $p.d.f.$ so that we can employ the saddlepoint approximation formulas for density and tail probability that appeared in §1.3.

In the Vasicek model obligors are modeled as independent Bernoulli random variables conditional on the common factor, with

$$p_i(Y) = \mathbb{P}(D_i = 1|Y) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right). \tag{2.1}$$

The application of the saddlepoint approximation is therefore straightforward. The conditional MGF of L is given by

$$\mathcal{M}(t, Y) = \prod_{i=1}^{n} \left(1 - p_i(Y) + p_i(Y)e^{\omega_i t}\right). \tag{2.2}$$

The conditional CGF and its derivatives up to fourth order are defined as follows:

$$\mathcal{K}(t, Y) = \sum_{i=1}^{n} \log\left(1 - p_i(Y) + p_i(Y)e^{\omega_i t}\right), \tag{2.3}$$

$$\mathcal{K}'(t, Y) = \sum_{i=1}^{n} \frac{\omega_i p_i(Y)e^{\omega_i t}}{1 - p_i(Y) + p_i(Y)e^{\omega_i t}}, \tag{2.4}$$

$$\mathcal{K}''(t, Y) = \sum_{i=1}^{n} \frac{(1 - p_i(Y))\omega_i^2 p_i(Y)e^{\omega_i t}}{[1 - p_i(Y) + p_i(Y)e^{\omega_i t}]^2} \tag{2.5}$$

$$\mathcal{K}'''(t,Y)=\sum_{i=1}^{n}\left\{\frac{(1-p_i(Y))\omega_i^3 p_i(Y)e^{\omega_i t}}{[1-p_i(Y)+p_i(Y)e^{\omega_i t}]^2}-\frac{2(1-p_i(Y))\omega_i^3 p_i^2(Y)e^{2\omega_i t}}{[1-p_i(Y)+p_i(Y)e^{\omega_i t}]^3}\right\}, \tag{2.6}$$

$$\begin{aligned}\mathcal{K}^{(4)}(t,Y)=\sum_{i=1}^{n}\Bigg\{&\frac{(1-p_i(Y))\omega_i^4 p_i(Y)e^{\omega_i t}}{[1-p_i(Y)+p_i(Y)e^{\omega_i t}]^2}-\frac{6(1-p_i(Y))\omega_i^4 p_i^2(Y)e^{2\omega_i t}}{[1-p_i(Y)+p_i(Y)e^{\omega_i t}]^3}\\&+\frac{6(1-p_i(Y))\omega_i^4 p_i^3(Y)e^{3\omega_i t}}{[1-p_i(Y)+p_i(Y)e^{\omega_i t}]^4}\Bigg\}.\end{aligned} \tag{2.7}$$

With $\mathcal{K}(t,Y)$ available, we are able to calculate the conditional loss density $f_L(x|Y)$ and the conditional tail probability $\mathbb{P}(L>x|Y)$ for loss level x by the saddlepoint approximation. Since $\mathcal{K}'(t,Y)$ is a monotonically increasing function of t and it is bounded in the interval $[0,\sum\omega_i]$, the equation $\mathcal{K}'(t,Y)=x$ admits a unique solution T for $x\in[0,\sum\omega_i]$. Integrating over Y gives the unconditional loss density and tail probability. For example, the tail probability is given by

$$\mathbb{P}(L>x)=\mathbb{E}_Y[\mathbb{P}(L>x|Y)]. \tag{2.8}$$

The VaR can then be found by inverting the loss distribution. Moreover, to obtain the VaR contribution, we differentiate $\mathbb{P}(L>x)$ with respect to the exposure:

$$\begin{aligned}&\frac{\partial}{\partial\omega_i}\mathbb{P}(L>x)\\=&\mathbb{E}_Y\left\{\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\left[\frac{1}{t}\frac{\partial\mathcal{K}(t,Y)}{\partial\omega_i}-\frac{\partial x}{\partial\omega_i}\right]\exp(\mathcal{K}(t,Y)-tx)\mathrm{d}t\right\}.\end{aligned} \tag{2.9}$$

Here we replace x by VaR_α. Since the tail probability $\mathbb{P}(L>\mathrm{VaR}_\alpha)$ is fixed at $1-\alpha$, the left hand side should vanish and we obtain

$$\begin{aligned}\omega_i\frac{\partial\mathrm{VaR}_\alpha}{\partial\omega_i}&=\omega_i\frac{\mathbb{E}_Y\left[\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\frac{\partial\mathcal{K}(t,Y)}{\partial\omega_i}\frac{1}{t}\exp\left(\mathcal{K}(t,Y)-t\mathrm{VaR}_\alpha\right)\mathrm{d}t\right]}{\mathbb{E}_Y\left[\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\exp(\mathcal{K}(t,Y)-t\mathrm{VaR}_\alpha)\mathrm{d}t\right]}\\&=\omega_i\frac{\mathbb{E}_Y\left[\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\frac{p_i(Y)e^{\omega_i t}\exp\left(\mathcal{K}(t,Y)-t\mathrm{VaR}_\alpha\right)}{1-p_i(Y)+p_i(Y)e^{\omega_i t}}\mathrm{d}t\right]}{\mathbb{E}_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\right]}.\end{aligned} \tag{2.10}$$

If we define

$$\hat{\mathcal{K}}^i(t,Y)=\log\left(p_i(Y)e^{\omega_i t}\right)+\sum_{j\neq i}\log\left(1-p_j(Y)+p_j(Y)e^{\omega_j t}\right),$$

which can be thought of as the CGF of L given Y and $D_i = 1$, (2.10) is rewritten as

$$\omega_i \frac{\partial \mathrm{VaR}_\alpha}{\partial \omega_i} = \omega_i \frac{\mathbb{E}_Y\left[\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} \exp\left(\hat{\mathcal{K}}^i(t,Y) - t\mathrm{VaR}_\alpha\right) \mathrm{d}t\right]}{\mathbb{E}_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\right]}. \tag{2.11}$$

Both the numerator and the denominator can be approximated by the saddlepoint method.

The VaR contribution can also be derived differently. Let us write $\hat{L}^i = \sum_{j\neq i} \omega_j D_j$. We have

$$\begin{aligned} \omega_i \mathbb{E}(D_i|L = \mathrm{VaR}_\alpha) &= \omega_i \frac{f(L = \mathrm{VaR}_\alpha; D_i = 1)}{f_L(\mathrm{VaR}_\alpha)} \\ &= \omega_i \frac{\mathbb{E}_Y\left[f\left(\hat{L}^i = \mathrm{VaR}_\alpha - \omega_i \middle| Y\right) p_i(Y)\right]}{\mathbb{E}_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\right]}. \end{aligned} \tag{2.12}$$

The conditional density in the numerator is the conditional loss density of a portfolio excluding obligor i and can again be calculated by the saddlepoint approximation. We note that (2.11) and (2.12) are essentially the same because both formulas use the saddlepoint T that solves

$$\sum_{j\neq i} \frac{\omega_j p_j(Y) e^{\omega_j t}}{1 - p_j(Y) + p_j(Y) e^{\omega_j t}} = \mathrm{VaR}_\alpha - \omega_i. \tag{2.13}$$

Similarly, the ES contributions are given by

$$\omega_i \mathbb{E}(D_i|L \geq \mathrm{VaR}_\alpha) = \omega_i \frac{\mathbb{E}_Y\left[\mathbb{P}\left(\hat{L}^i \geq \mathrm{VaR}_\alpha - \omega_i \middle| Y\right) p_i(Y)\right]}{\mathbb{E}_Y\left[\mathbb{P}(L \geq \mathrm{VaR}_\alpha|Y)\right]}. \tag{2.14}$$

In this chapter ES will be estimated by simply summing up all the ES contributions, i.e.,

$$ES_\alpha = \sum \omega_i \mathbb{E}(D_i|L \geq \mathrm{VaR}_\alpha). \tag{2.15}$$

Note that explicit saddlepoint approximations to ES can also be obtained without first computing ES contributions. The approximate formulas will be derived in Chapter 7.

Remark 2.1. Although the obligors in a portfolio are assumed to be completely heterogeneous, for the sake of computational efficiency, it is advisable to group obligors as much as possible into homogeneous buckets with similar

characteristics, esp. for large portfolios. The main advantages of doing this are (i) the expedition of the calculation of conditional CGF and its partial derivatives and (ii) a reduction of the amount of risk contributions that need to be computed.

Remark 2.2. Martin et al. (2001*b*) proposed a simple estimate to the VaR contribution, which reads

$$\mathrm{VaRC}_{i,\alpha} \approx \frac{E_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\frac{\omega_i}{T}\frac{\partial \mathcal{K}(t,Y)}{\partial \omega_i}\Big|_{t=T}\right]}{E_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\right]} \tag{2.16}$$

in the Vasicek model. In our numerical examples we show, however, that this approximation may be inaccurate.

2.3 Numerical results

We now illustrate the performance of the saddlepoint approximation in the Vasicek one-factor model. For the implementation of the saddlepoint approximation, we always employ the Lugannani-Rice formula (1.23) for the tail probability. We truncate the common factor Y in the interval $[-5, 5]$ so that the probability of Y falling out of this interval is merely 5.7×10^{-7}. Discretization of Y is done by Gauss-Legendre quadrature, generating 1000 abscissas and weights. The four examples evaluated are:

- Example 1: A homogeneous portfolio with $n = 1000$ obligors, each with EAD=1, PD=0.01 and $\rho = 0.2$.
- Example 2: A portfolio consisting of $n = 100$ obligors with $\omega_i = i$, $k = 1, 2, \ldots 100$, PD=0.1, $\rho = 0.2$.
- Example 3: A portfolio consisting of 1 obligor with $\mathrm{EAD}_1 = 100$ and 10,000 obligors with $\mathrm{EAD}_2 = 1$. All obligors have PD=0.005 and $\rho = 0.2$.
- Example 4: This portfolio is taken from Glasserman (2006). All 100 obligors have PD=0.01 and $\rho = 0.5$. The exposures are

$$\omega_i = \begin{cases} 1, & i = 1, \ldots, 20 \\ 4, & i = 21, \ldots, 40 \\ 9, & i = 41, \ldots, 60 \\ 16, & i = 61, \ldots, 80 \\ 25, & i = 81, \ldots, 100 \end{cases}$$

The HHI values for the four portfolios are as follows,

Example	1	2	3	4
HHI	0.001	0.0133	0.0002	0.0162
$1/n$	0.001	0.01	0.0001	0.01

with $1/n$ being the HHI value in case of no exposure concentration. So, only Example 1 deals with a homogeneous portfolio.

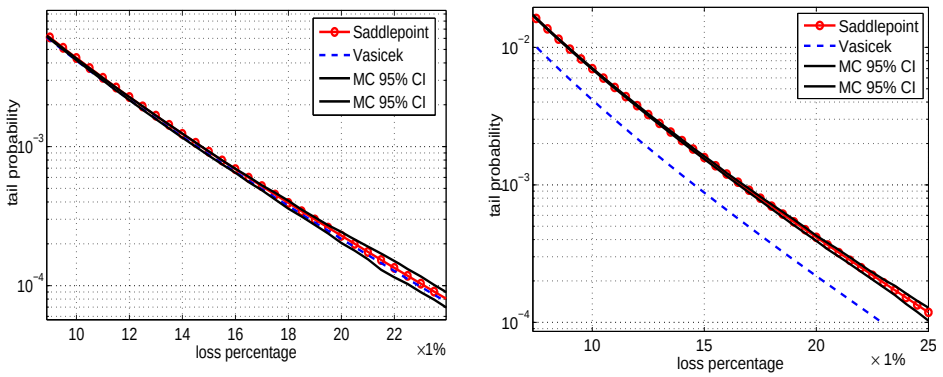
We compare the loss distribution from the saddlepoint approximation to results from the analytic Vasicek formula and from Monte Carlo simulation in the first two examples. Our benchmark is the sample mean and the accompanying 95% confidence intervals obtained from 10 subsamples of Monte Carlo simulation with 4 million replications. The loss distribution corresponding to the Vasicek model is obtained by inverting the VaR given by Vasicek's formula (1.10) for a series of quantiles α.

Example 1 is an ideal case for the Vasicek formula (1.10) to be accurate. The loss distributions from different methods are presented in Figure 2.1(a). The x-axis represents the loss percentage, i.e., the loss amount in proportion to the total exposure. The y-axis, the tail probability $\mathbb{P}(L > x)$, is in log-scale. It can be seen that both the Vasicek formula and the saddlepoint approximation follow our benchmark very well.

In Example 2, the Vasicek's formula significantly underestimates the risk, as is demonstrated in Figure 2.1(b). This implies the presence of exposure concentration in the portfolio. We observe however that the saddlepoint approximation gives results comparable to simulation in this example.

We show more details of the errors made by the saddlepoint approximation for Examples 1 and 2 in Figure 2.2. Concentrating on the loss percentages from 15% to 25%, that roughly correspond to quantiles from 99.9% to 99.99% for both examples, we report relative errors compared to the sample means of the ten subsample estimates. The normalized standard deviations of Monte Carlo simulation with 4 million replications are also provided for comparison. We find that in the tail of the distribution, the relative errors of the saddlepoint approximation are typically smaller than the standard deviations. Furthermore we see that as the loss level increases the standard deviation of the Monte Carlo simulation increases significantly, whereas the accuracy of the saddlepoint approximation does not seem affected. This is highly desirable because the tail of the distribution is the center of interest.

Example 3 is a particular test case for which the VaR and VaR contributions can be computed almost exactly by the binomial expansion method (BEM) if we treat the portfolio loss as a discrete variable. It is therefore



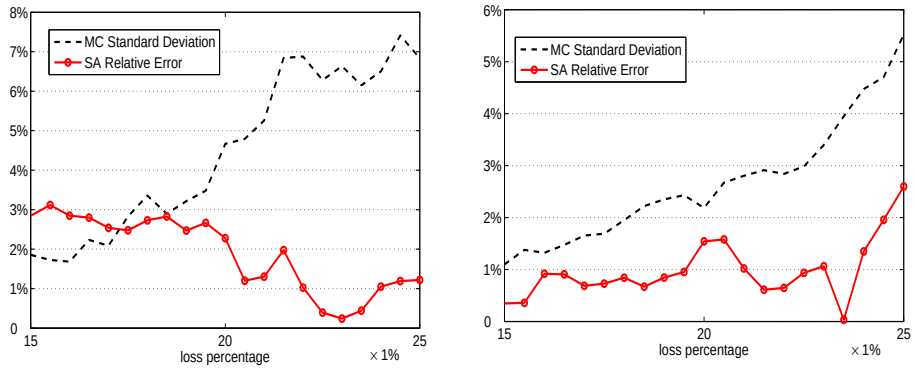
(a) Example 1 (b) Example 2

FIGURE 2.1: Comparison of the Saddlepoint approximation, Vasicek's formula and Monte Carlo simulation for the loss distribution in Examples 1 and 2. The Monte Carlo 95% confidence interval (CI) is constructed using 10 subsamples of 4 million replications each.

a suitable test portfolio for the calculation of VaR contributions. BEM will serve as the benchmark for both the VaR and the VaR contributions. More details on BEM can be found in the Appendix 2.A. The loss distribution of this portfolio given by the saddlepoint approximation and the BEM are shown in Figure 2.3. The saddlepoint approximations again follow our benchmark very well.

As for the VaR contribution, we first consider a fixed loss level $L = 922$, which lies around the 99.9% quantile. We compute the VaR contributions of both the large obligor (VaRC_1) and any small obligor (VaRC_2). We use both the standard and higher order saddlepoint approximation given by (1.20) and (1.22), respectively (denoted by SA2 and SA4). Results are shown in Table 2.1 and in parenthesis are the relative errors of the saddlepoint approximation to the benchmark. Besides, we compute the Vasicek VaR contribution, the saddlepoint approximation for the VaR contribution as given by (2.16) (denoted by SA_M) for comparison.

The results given by the benchmark BEM show that the VaR contribution increases non-linearly with the size of the exposure. Both the standard and higher order saddlepoint method successfully capture this feature and give the VaR contributions with small relative errors. The higher order approximation, with relative error less that 1%, outperforms the standard approximation. The only (negligible) problem is that the VaR contributions do not add up to the total VaR exactly. It is also clear that the VaR



(a) Example 1 (b) Example 2

FIGURE 2.2: Comparison of the normalized standard deviations of Monte Carlo simulation with 4 million replications and relative errors of the saddlepoint approximations.

contributions of the large obligor (VaRC_1) obtained from Vasicek and SA_M are both relatively far from the true value. The Vasicek contribution is proportional to the effective exposure and therefore it underestimates the large obligor's risk contribution. SA_M penalizes large exposure too much.

Next we consider a fixed confidence level $\alpha = 99.99\%$ in Example 3, which is truly far in the tail. The Lugannani-Rice formula will be used to compute the loss distribution and the ES contributions. The higher order saddlepoint approximation is used for the VaR contributions. Results are shown in Table 2.2. The accuracy of the saddlpoint approximation is highly satisfactory for all estimates of VaR contributions, ES contributions and ES. The table suggests that the approximation is slightly more accurate for the VaR contribution than for the ES contribution. This can be understood roughly because the relative error of Daniels formula is $O(n^{-2})$ and that of Lugannani-Rice formula is $O(n^{-3/2})$, with n being the number of *i.i.d.* random variables (although in our example L_i are not really identically distributed).

We remark that in Example 3 the skewness and kurtosis in exposure size are 99.985 and 9998, respectively. They are much higher than in the portfolios 4 and 5 given in Annaert et al. (2006), where it is shown that the accuracy and reliability of the saddlepoint approximation obtained from Gordy's (2002) procedure may deteriorate. In our approach high skewness and kurtosis do not pose any problem with respect to accuracy.

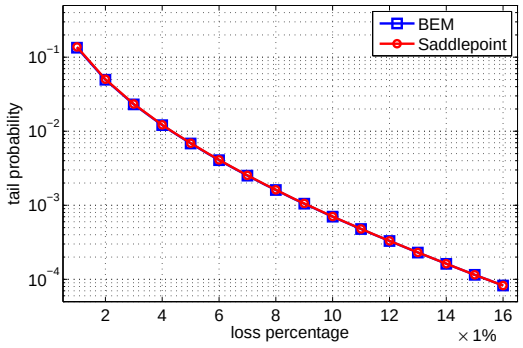


FIGURE 2.3: Tail probability given by the saddlepoint approximation and the BEM for portfolio in Example 3.

For the portfolio in Example 4 we report the expected shortfall contributions at the loss level $L = 100$. In Glasserman (2006) both the VaR contributions and ES contributions estimates from importance sampling are provided. We opt for the ES contributions since they are supposedly more accurate and robust: they are conditioned on a relatively less rare event, $\{L \geq 100\}$ compared to $\{L = 100\}$ for VaR contributions. The results are illustrated in Table 2.3. Those estimates from importance sampling, using as many as 250,000 replications, are taken from Table 2 in Glasserman (2006) and seen as our benchmark. We observe only marginal differences between the ES contributions given by the saddlepoint approximation and the benchmark.

It is also interesting to explore the efficiency of the two methods as both appropriately accommodate exposure concentration. Note that both saddlepoint approximation and importance sampling with exponential twisting involve finding the saddlepoint as a solution to Eq. (1.18) for each realization of the common factor. With the saddlepoint at hand, saddlepoint approximation can be obtained analytically, while importance sampling needs to simulate idiosyncratic risks for all obligors in the portfolio. It is more likely than not that the saddlepoint approximation is faster than importance sampling, especially for portfolios with a large amount of obligors. Besides, we find that in a one-factor model generally a draw of 100-1000 common factors based on the Gauss-Legendre quadrature is sufficient in terms of accuracy for the saddlepoint approximation, while for importance sampling many more points are necessary to obtain an estimate with a small variance, particularly for the estimation of the VaR contributions. The advantage of importance sampling is that it can compute tail probabilities for a series of

loss levels with one set of generated scenarios (see Glasserman & Li 2005). For the saddlepoint approximation different saddlepoints need to be found for different loss levels. A final remark is that in all the four examples the saddlepoint approximation is able to find the portfolio VaR in less than 5 seconds on a Pentium 4 2.8 GHz desktop.

2.4 Further discussion

2.4.1 Adaptive integration in a one-factor model

The efficiency of saddlepoint approximations can be improved significantly by using an adaptive integration technique.

A key observation is that for a fixed loss level x, the function $f(y) = \mathbb{P}(L > x|Y = y)$ is a non-increasing function of y. A formal proof of the monotonicity can be found in Chapter 5, in the more general setup of multi-factor models.

Take as an example a homogeneous portfolio consisting of 1000 obligors with $\omega_i = 1$, $p_i = 0.0033$ and $\rho_i = 0.2$, $i = 1, \ldots, 1000$. The integrand $\mathbb{P}(L > x|Y)$ with $x = 100$ is illustrated in Figure 2.4. It is indeed a non-increasing function. Furthermore, it decreases rapidly from its upper bound 1 to its lower bound 0 for Y in a narrow band (between the two dashed vertical lines in Figure 2.4) much smaller than the domain of Y. Note that the band will move toward the left tail of Y as the loss level x increases. Moreover the width of the band should further decrease as the number of the obligors n increases. Asymptotically, as $n \to \infty$, $\mathbb{P}(L/\sum \omega_i > x|Y)$ approaches a Heaviside step function. Due to the law of large numbers, we have $L(Y)/\sum \omega_i \to \sum \omega_i p_i(Y)/\sum \omega_i$ *a.s.* and $\mathbb{P}(L/\sum \omega_i > x|Y) \to \mathbf{1}_{\{\sum \omega_i p_i(Y)/\sum \omega_i > x\}}$.

Generally an N-point quadrature rule demands N integrand evaluations. However since in our problem the integrand is monotone and bounded in $[0, 1]$, significantly fewer evaluations are required for the same accuracy with an adaptive integration algorithm. We propose a simple procedure that utilizes the nodes of an N-point Gauss-Legendre quadrature rule. It produces identical results for the tail probability $\mathbb{P}(L > x)$ as the N-point quadrature but the number of integrand evaluations is substantially reduced.

Denote the Gauss nodes and weights by Y_k with $Y_1 > Y_2 > ... > Y_N$ and u_k, $k = 1, \ldots, N$, respectively. First identify the smallest node y_1 giving $f(y_1) = 0$. Then discard all nodes larger than y_1 and proceed sequentially with decreasing Y until we find a y_2 such that $f(y_2) = 1$. For all $Y < y_2$ we set $f(Y) = 1$. Finally sum over. The procedure is summarized in Algorithm

	VaRC_1	VaRC_2	$\sum \text{VaRC}$
BEM	12.61	0.0909	921.95
Vasicek	9.13	0.0913	922
SA_M	21.82	0.0900	921.99
SA2	12.24(2.93%)	0.0904(0.55%)	916.64(0.58%)
SA4	12.65(0.32%)	0.0907(0.22%)	920.00(0.21%)

TABLE 2.1: VaR contributions at the loss level $L = 922$. In parenthesis are the relative errors of the saddlepoint approximation to the benchmark. The portfolio is given in Example 3.

	$\text{VaR}_{99.99\%}$	VaRC_1	VaRC_2	ESC_1	ESC_2	ES
BEM	1558	19.79	0.1538	23.14	0.1839	1862.51
SA	1558	19.71 (0.4%)	0.1537 (0.06%)	23.18 (0.17%)	0.1848 (0.49%)	1871 (0.46%)

TABLE 2.2: VaR contributions and ES contributions at the loss level $\text{VaR}_{99.99\%}$. In parenthesis are the relative errors of the saddlepoint approximation to the benchmark. The portfolio is given in Example 3.

Obligor	1-20	21-40	41-60	61-80	81-100
IS	0.10	0.42	1.02	2.03	3.67
SA	0.1017	0.4254	1.0327	2.0453	3.6835

TABLE 2.3: Comparison of importance sampling (IS) and saddlepoint approximation (SA) for expected shortfall contributions in Example 4 at the loss level $x = 100$.

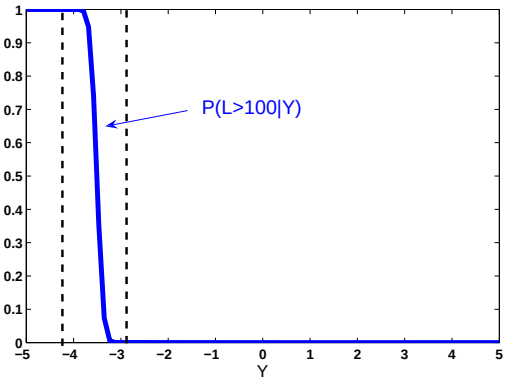


FIGURE 2.4: The integrand $\mathbb{P}(L > 100|Y)$ as a function of the common factor Y. The portfolio consists of 1000 obligors with $\omega_i = 1$, $p_i = 0.0033$ and $\rho_i = 0.2$, $i = 1, \ldots, 1000$.

2.1 where $\mathbb{P}(L > x)$ is approximated by I.

Algorithm 2.1 Adaptive integration in a one-factor model

Generate the N-degree Gauss-Legendre nodes $Y_1, \ldots, Y_N$ and weights $u_1, \ldots, u_N$
Find $Y_i = \min\{Y_k | f(Y_k) = 0, k = 1, \ldots, N\}$
$j = i + 1$, $I = 0$
while $j \leq N, f(Y_j) < 1$ **do**
  $I = I + f(Y_j) \cdot \phi(Y_j) \cdot u_j$
  $j = j + 1$
end while
$I = I + \sum_{k=j}^{N} \phi(Y_k) \cdot u_k$

For the above example with $N = 100$ this algorithm results in less than 20 integrand evaluations. It is evident that an adaptive integration algorithm is able to effectively reduce the amount of computations in a one-factor model.

2.4.2 Multi-factor models

We further present an Example 5 under a Gaussian multi-factor model, taken from Glasserman & Li (2005). It is a 21-factor model with $n = 1000$ heterogeneous obligors. The exposures ω_i increase linearly from 1 to 100 as

i increases from 1 to 1000. PDs have the following form:

$$p_i = 0.01 * (1 + \sin(16\pi i/n)), i = 1, \ldots, n.$$

The matrix of factor loadings, $\mathbf{A} = (a_{ij}, i = 1, \ldots, 1000, j = 1, \ldots, 21)$, has the following block structure:

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} & F & & & G \\ R & & \ddots & & \vdots \\ & & & F & G \end{array} \right), G = \begin{pmatrix} g & & \\ & \ddots & \\ & & g \end{pmatrix},$$

with R a column vector of 1000 entries, all equal to 0.8; F a column vector of 100 entries, all equal to 0.4; G a 100×10 matrix; and g a column vector of 10 entries, all equal to 0.8.

In a multi-factor model with more than three factors, instead of Gauss quadrature, Monte Carlo simulation or low discrepancy sequences can be employed for the integration. The saddlepoint approximation itself is however not affected, since all the information of the common factors is encapsulated in $p_i(Y)$ before starting the approximation procedure. We note that when x is large and Y is large and positive, $\mathbb{P}(L > x|Y)$ will tend to zero and thus the integration in (2.8) will not be efficient. It is a natural idea to resort to importance sampling for a significant improvement in such cases. By choosing a suitable $\mathbb{P}$-equivalent probability measure $\mathbb{Q}$ under which the mean of the common factor Y is shifted, the tail probability can be rewritten as

$$\mathbb{P}(L > x) = \mathbb{E}_Y[\mathbb{P}(L > x|Y)] = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{P}(L > x|Y)\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right].$$

Several procedures to find an optimal measure $\mathbb{Q}$ are suggested by Glasserman & Li (2005), Glasserman (2006).

A hybrid method of saddlepoint approximation and importance sampling is more efficient than pure importance sampling, since in the simulation only the common factors need to be generated and not the idiosyncratic risks. This is more advantageous for large portfolios where the number of obligors is considerably larger than the number of factors. Moreover for the calculation of the VaR contributions importance sampling can only use the few replications $L = x$, whereas the hybrid method need not condition on this rare event.

We employ the hybrid method for Example 5. It is carried out as follows. Taking the same mean shift in the common factors as in Glasserman & Li (2005), i.e., 2.46 for the first factor and 0.2 for the other components, we generate 10 subsamples of simulation with 1000 replications each.

The saddlepoint approximation is employed to compute the tail probability conditional on each realization of the common factors. Afterwards the conditional tail probabilities are aggregated with proper likelihood ratios. Table 2.4 shows the resulting tail probabilities and their standard deviations for five loss levels from 10000 to 30000, with reference to those point estimates reported in Glasserman & Li (2005), which are based on pure importance sampling. Differences between the two methods appear to be immaterial. In addition we produce a column of standard deviations (std) of pure importance sampling with 10×1000 replications alongside the point estimates. It is immediate to see that the hybrid method yields smaller standard deviations. The larger variance associated to the pure importance sampling method can be attributed to variations in the idiosyncratic risks.

x	IS	std	Hybrid	std
10000	0.0114	8.17×10^{-4}	0.01139	4.55×10^{-4}
14000	0.0065	3.85×10^{-4}	0.00641	3.41×10^{-4}
18000	0.0037	2.57×10^{-4}	0.00367	1.49×10^{-4}
22000	0.0021	1.89×10^{-4}	0.00210	8.12×10^{-5}
30000	0.0006	6.53×10^{-5}	0.00063	4.81×10^{-5}

TABLE 2.4: Comparison of importance sampling (IS) and the hybrid method for point estimates and standard deviations (std) of tail probability $\mathbb{P}(L > x)$ at various loss levels. The portfolio is given in Example 5.

2.4.3 Non-Gaussian factor models

Although we confine our numerical experiments to the Gaussian factor models in previous sections, the saddlepoint approximation technique can be readily applied to all conditionally independent models with common factors that follow any distribution. Recall that in our approach the distribution of portfolio loss L is obtained by integrating the conditional distribution of $L(Y)$ and the saddlepoint approximation only deals with $L(Y)$. A different choice of mixture model gives a difference in the form of the conditional default probability $p_i(Y)$, eg., in the Vasicek one-factor model $p_i(Y)$ is given by (2.1) and in CreditRisk$^+$

$$p_i(Y) = p_i \left(w_{i0} + \sum Y_k w_{ik} \right),$$

where Y_k are assumed to be independently gamma distributed (see Gordy 2002). But then the conditional portfolio loss $L(Y) = \sum \omega_i D_i(Y)$ reduces

to a weighted sum of independent Bernoulli random variables, whose MGF always exists. As the main requirement for calculating a saddlepoint approximation is the existence of a MGF, the saddlepoint approximation is applicable in any factor model.

2.4.4 Random LGD

Although discussed in great detail in Chapter 4, here we will provide a first flavor of the use of saddlepoint approximations in models with random loss given default (LGD).

When the LGD, which was assumed to be constant in the Vasicek model, becomes a random variable, the conditional CGF reads

$$\mathcal{K}(t,Y) = \sum \log\left[1 - p_i(Y) + p_i(Y)\mathbb{E}(e^{\omega_i \Lambda t}|Y)\right]. \tag{2.17}$$

Various forms of distribution of LGD can be found in the literature. For example, in Frye's (2000) model, the LGD is modeled as a normal random variable with mean μ and standard deviation σ such that

$$\Lambda_i = \mu + \sigma\left(-b_i Y + \sqrt{1-b_i^2}\epsilon_i\right).$$

Here the ϵ_i, independent to Y, are assumed to be *i.i.d.* standard normal variables and the b_i are assumed to be positive to insure the correct qualitative effect of LGD, which is mostly determined by the value of collateral. It should tend to be higher when the economy is weak and lower when the economy is strong. It follows that

$$\begin{aligned}\mathbb{E}(e^{\omega_i \Lambda t}|Y) &= e^{\omega_i(\mu-\sigma b_i Y)t}\mathbb{E}(e^{\omega_i \sigma\sqrt{1-b_i^2}\epsilon_i t}) \\ &= \exp\left(\omega_i(\mu-\sigma b_i Y)t + \omega_i^2\sigma^2(1-b_i^2)t^2/2\right). \end{aligned} \tag{2.18}$$

After substitution of (2.18) into (2.17), we see that a random LGD will not complicate the problem further.

Other examples are given in §4.5 in detail along with numerical experiments.

2.5 Conclusions

We have described a new procedure to embed the saddlepoint approximation as a useful tool in portfolio credit loss modeling. The saddlepoint approximation is applied to the conditional moment generating function of

the portfolio loss given the common factor in the Vasicek one-factor model. The saddlepoint approximations, esp. the higher order approximations, are able to produce accurate results on both the VaR and the VaR contribution. The ES and ES contribution can also be computed satisfactorily. We have also illustrated that the saddlepoint approximation works well for small-sized portfolios and portfolios with exposure concentration, where Vasicek's asymptotic formulas fail. We further point out that the saddlepoint approximation is a flexible method which can be applied in quite general situations, for example, multi-factor models, non-Gaussian factor models and models with random LGD. The extensions will be the subject of forthcoming chapters.

2.A Binomial expansion method

The binomial expansion method is similar to the recursive method proposed by Andersen et al. (2003) in the sense that both methods treat the portfolio loss as a discrete variable. The former method is tailor-made for the portfolio in Example 3 we considered in §2.3, while the latter method can be applied to more general portfolios. An evaluation of the recursive method can be found in Glasserman & Ruiz-Mata (2006).

Consider a portfolio consisting of 1 obligor with $\text{EAD}_1 = k$, PD$= p_1$ and n obligors with $\text{EAD}_2 = 1$, PD$= p_2$. In a Bernoulli mixture model, the losses of the obligors are conditionally independent given the common factor Y. Let $p_1(\cdot)$ and $p_2(\cdot)$ be the conditional default probabilities, we have

$$\begin{aligned}\mathbb{P}(L=m) &= \int \mathbb{P}(L=m|y)\mathrm{d}F_Y(y) \\ &= \int p_1(y)\mathbb{P}(L^n=m-k|y)+(1-p_1(y))\mathbb{P}(L^n=m|y)\ \mathrm{d}F_Y(y),\end{aligned}$$

where $F_Y(\cdot)$ denotes the *c.d.f.* of Y and

$$\mathbb{P}(L^n=m|y)=\binom{n}{m}(p_2(y))^m(1-p_2(y))^{n-m}.$$

The VaR and VaR contributions are then given, respectively, by

$$\text{VaR}_\alpha = \inf\left\{x\Big|\sum_{m=0}^{x}\mathbb{P}(L=m)\geq\alpha\right\},$$

and

$$\mathrm{VaRC}_1 = \frac{\int p_1(y)\mathbb{P}(L^n = \mathrm{VaR}_\alpha - k|y)\mathrm{d}F_Y(y)}{\mathbb{P}(L = \mathrm{VaR}_\alpha)},$$

$$\begin{aligned}\mathrm{VaRC}_2 &= \frac{1}{\mathbb{P}(L = \mathrm{VaR}_\alpha)}\left\{\int p_2(y)p_1(y)\mathbb{P}(L^{n-1} = \mathrm{VaR}_\alpha - k - 1|y)\mathrm{d}F_Y(y) + \right. \\ &\left. + \int p_2(y)(1 - p_1(y))\mathbb{P}(L^{n-1} = \mathrm{VaR}_\alpha - 1|y)\mathrm{d}F_Y(y)\right\}.\end{aligned}$$

The ES contributions are computed according to (2.14) with

$$\mathbb{P}(L \geq x) = 1 - \sum_{m=0}^{x-1}\mathbb{P}(L = m)$$

and ES is obtained by (2.15).

Chapter 3

Computation of VaR and VaR Contribution in the Vasicek Portfolio Credit Loss Model: A Comparative Study

3.1 Introduction

This chapter provides a comparative study on different numerical methods for the estimation of the VaR and the marginal VaR contribution (VaRC) in the Vasicek one-factor portfolio credit loss model. We investigate each method in terms of speed, accuracy and robustness and in particular explore their abilities of dealing with exposure concentration.

A variety of methods to estimate the portfolio credit risk and the risk contributions have been proposed in the literature. Glasserman & Ruiz-Mata (2006) provide an interesting comparison of methods for computing credit loss distributions. The methods considered there are plain Monte Carlo simulation, a recursive method due to Andersen et al. (2003), the saddlepoint approximation, and the numerical transform inversion as in Abate et al. (2000). They conclude that the plain Monte Carlo method is the best method in a multi-factor setting in terms of speed and accuracy, followed by the saddlepoint approximation. They find that the recursive method performs well when the number of obligors is small but becomes slow as the number of obligors increases, particularly for high loss levels. This is be-

cause the recursive method computes the entire loss distribution and when the number of obligors increases, the maximum total loss increases in the meantime. They also find that the numerical transform inversion method gives acceptable estimates for small loss levels but the approximation worsens for higher loss levels. This is not surprising. This method numerically inverts the Bromwich integral, whose integrand becomes highly oscillatory and extremely difficult to handle for high loss levels.

The perspective of our comparison in this chapter is quite different from Glasserman & Ruiz-Mata (2006). First we concentrate on the one-factor model. Secondly we are mainly interested in VaR_α when α is close to 1, i.e., high loss levels. Thirdly we are also interested in the estimation of the marginal VaR contribution. Finally we would like to investigate how well the problem of exposure concentration can be handled.

We point out that the conclusions of Glasserman & Ruiz-Mata (2006) are based on portfolios with less than 1000 obligors. But in practice it will not be surprising that a bank's credit portfolio has more than tens of thousands of obligors. The plain Monte Carlo simulation will certainly become more demanding in computation time as the portfolio size increases. After all, a true problem with plain simulation is the estimation of the marginal VaR contribution, which is based on the scenarios that portfolio loss equals VaR. These are extremely rare events. We should for this reason consider importance sampling as in Glasserman & Li (2005), Glasserman (2006) instead of plain simulation. We will drop the recursive method and the numerical transform inversion method for obvious reasons given above. Note that Debuysscher & Szegö (2003) suggest that the numerical inversion can be expedited by fast Fourier transform. However a straightforward implementation of FFT also suffers from the same problem as the numerical transform inversion. We should instead include the normal approximation method as in Martin (2004), which is a direct application of the central limit theorem. In addition we also consider a simplified saddlepoint approximation for the estimation of VaRC.

The rest of the chapter is organized as follows. §3.2 reviews the various numerical methods we want to investigate, i.e., the normal approximation, the saddlepoint approximation, the simplified saddlepoint approximation and importance sampling. A stylized portfolio is considered in §3.3. §3.4 discusses the robustness of each method. The last section concludes along with some further discussions.

3.2 Numerical methods

In this section we briefly describe the numerical methods that we want to compare for the estimation of VaR and VaRC.

3.2.1 Normal approximation

The normal approximation (NA) is a direct application of the central limit theorem (CLT) and can be found in Martin (2004). When the portfolio is not sufficiently large for the law of large numbers to hold or not very homogeneous, unsystematic risk arises. We then need to take into account the variability of the portfolio loss L conditional on the common factor Y. This can easily be approximated due to the CLT. Conditional on the common factor Y, the portfolio loss L is normally distributed with mean $\mu(Y)$ and variance $\sigma^2(Y)$ such that

$$\mu(Y) = \sum_{i=1}^{n} \omega_i p_i(Y), \quad \sigma^2(Y) = \sum_{i=1}^{n} \omega_i^2 p_i(Y)(1 - p_i(Y)), \tag{3.1}$$

where $p_i(Y) = \mathbb{P}(D_i = 1|Y) = \Phi\left(\frac{\Phi^{-1}(p_i)-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right)$ and p_i is the default probability of obligor i. It follows that the conditional tail probability reads

$$\mathbb{P}(L > x|Y) = \Phi\left(\frac{\mu(Y) - x}{\sigma(Y)}\right).$$

The unconditional tail probability can then be obtained by integrating over Y, i.e.,

$$\mathbb{P}(L > x) = \mathbb{E}_Y\left[\Phi\left(\frac{\mu(Y) - x}{\sigma(Y)}\right)\right] = \int \Phi\left(\frac{\mu(y) - x}{\sigma(y)}\right)\phi(y)\mathrm{d}y. \tag{3.2}$$

We will in this chapter approximate the integral by the Gauss quadrature, but there are also attempts to find an analytic approximation to Eq. (3.2), under the name of the granularity adjustment (cf. Gordy 2003, Wilde 2001).

To obtain the VaR contribution in the current setting, we first differentiate $\mathbb{P}(L > x)$ with respect to the effective exposure:

$$\begin{aligned}\frac{\partial}{\partial \omega_i}\mathbb{P}(L > x) =& \mathbb{E}_Y\left\{\phi\left(\frac{\mu(Y) - x}{\sigma(Y)}\right)\right. \\ &\left.\times\left[\frac{1}{\sigma(Y)}\left(\frac{\partial \mu(Y)}{\partial \omega_i} - \frac{\partial x}{\partial \omega_i}\right) - \frac{\mu(Y) - x}{\sigma^2(Y)}\frac{\partial \sigma(Y)}{\partial \omega_i}\right]\right\},\end{aligned} \tag{3.3}$$

with

$$\frac{\partial \mu(Y)}{\partial \omega_i} = p_i(Y), \quad \frac{\partial \sigma(Y)}{\partial \omega_i} = \omega_i p_i(Y)(1 - p_i(Y))/\sigma(Y). \tag{3.4}$$

Then, we replace x by VaR_α in formula (3.3). Since $\mathbb{P}(L > \mathrm{VaR}_\alpha) \equiv 1 - \alpha$, the left-hand side of Eq. (3.3) becomes zero and by rearranging terms we obtain the following VaR contribution

$$\begin{aligned} &\omega_i \frac{\partial \mathrm{VaR}_\alpha}{\partial \omega_i} \\ =&\omega_i \frac{\mathbb{E}_Y\left[\left(\frac{1}{\sigma(Y)}\frac{\partial \mu(Y)}{\partial \omega_i} - \frac{\mu(Y) - \mathrm{VaR}_\alpha}{\sigma^2(Y)}\frac{\partial \sigma(Y)}{\partial \omega_i}\right)\phi\left(\frac{\mu(Y) - \mathrm{VaR}_\alpha}{\sigma(Y)}\right)\right]}{\mathbb{E}_Y\left[\frac{1}{\sigma(Y)}\phi\left(\frac{\mu(Y) - \mathrm{VaR}_\alpha}{\sigma(Y)}\right)\right]}. \end{aligned} \tag{3.5}$$

The normal approximation is also applied in Shelton (2004) for CDO and CDO-squared pricing. Zheng (2006) employs higher order approximations as an improvement to the central limit theorem to compute CDS and CDO-squared transactions. In this chapter we will restrict ourselves to the standard normal approximation as in Martin (2004).

3.2.2 Saddlepoint approximation

The details of the saddlepoint approximations in the Vasicek model can be found in §2.2. Here we only want to emphasize that

- The estimation of the tail probability and the VaR involves finding, for each realization of the common factor Y, the saddlepoint T such that

$$\mathcal{K}'_L(T, Y) = x, \tag{3.6}$$

  where $\mathcal{K}_L(t) = \log \mathbb{E}\left[e^{tL}\right]$ is the cumulant generating function of L.

- The computation of VaRC requires finding for each obligor i and each Y a saddlepoint T in addition to Eq. (3.6) which solves

$$\sum_{j \neq i} \frac{\omega_j p_j(Y) e^{\omega_j t}}{1 - p_j(Y) + p_j(Y) e^{\omega_j t}} = \mathrm{VaR}_\alpha - \omega_i. \tag{3.7}$$

3.2.3 Simplified saddlepoint approximation

The approximation given below has been discussed briefly in §2.2 and its performance was compared to the saddlepoint approximation in §2.3. Here we shall provide more details.

For the calculation of VaRC, Martin et al. (2001*b*) propose the following estimate, also under the name of a saddlepoint approximation,

$$\mathrm{VaRC}_{i,\alpha} = \frac{\omega_i}{T}\frac{\partial \mathcal{K}_L(t)}{\partial \omega_i}\Big|_{t=T} = \frac{\omega_i p_i e^{\omega_i T}}{1 - p_i + p_i e^{\omega_i T}} \tag{3.8}$$

in the case of independent obligors. Here T is again the solution of $\mathcal{K}'_L(t) = \mathrm{VaR}_\alpha$. This estimate is also derived by Thompson & Ordovás (2003) based on the idea of an ensemble and Glasserman (2006) as a result of an asymptotic approximation.

It is straightforward to extend the independent case to the conditionally independent case as in the Vasicek model, which reads

$$\mathrm{VaRC}_{i,\alpha} \approx \frac{\mathbb{E}_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\dfrac{\omega_i p_i(Y) e^{\omega_i T}}{1 - p_i(Y) + p_i(Y) e^{\omega_i T}}\right]}{\mathbb{E}_Y\left[f_L(\mathrm{VaR}_\alpha|Y)\right]}, \tag{3.9}$$

where $f_L(\mathrm{VaR}_\alpha|Y)$ can be computed efficiently by the saddlepoint approximations. This formula can also be found in Antonov et al. (2005).

We call the estimate given by (3.9) a simplified saddlepoint approximation (SSA), in the sense that it is a simplified version of the double saddlepoint approximation to (2.10). For a portfolio with n distinct obligors, the double saddlepoint approximation requires solving (3.6) once and n times (3.7) for each realization of the common factor Y, whereas the SSA only needs the solution T to (3.6). It then assumes that T and T_i, the solutions to (3.6) and (3.7), are more or less the same for each obligor i and simply replace all T_i by the saddlepoint T. Consequently the SSA is generally faster than the SA, but it may give less accurate results if the above assumption is violated.

3.2.4 Importance sampling

Monte Carlo (MC) simulation is an all-around method which is very easy to implement. However, Monte Carlo simulation can be extremely time-consuming. The typical error convergence rate of plain Monte Carlo simulation is $O(1/\sqrt{N})$, where N is the number of simulations, requiring a large number of simulations to obtain precise results. See Boyle et al. (1997) for a review in the finance context.

Two main variance reduction techniques for Monte Carlo methods applied to portfolio credit loss can be found in the literature. Control variates are employed by Tchistiakov et al. (2004) where the Vasicek distribution is considered as a control variable. Importance sampling (IS) is adopted by Kalkbrener et al. (2004), Merino & Nyfeler (2005) for the calculation of Expected Shortfall contribution and by Glasserman & Li (2005), Glasserman (2006) for the calculation of VaR and VaRC. We note that the difficulty with Monte Carlo simulation mainly concerns the determination of VaRC since the estimate expressed in formula (1.2) is based on the very rare event that portfolio loss $L = \text{VaR}$. In this respect control variates do not provide any improvement. IS as suggested in Glasserman & Li (2005), Glasserman (2006) seems a more appropriate choice and will be adopted here.

The importance sampling procedure consists of two steps:

- Mean shifting - shift of the mean of common factors,
- Exponential twisting - change of distribution to the (conditional) default probabilities.

With mean shifting the common factor Y is sampled under probability measure $\mathbb{S}$ which is equivalent to the original measure $\mathbb{P}$ such that under $\mathbb{S}$, Y is normally distributed with mean $\mu \neq 0$ and variance 1. The tail probability is then given by

$$\mathbb{P}(L > x) = \mathbb{E}_{\mathbb{S}}\left[\mathbf{1}_{\{L>x\}} e^{-\mu Y + \mu^2/2}\right]. \tag{3.10}$$

This step will increase the likelihood of $\{L > x\}$, making a rare event less rare.

The idea of exponential twisting is to choose

$$q_{i,\theta(Y)}(Y) = \frac{p_i(Y)e^{\theta(Y)\omega_i}}{1 + p_i(Y)(e^{\theta(Y)\omega_i} - 1)}, \tag{3.11}$$

which increases the default probability if $\theta > 0$. This step will cluster the losses around x, which is particularly useful for the estimation of VaRC. With these two techniques the tail probability can be formulated as

$$\begin{aligned}
\mathbb{P}(L > x) &= \mathbb{E}\left\{\mathbb{E}_{\mathbb{Q}}\left[\mathbf{1}_{\{L>x\}} \prod_i \left(\frac{p_i(Y)}{q_i(Y)}\right)^{D_i} \left(\frac{1-p_i(Y)}{1-q_i(Y)}\right)^{1-D_i} \Big| Y\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}_{\mathbb{Q}}\left[\mathbf{1}_{\{L>x\}} e^{-\theta(Y)L + K(\theta(Y),Y)} \Big| Y\right]\right\} \\
&= \mathbb{E}_{\mathbb{S}}\left\{e^{-\mu Y + \mu^2/2}\mathbb{E}_{\mathbb{Q}}\left[\mathbf{1}_{\{L>x\}} e^{-\theta(Y)L + K(\theta(Y),Y)} \Big| Y\right]\right\}.
\end{aligned} \tag{3.12}$$

To find suitable parameters for the procedures of exponential twisting and mean shifting, Glasserman & Li (2005) and Glasserman (2006) propose to choose $\hat{\theta}(y)$ that solves

$$\mathcal{K}'(\theta(y), y) = x. \tag{3.13}$$

and $\mu = \arg_y \max \mathcal{K}(\hat{\theta}(y), y) - \hat{\theta}(y)x - \frac{1}{2}y^2$, where $\hat{\theta}(y)$ is given by (3.13) and the exponential of which to be maximized is the upper bound of $\mathbb{P}(L > x|Y = y)\exp(-y^2)$. Note that Eq. (3.13) is identical to Eq. (3.6) in the saddlepoint approximation since both methods employ the idea of an Esscher transform.

The estimation of the VaR contribution is trivial. It is given by

$$\text{VaRC}_i = \omega_i \frac{\sum_k D_i \ell^k \mathbf{1}_{\{L^k = \text{VaR}\}}}{\sum_k \ell^k \mathbf{1}_{\{L^k = \text{VaR}\}}},$$

where the superscript k denotes the k-th simulated scenario and ℓ is the likelihood ratio $e^{-\mu Y + \mu^2/2 - \theta(Y)L + \mathcal{K}(\theta(Y), Y)}$.

3.3 A stylized portfolio

To examine the performance of the methods described above, we first consider a stylized portfolio A consisting of 11,325 obligors which only differ in exposure size. They are categorized in 6 buckets and the exposure per obligor and the number of obligors in each bucket are the following,

bucket	1	2	3	4	5	6
Exposure	1	10	50	100	500	800
# of obligors	10000	1000	200	100	20	5

Asset correlation and PD are given by

$$\rho = 20\%, \quad p = 0.33\%. \tag{3.14}$$

The portfolio has a total exposure of 54000. It is a portfolio of so-called lower granularity since the largest obligor has an exposure 800 times larger than the smallest obligor. Exposure concentration is not really significant as the weight of the largest obligor is less than 1.5% of the total exposure.

Both the normal approximation and saddlepoint approximation calculate the tail probability instead of the VaR directly. The VaR can then be obtained by inverting the loss distribution. A not very sophisticated

iterative solver, the bisection method, is used here for this purpose. We search the VaR in the interval with as a lower bound the portfolio expected loss $E(L)$ and as an upper bound the total portfolio exposure. The two approaches also require the discretization of the common factor Y. In a one-factor setting, numerical integration methods rather than simulation should be used for efficient and accurate calculation of the unconditional loss density and tail probability. As is done in §2.3 we truncate the domain of Y to the interval $[-5, 5]$ and employ the (non-adaptive) Gaussian quadrature method. The speed of saddlepoint methods strongly depends on the number of abscissas N in the discretization of Y. Most of the CPU time is spent in finding the saddlepoints. The same holds for IS with exponential twisting. We find generally that $N = 100$ abscissas are sufficient in terms of accuracy for the saddlepoint methods, while for IS many more points are necessary to obtain an estimate with small variance. For the normal approximation we also adopt $N = 100$.

In the tables that follow "Vasicek" denotes the asymptotic approximation of the Vasicek model and "NA" denotes the normal approximation. The results given by the saddlepoint approximations are labeled by "SA". "IS-10K" stands for importance sampling with ten thousand scenarios. Its VaR estimate and the sample standard deviations are computed by subdividing the ten thousand scenarios into 10 equally-sized subsamples.

Table 3.1 presents both the $\text{VaR}_{99.9\%}$ and $\text{VaR}_{99.99\%}$ of the portfolio given by various methods. CPU times are in seconds. Monte Carlo simulation based on 10 subsamples with 16 million scenarios each serves as our benchmark. We also report on the standard deviation and the 95% confidence intervals (CI) beside the point estimates. The standard deviations of $\text{VaR}_{99.9\%}$ and $\text{VaR}_{99.99\%}$ are 7.7 (0.1% of the corresponding VaR) and 38.4 (0.5% of the corresponding VaR), respectively.

Even though the portfolio has no serious exposure concentration, the VaR estimates at both confidence levels obtained from the asymptotic Vasicek approximation are far from the benchmark VaR (relative errors around 5%). The normal approximation provides a significant improvement in accuracy with only little additional computational time. The relative errors for both VaR estimates are less than 1%. The saddlepoint approximation is even more accurate than the normal approximation while remaining fast. Both VaR estimates, which can be obtained in several seconds, fall within the 95% confidence interval and have relative errors less than 0.2%. The variance reduction of IS compared to plain simulation is especially effective in the far tail. With only one thousand scenarios in each subsample, the standard deviations of the VaR estimate are not really small. Although

the VaR estimates given by IS are comparable to those given by SA, IS is significantly more computational intensive.

	$\text{VaR}_{99.9\%}$	$\text{VaR}_{99.99\%}$	time
Benchmark	3960.3(7.7)	6851.6(38.4)	
95% CI	[3945.2, 3975.3]	[6776.3, 6926.9]	
Vasicek	3680.5	6477.0	8E-4
NA	3924	6804	2E-2
SA	3965	6841	6E+0
IS-10K	3975.3(56.4)	6836.8(84.9)	2E+3

TABLE 3.1: $\text{VaR}_{99.99\%}$ of portfolio A. The Benchmark and IS-10K sample standard deviations (in parentheses) are calculated using 10 simulated subsamples of 16 million and one thousand scenarios each, respectively.

Regarding the VaR contribution, we in fact compute the VaRC of an obligor scaled by its effective weight ω_i, i.e.,

$$\frac{\partial \text{VaR}_\alpha}{\partial \omega_i}(L) = \mathbb{P}(D_i = 1 | L = \text{VaR}).$$

This represents the VaRC of an obligor as a percentage of its own effective exposure. Expressed as a probability, it always lies in the interval $[0, 1]$.

VaRCs of the obligors in each bucket at loss level $L = 4000$ and $L = 6800$ are given in Table 3.2. The simulated portfolio loss L is so sparse in the vicinity of the VaR, that we have to replace the event $\{L = \text{VaR}\}$ by

$$\frac{|L - \text{VaR}|}{\text{VaR}} < \gamma \tag{3.15}$$

to make our VaRC estimates meaningful. (An alternative is to use the Harrell-Davis estimate, which computes a quantile estimate as a weighted average of multiple order statistics, as in Mausser & Rosen (2004).) We face a dilemma here. A small γ reduces bias but at the expense of having only very few useful scenarios. We here choose $\gamma = 0.5\%$ for $L = 4000$ and $\gamma = 1\%$ for $L = 6800$. The former event has a probability around 0.004% and the latter around 0.001%. Our benchmark VaRC estimates are both based on 12000 such events, resulting from roughly 300 million and 1200 million scenarios, respectively. The benchmark standard deviations (in parenthesis) and confidence intervals are computed by dividing the 12000 scenarios into ten equally-sized subsamples. For IS we simply use the same

γ as MC for both loss levels. There are 316 and 772 out of ten thousand IS scenarios, hence 3.16% and 7.72% respectively, for which L falls in the desired ranges. The effect of clustering losses around the level of interest by IS is truly significant compared to plain Monte Carlo simulation.

It appears that SA is the only method that is able to give all VaRC estimates within the 95% confidence interval. Its maximum absolute error of 0.33% is also the smallest among all methods. The estimates from SSA are similar to those with SA, especially for small exposures. In terms of speed SSA is about seven times faster than SA. At the same time it has 2 estimates outside the 95% confidence interval. The normal approximation and importance sampling have 7 and 5 estimates outside the 95% confidence interval, respectively. The differences to the benchmark for all the three methods are quite small though, with maxima 1.14% (SSA), 1.27% (NA) and 1.24% (IS). NA overestimates the VaRC of small exposures and underestimates the VaRC of large exposures, whereas SSA overestimates the VaRC of large exposures. A problem with IS is that the VaRCs are not monotonically increasing with the effective weight w, which is counterintuitive. From this perspective ten thousand scenarios do not seem enough.

It must be finally noted that the above observations on the performance of the various methods are not restricted to portfolios with uniform PD as we impose. As an example we vary the PDs of obligors in each bucket in portfolio A more realistically as follows,

bucket	1	2	3	4	5	6
PD	2.5%	1%	0.5%	0.33%	0.05%	0.01%

In Table 3.3 we report the estimated portfolio $\text{VaR}_{99.9\%}$. It turns out that the variation in the PDs among individual obligors has virtually no impact on the performance of a method. All the three methods other than the Vasicek formula again give satisfactory approximations. Further results on CPU time, $\text{VaR}_{99.99\%}$ and VaR contributions will not be shown as we did not find anything significantly different from the results for the original portfolio A.

3.4 Analysis of robustness

Both the normal approximation and the saddlepoint approximation are asymptotic approximations that become more accurate when the portfolio size increases. The normal approximation stems from the central limit theorem and uses merely the first two moments of the conditional portfolio

(a) VaRC at the loss level $L = 4000$

	VaRC_1	VaRC_2	VaRC_3	VaRC_4	VaRC_5	VaRC_6	time
BM	6.33(0.04)	6.38(0.05)	6.54(0.03)	6.86(0.08)	9.36(0.17)	11.32(0.38)	
95% CI	[6.25, 6.41]	[6.28, 6.48]	[6.49, 6.59]	[6.70, 7.02]	[9.02, 9.70]	[10.58, 12.06]	
Vasicek	7.41	7.41	7.41	7.41	7.41	7.41	3E-3
NA	6.55	6.59	6.78	7.02	8.92	10.35	1E-2
SA	6.35	6.39	6.58	6.82	9.21	11.65	2E+0
SSA	6.35	6.37	6.5	6.68	9.12	12.46	3E-1
IS-10K	6.54	6.46	6.77	6.7	9.4	10.33	1E+3

(b) VaRC at the loss level $L = 6800$

	VaRC_1	VaRC_2	VaRC_3	VaRC_4	VaRC_5	VaRC_6	time
BM	11.23(0.09)	11.29(0.09)	11.56(0.11)	11.87(0.12)	14.89(0.21)	17.86(0.59)	
95% CI	[11.06, 11.41]	[11.11, 11.48]	[11.35, 11.77]	[11.63, 12.11]	[14.48, 15.30]	[16.70, 19.03]	
Vasicek	12.59	12.59	12.59	12.59	12.59	12.59	3E-3
NA	11.42	11.48	11.74	12.06	14.65	16.59	1E-2
SA	11.23	11.29	11.55	11.88	14.94	17.78	2E+0
SSA	11.23	11.27	11.48	11.75	14.89	18.44	3E-1
IS-10K	11.34	11.52	11.62	12.03	14.83	16.62	1E+3

TABLE 3.2: VaRC of portfolio A at the loss levels $L = 4000$ and $L = 6800$. All the numbers except CPU time are in percentage terms. The benchmark (BM) sample standard deviations (in parentheses) are calculated using 10 simulated subsamples of 1,200 relevant scenarios each.

	Benchmark	Vasicek	NA	SA	IS-10K
$\text{VaR}_{99.9\%}$	5888(12.5)	5819	5882	5886	5871(63.6)
95% CI	[5863.5,5912.5]				

TABLE 3.3: $\text{VaR}_{99.9\%}$ of portfolio A with non-uniform PD ranging from 2.5% to 0.01%. The Benchmark and IS-10K sample standard deviations (in parentheses) are calculated using 10 simulated subsamples of 16 million and one thousand scenarios each, respectively.

loss $L(Y)$. Higher order approximations such as the Edgeworth expansion provide an improvement as they take the higher cumulants of $L(Y)$ into account. As for the saddlepoint approximations, the Daniels formula can be considered as a generalization of the Edgeworth expansion that makes use of the explicit knowledge of the moment generating function (see Jensen 1995). In this respect it is expected that the saddlepoint approximations are generally more accurate than the normal approximation, which is confirmed by our example above. A drawback is that the tail probability given by the Edgeworth expansion is not necessary in the range of $[0, 1]$ and is not always monotone. Similarly the quantile approximations are not always monotone in the probability levels (cf. Wallace 1958). The Lugannani-Rice formula may also suffer from the same problems. On the contrary, importance sampling/simulation always gives estimates to a probability in $[0, 1]$.

An important concern is whether the conditions of the central limit theorem hold if severe exposure concentration is present in a portfolio. Apparently if the conditions do not hold the normal approximation will fail. Let us now consider a portfolio B consisting of a bucket of 1000 obligors with effective exposure $w_1 = 1$ and one large obligor with effective exposure $w_2 = S$, $S \in \{20, 100\}$, i.e.,

bucket	1	2
Exposure	1	S, $S \in \{20, 100\}$
# of obligors	1000	1

For the other parameters ρ and PD we adopt (3.14). The weight of the large obligor relative to the total exposure is almost 2% when $S = 20$ and 10% when $S = 100$. The latter should be considered as serious exposure concentration. The Binomial Expansion Method (see §2.A), by which the VaR and VaRC can be computed almost exactly, will be used as the benchmark.

We consider the quantile $\alpha = 99.99\%$. Table 3.4 gives the VaR of portfolio B obtained by various methods. The approximation error of VaR is measured by the relative error (RE) defined as

$$\text{RE=(Estimate-Benchmark)/Benchmark.}$$

When $S = 20$, we see that all methods, except Vasicek, have relative errors of less than 2%. When S is increased to 100, both Vasicek and NA become erratic (relative errors $> 10\%$), whereas the effect of a large S on the accuracy of SA is marginal. We remark that we have tested for even larger S up to 1000 (50% of the total exposure of correponding portfolio), and SA manages to consistenly give $\text{VaR}_{99.99\%}$ estimates with $|\text{RE}| < 2\%$. IS is also insusceptible to the size of S. It is as accurate as SA, but demands much more CPU time.

	$S = 20$			$S = 100$		
	VaR	error	time	VaR	error	time
Exact	125			170		
Vasicek	122.3	-2.13%	6E-4	131.9	-22.39%	1E-3
NA	125	0.00%	1E-2	149	-12.35%	9E-3
SA	126	0.80%	3E+0	168	-1.18%	3E+0
IS-10K	124.1(1.7)	-0.72%	2E+2	170.5(3.1)	0.29%	2E+2

TABLE 3.4: $\text{VaR}_{99.99\%}$ of portfolio B. Errors reported are relative errors.

The reason why the normal approximation does not work for $S = 100$ is not difficult to explain. Conditional on the common factor Y, NA tries to approximate the loss density by a normal distribution (due to the central limit theorem). This works quite well when S is as large as 20. However, when we have $S = 100$, which is almost 10% of the total exposure, the loss density will no longer be unimodal. A normal approximation is not able to capture this pattern and therefore can be problematic. This is illustrated in Figure 3.1.

It is also worthwhile explaining how the exposure concentration is handled by the saddlepoint approximation. Therefore, instead of computing only a quantile of the portfolio loss, we calculate the whole loss distribution when $S = 100$ using our benchmark and the SA. This is demonstrated in Figure 3.2(a). We notice that the true loss distribution is not smooth in the vicinity of 100, which is precisely the size of the large exposure S.

Recall that the SA relies on the inversion formula (1.17) representing the tail probability. It is thus implicitly assumed that the portfolio loss L,

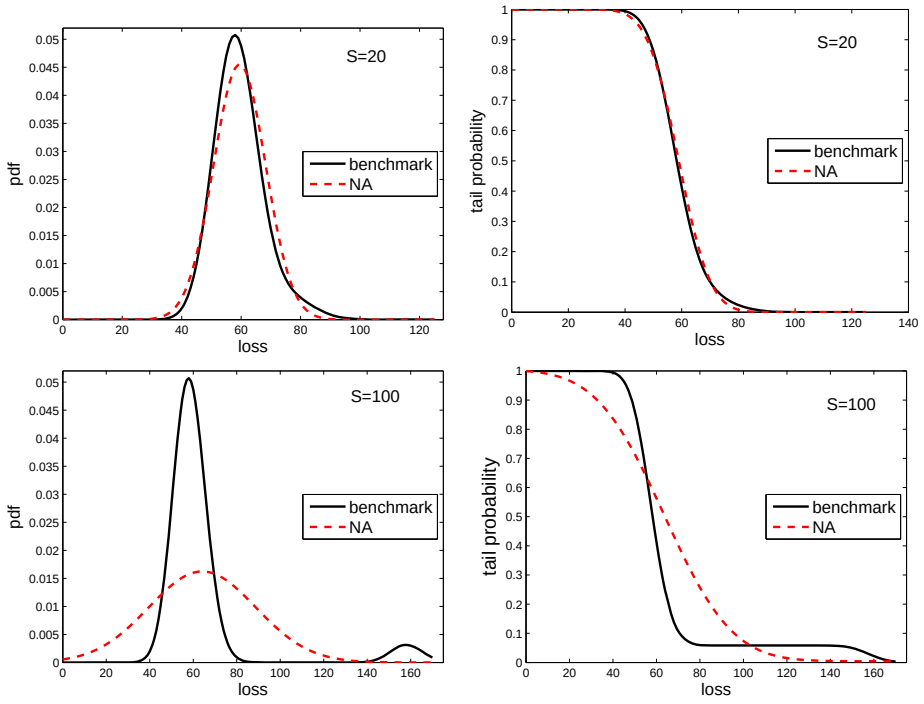


FIGURE 3.1: Loss density and tail probability of Portfolio B given by the normal approximation conditional on an arbitrarily chosen common factor Y.

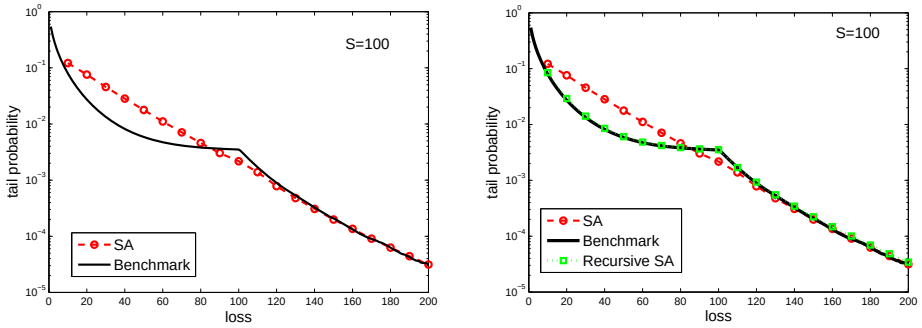
which is discrete when LGD is constant, can be well approximated by a continuous random variable which has an absolutely continuous cumulative distribution function. The saddlepoint method thus produces a smoothed version of the loss distribution. A more detailed discussion of the saddlepoint approximations as smoothers is in Davison & Wang (2002). We see in Figure 3.2(a) that the saddlepoint approximation to the tail probabilities is incorrect for almost all quantiles preceding the point of non-smoothness (around the 99.6%-quantile) but is again accurate for higher quantiles. It entails that, with one or a few exceptional exposures in the portfolio, a uniform accuracy of the loss distribution may not be achieved by a straightforward saddlepoint approximation. This can be a problem if the quantile we are interested in precedes the non-smoothness in the loss distribution, which usually occurs at the size of large exposures.

A very easy algorithm can be used to retain the uniform accuracy. Suppose a portfolio has m large exposures $S_i, i = 1, \ldots, m$ with $S_1 \leq S_2 \leq \cdots \leq S_m$. For any loss level $x < S_k$ the tail probability conditional on Y

can be written as

$$\mathbb{P}(L > x|Y) = 1 - \mathbb{P}(L - \sum_{i \geq k} L_i \leq x|Y) \prod_{i \geq k} \mathbb{P}(D_k = 0|Y). \tag{3.16}$$

The above reformulation takes into account the implicit information that when $L < x$ the obligors with exposure larger than x must not default. An application of SA to the probability $\mathbb{P}(L - \sum_{i \geq k} L_i > x|Y)$ rather than directly to $\mathbb{P}(L > x|Y)$ furthermore removes the exceptional exposure concentration $S_i, i \geq k$. It is apparently more accurate than a direct SA to $\mathbb{P}(L > x|Y)$. A similar idea is discussed in Beran & Ocker (2005). We call this method the *adaptive saddlepoint approximation* (ASA) here. As an experiment we apply the ASA to portfolio B with $S = 100$ and plot in Figure 3.2(b) the loss distribution for loss levels up to but excluding $L = S$ (in the estimation of the tail probabilities the ASA only differs from a direct SA for loss levels $L < S$.). The loss distribution given by the ASA matches the benchmark almost exactly for all $L < S$.



(a) SA (b) ASA

FIGURE 3.2: The loss distribution of portfolio B given by the saddlepoint approximation (SA) and adaptive saddlepoint approximation (ASA) when the loss distribution is not smooth at the vicinity of S. PD= 0.0033, $\rho = 0.2$, $S = 100$.

Now we consider the VaR contribution. Table 3.5 presents the VaRC of both a small obligor (VaRC_1) and the large obligor (VaRC_2). For the cases $S = 20$ and $S = 100$, we report four estimates to the VaRC given by each method. The error we report here is absolute error. NA gives fair VaRC estimates for both VaRC_1 and VaRC_2 when $S = 20$ but deviates dramatically from the benchmark when $S = 100$. This is in line with its performance on the VaR estimation. SA is quite accurate for VaRC_1 but becomes less accurate for VaRC_2 as S increases. SSA resembles SA in the estimates of

VaRC_1 but does not give satisfactory estimates to VaRC_2 at all: both errors are larger than 5%. This can be understood by the fact that, as mentioned in §3.2.3, the solutions to (3.6) and (3.7) are indeed close for small exposures but can differ substantially for large exposures. Further experiments show that NA, SA and SSA may all give VaRC values that are not in the interval $[0, 1]$ in the presence of more exceptional exposure concentrations (as is pointed out at the beginning of this section). IS appears to be the best method in terms of accuracy and robustness in this case.

(a) $S = 20$

$S = 20$	VaRC_1	error	VaRC_2	error	time
Benchmark	12.06		21.78		
Vasicek	12.25	0.19	12.25	-9.53	3E-3
NA	12.12	0.06	18.94	-2.84	1E-2
SA	12.05	-0.01	21.70	-0.08	8E-1
SSA	11.96	-0.10	27.06	5.28	3E-1
IS-10K	12.04	-0.02	22.89	1.11	1E+3

(b) $S = 100$

$S = 100$	VaRC_1	error	VaRC_2	error	time
Benchmark	8.29		87.07		
Vasicek	15.45	7.16	15.45	-71.62	3E-3
NA	12.68	4.39	43.18	-43.89	2E-1
SA	8.89	0.60	90.79	3.72	8E-1
SSA	9.15	0.86	78.52	-8.55	3E-1
IS-10K	8.12	-0.17	88.85	1.78	1E+3

TABLE 3.5: $\text{VaRC}_{99.99\%}$ of portfolio B. All the numbers except CPU time are in percentage terms. Errors reported are absolute errors.

In both portfolios A and B importance sampling seems to perform fine for determining VaRC. The reason for this is that the obligors in a bucket are considered identical and we are able to take the average of all obligors in the same bucket when estimating VaRC. This makes the simulated VaRC estimates much less volatile. We must point out that even though IS is able to cluster the simulated losses around the VaR of interest and thus significantly increases the probability $\mathbb{P}(L = \text{VaR})$, a rather large number of replications are still necessary.

Let us consider a portfolio C consisting of 100 obligors with exposures all different from each other such that

$$\omega_i = i, \quad i = 1, \ldots, 100. \tag{3.17}$$

The parameters ρ and PD are again the same as in (3.14).

Figure 3.3 gives scatterplots of the (scaled) VaRC (y-axis) at the loss level $L = 700$, which is around $\text{VaR}_{99.99\%}$, against the EAD (x-axis). In the left-side figure we show the results given by the saddlepoint approximation, the simplified saddlepoint approximation and the normal approximation. All methods clearly show that the VaRC increases as the EAD increases, which is highly desirable for practical purposes. SSA again gives results very close to the saddlepoint approximation. Compared to the SA, the NA overestimates the VaRC of small exposures and underestimates the VaRC of large exposures. This is consistent with the pattern shown in Portfolio A. The estimates given by importance sampling with ten thousand scenarios (IS-10K) and one hundred thousand scenarios (IS-100K) are presented in the right-side of Figure 3.3 along with those given by SA. γ as in (3.15) is set to be 1%. The relation between the VaRC and EAD is not clear at all with only ten thousand simulated scenarios. The estimates, resulting from 256 relevant scenarios, disperse all over the area. Improvement in the performance of the VaRC estimation is discernable when we increase the number of scenarios of IS by ten times. The VaRC estimates are then based on 2484 relevant scenarios and the upward trend of VaRC with increasing EAD is evident. However due to simulation noise the curve remains highly oscillatory and an even higher number of scenarios seems necessary.

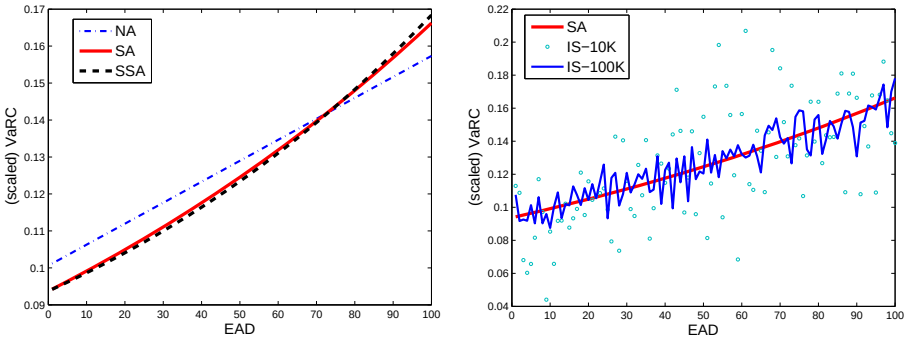


FIGURE 3.3: VaR contribution of Portfolio C as a function to EAD at the loss level $L = 700$.

3.5 Conclusions and discussions

We have examined various numerical methods for the purpose of calculating the credit portfolio VaR and VaRC under the Vasicek one-factor model. Each method provides a viable solution to VaR/VaRC estimation for lower granular portfolios and portfolios with medium exposure concentration. However there is no perfect method that prevails under all circumstances and the choice of preferred method turns out to be a trade-off among speed, accuracy and robustness.

The normal approximation is the fastest method and is able to achieve a fair accuracy. It is however rather vulnerable because it is incapable of handling portfolios dominated by one or a few obligors (or, portfolios with multi-modal loss density). The simplified saddlepoint approximation is second to the normal approximation in speed and may suffer from the same problem when estimating the VaRC.

Importance sampling does not guarantee to be the most accurate method but it always works fine provided a sufficient number of scenarios are drawn. It makes no assumption on the composition of a portfolio and thus is certainly the best choice from the perspective of robustness. Unlike the other methods, it always gives estimates to the scaled VaRC in $[0, 1]$. The downside of IS is that it is rather time-consuming when compared to the other methods. Moreover IS is not strong in the estimation of VaRC, which is really demanding in the number of simulated scenarios.

The saddlepoint approximation is generally more accurate than the normal approximation and the simplified saddlepoint approximation. It is also more reliable in the sense that it can handle more extreme exposure concentration. Consequently it may well serve as a fast alternative to IS with a good balance between accuracy and speed. It must be emphasized that, if the loss distribution is not smooth due to exceptional exposure concentrations and the target quantile precedes the non-smoothness in the loss distribution, a straightforward implementation of SA is likely to be insufficient. The adaptive saddlepoint approximation should be employed in this situation.

We would like to point out again that the normal approximation and the saddlepoint approximation methods are all based on asymptotic approximations. They become more accurate when the portfolio size increases. On the other hand, importance sampling becomes substantially more demanding in computation time when the portfolio size increases.

Although we mainly concentrate on the VaR-based risk contribution, we would like to point out that all the four methods evaluated can be readily

extended to compute the risk contribution with respect to the Expected Shortfall (ES). A thorough discussion on the ES and ES contribution can be found in Acerbi & Tasche (2002), Tasche (2002). The estimation of ES contributions by importance sampling is developed in Glasserman (2006). It is shown that importance sampling is equally effective for the estimation of the ES contributions as for the VaR contributions. The saddlepoint approximation to the ES and ES contributions has been discussed in Chapter 2. A numerical experiment therein shows that IS and SA give comparable results for ES contributions. Formulas for the normal approximation to the ES and ES contributions are derived in Martin (2004). The approximations are likely to be satisfactory when the normal approximations to the tail probability and VaR contributions work well, as all approximations hinge on the central limit theorem.

A final remark is that it is straightforward to extend the use of the four methods to multi-factor models. The only problem with multi-factor models is that the efficiency of the normal/saddlepoint approximation can no longer be maintained: due to the curse of dimensionality the Gaussian quadrature rule becomes impractical as the number of factors increases. It is therefore even desirable to combine importance sampling and the other methods, as was a hybrid method of importance sampling and saddlepoint approximation presented in §2.4.2. Another efficient high-dimensional integration method is proposed in Chapter 5.

Chapter 4

Generalized Beta Regression Models for Random Loss Given Default

4.1 Introduction

In the context of credit portfolio losses, the quantity Loss-Given-Default (LGD) is the proportion of the exposure that will be lost if a default occurs. The uncertainty about the actual LGD constitutes an important source of the credit portfolio risk in addition to the default risk. In practice, e.g., in both CreditMetrics (Gupton et al. 1997) and KMV Portfolio Manager (Gupton & Stein 2002), the uncertainty in the LGD rates of defaulted obligors is assumed to be a beta random variable independent for each obligor. The beta distribution is well-known to be very flexible, modeling quantities constrained in the interval $[0, 1]$. Depending on the choice of parameters, the probability density function can be unimodal, U-shaped, J-shaped or uniform.

However, extensive empirical evidence, see e.g., Hu & Perraudin (2002), Altman et al. (2005), has shown that this simple approach is insufficient. It is now well understood that LGD is positively correlated to the default rate, in other words, LGD is high when the default rate is high, which suggests that there is also systematic risk in LGD, just like in the default rates. A heuristic justification is that the LGD is determined by the collateral value which is sensitive to the state of the economy.

Based on results of a non-parametric estimation procedure, Hu & Perraudin (2002) further showed that without taking into account the PD/LGD

correlation the economic capital, or Value at Risk (VaR), of a loan portfolio can be significantly underestimated. This has a critical consequence for risk management practice. In the Basel II Accord this issue is addressd by the notion of "downturn LGD".

The insight of LGD being subject to systematic risk dates back to Frye (2000), in which the LGD is modeled by a normal distribution. An obvious problem with this model is that it allows the LGD to be negative which cannot be the case. To ensure the nonnegativity of LGD, Pykhtin (2003) employs a truncated log-normal distribution for the LGD. Andersen & Sidenius (2004) propose the use of a probit transform of the LGD such that the transformed LGD is normally distributed. The probit transformation guarantees that the LGD stays in the interval $[0, 1]$. In a similar manner Düllmann & Trapp (2004), Rösch & Scheule (2005) employ a logit transform of the LGD. Rather different from the above approaches, Giese (2006) and Bruche & González-Aguado (2008) extend the static beta distribution assumption in CreditMetrics and KMV Portfolio Manager by modeling the LGD as a mixture of beta distributions that depend on the systematic risk.

In this chapter we propose a *Generalized Beta Regression* (GBR) *framework* to model the Loss-Given-Default. This framework generalizes the Beta Regression model proposed by Ferrari & Cribari-Neto (2004) and is very similar to a class of models derived from Generalized Linear Models (GLM). Our models are called Generalized Beta Regression Models since the LGD is always assumed to be (conditionally) beta distributed. The models by Giese (2006), Bruche & González-Aguado (2008) can be regarded as special examples in our GBR framework. The quantities in our models have simple economic interpretation as the quantity and quality of the LGD. In contrast with the transformed LGD models, GBR models do not require *normality* and *homoscedasticity*. Inference in this framework can be unified for models with a variety of link functions and different degrees of complexity with least squares method and maximum likelihood estimation, making model selection a straightforward task. Moreover, the GBR framework allows both the normal approximation and the saddlepoint approximation to efficiently calculate the portfolio loss distribution. This is the first time that the numerical approximation methods have been used successfully to calculate portfolio loss distribution in the presence of random LGD.

The rest of the chapter is organized as follows. In section 2 we introduce the Vasicek's Gaussian one-factor model as the default model and give a brief summary of existing random LGD models. Section 3 elaborates on the GBR framework including the basic Beta regression model and two extensions. In Section 4 we discuss methods for parameter estimation and

provide a calibration example. Section 5 explains techniques for efficient loss distribution approximation in the GBR framework.

4.2 Existing LGD models

A variety of models in which LGD is subject to systematic risk can be found in the literature. Within a one-factor framework, Frye (2000) proposed a model in which the LGD is normally distributed and influenced by the same systematic factor Y that drives the PD, so that

$$\begin{aligned}\Lambda &= \mu + \sigma\xi, \\ \xi &= -\sqrt{\tilde{\rho}}Y + \sqrt{1-\tilde{\rho}}\epsilon,\end{aligned}$$

where ξ and ϵ are both standard normally distributed. The minus sign in front of $\sqrt{\tilde{\rho}}$ reflects the empirical findings that LGD tends to be higher when the economy is weak and lower when the economy is strong. This way the dependence between LGDs and the dependence between PD and LGD are modeled simultaneously. The parameters μ and σ can be understood to be the expected LGD and the LGD volatility, respectively. Unfortunately, the LGD is unbounded in $\mathbb{R}$ and can thus be negative. To ensure the nonnegativity of LGD, Pykhtin (2003) employs a log-normal distribution for the LGD,

$$\Lambda = \left(1 - e^{\mu+\sigma\xi}\right)^+.$$

Other extensions include Andersen & Sidenius (2004), choosing a probit transformation

$$\Lambda = \Phi(\mu + \sigma\xi),$$

where Φ is again the cdf of the standard normal distribution and Düllmann & Trapp (2004), Rösch & Scheule (2005) that employ a logit transformation

$$\Lambda = \frac{1}{1+\exp(\mu+\sigma\xi)}.$$

All three transformations for the LGD above guarantee that the LGD lies in the interval $[0,1]$. However the parameters μ and σ do not have a convenient economic interpretation as in Frye's model.

The above models basically are all linear models of the transformed LGD in the form of

$$g(\Lambda(Y)) = \mu - \sigma\sqrt{\tilde{\rho}}Y + \sigma\sqrt{1-\tilde{\rho}}\epsilon \tag{4.1}$$

so that $g(\Lambda(Y))$ is normally distributed with mean $\mu + \sigma\sqrt{\tilde{\rho}}Y$ and variance $\sigma^2(1-\tilde{\rho})$. Hence the transformed LGD $g(\Lambda(Y))$ is required to be symmetric and homoscedastic, i.e., its variance must not vary with the mean. This is in contradiction with an empirical study in Düllmann & Trapp (2004), at least for the Pykhtin (2003) model, where the Shapiro-Wilk test for normality to $\log(1-\Lambda)$ gives a p-value of 0.05.

A more flexible approach extends the static beta distribution assumption as it is present in CreditMetrics and KMV Portfolio Manager. Giese (2006), Bruche & González-Aguado (2008) model the LGD by a mixture of beta distributions

$$\Lambda \sim Beta(\alpha, \beta)$$

where both α and β are functions of common factor Y. However, α and β are both shape parameters and an economic interpretation of such models is very difficult.

We here propose the Generalized Beta Regression (GBR) framework for random LGD. The GBR framework includes Giese (2006) and Bruche & González-Aguado (2008) as special examples but calls for a different parameterization of the beta distribution. The class of models is very flexible and the quantities in our models have an easy economic interpretation as the quantity and quality of the LGD. Inference of models in this framework can be unified. Compared to the transformed LGD models given by (4.1), the GBR models accommodate better skewness and heteroscedastic errors.

4.3 Generalized Beta Regression Models

4.3.1 Parameterization of a beta distribution

Recall that the probability density function of a beta distribution with parameters $\alpha > 0$, $\beta > 0$ reads

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $B(\cdot,\cdot)$ denotes the beta function and $\Gamma(\cdot)$ the gamma function.

The beta distribution is well-known to be very flexible, modeling quantities constrained in the interval $[0,1]$. Depending on the choice of parameters, the probability density function can be unimodal, U-shaped, J-shaped or uniform. The expectation and variance of a beta distributed variable X are given by

$$\mu = \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}, \tag{4.2}$$

$$\sigma^2 = \mathbb{V}ar[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\mu(1-\mu)}{\alpha+\beta+1}. \tag{4.3}$$

Note here that the variance is permitted to vary with its mean. Let $\varphi = \alpha + \beta$, then φ can be regarded as a *dispersion parameter* in the sense that, for a given μ, the variance is determined by the size of φ.

The parameters α and β can be formulated in terms of the mean and dispersion in the following way

$$\alpha = \mu\varphi, \quad \beta = (1-\mu)\varphi. \tag{4.4}$$

Therefore, a beta distribution can also be uniquely determined by its mean and dispersion.

4.3.2 Beta Regression Model

The Generalized Beta Regression framework proposed here is characterized by the following elements,

1. the LGD is assumed to be beta distributed, conditional on some covariates,

2. the beta distribution is parameterized by its mean and dispersion, rather than its natural parameters (α, β). The parameters mean and dispersion carry the economic interpretation as the quantity and quality of the LGD, respectively.

This framework generalizes the Beta Regression model proposed by Ferrari & Cribari-Neto (2004) for modeling rates and proportions. The models from the GBR framework are similar to a class of models derived from Generalized Linear Models (GLM).

The Generalized Linear Models have been developed since the seminal paper Nelder & Wedderburn (1972) as an extension to the classical linear regression models. In a GLM, the response variable X is in the exponential family. Its density can be represented in the form

$$f(x;\theta,\varphi) = e^{a(\varphi)[x\theta - b(\theta)] + c(\varphi,x)}, \tag{4.5}$$

For a comprehensive exposition of GLM we refer to McCullagh & Nelder (1989).

We start the explanation of our GBR framework with the basic Beta Regression model proposed in Ferrari & Cribari-Neto (2004). This basic approach only models the mean μ and treats the dispersion parameter φ as

a nuisance parameter. With some abuse of language we also call the model for the mean, μ, a GLM (although the probability density function of the beta distribution cannot be written in the form (4.5) and therefore it does not fit in the framework of GLM). The model for the mean in the GBR framework has the following two components:

- a *linear predictor* η

$$\eta = a\zeta \tag{4.6}$$

  where ζ is a vector of explanatory variables and a is a vector of the corresponding regression coefficients. As convention the first element of ζ is set to be 1, so that the first element of a is an intercept term.

- a monotonic, differentiable *link function* g

$$g(\mu) = \eta, \tag{4.7}$$

  where $\mu = E[\Lambda]$.

Potential covariates in the linear predictor can be seniority, collateral, type of industry and timing of business cycle. According to Schuermann (2004), these factors drive significant differences on LGD. Meanwhile, the inverse of the link function, $g^{-1}(\cdot)$, should form a mapping from $\mathbb{R}$ to $[0, 1]$, which is exactly the range of μ. This can be achieved by a variety of link functions, such as the logit link

$$\mu = \frac{e^{\eta}}{1 + e^{\eta}},\ \eta = \log\left(\frac{\mu}{1-\mu}\right), \tag{4.8}$$

or the probit link

$$\mu = \Phi(\eta),\ \eta = \Phi^{-1}(\mu). \tag{4.9}$$

Both the logit and probit link functions have a symmetric form about $\mu = 1/2$. If however it is believed that symmetric links are not justified, asymmetric link functions like the scaled probit link and the complementary log-log link can be used instead.

A remark is that our model can be very different from the transformed models characterized by (4.1) as we take $g(\mathbb{E}[\Lambda])$ to be linear to the covariates, rather than $\mathbb{E}[g(\Lambda)]$.

The most parsimonious model for LGD subject to systematic risk is a one-factor model with $\zeta = [1, Y]^T$, where Y is the common factor that also drives the default process. An example of such a model is given in Giese (2005), where the mean is modeled by

$$\mu = 1 - a_0\left(1 - p_i(Y)^{a_1}\right)^{a_2}, \tag{4.10}$$

and φ is considered a nuisance parameter.

Another special case is the static beta distribution model adopted by CreditMetrics and KMV, which is a degenerated version of the Beta Regression model, in which the coefficient in front of Y equals 0.

4.3.3 Extensions

The basic beta regression model above can be readily extended in various ways. One extension is to model the mean and dispersion jointly, rather than treating the dispersion parameter φ as a nuisance parameter which is either fixed or known. This is in the same spirit as the Joint Generalized Linear Model (JGLM) from the GLM framework, see e.g., Nelder & Lee (1991), Lee & Nelder (1998).

The dispersion φ can be modeled by a separate GLM,

$$h(\varphi) = b\zeta,$$

where h is also a link function. A simple way to ensure $\varphi > 0$ is to use a log link so that

$$\varphi = e^{b\zeta}. \tag{4.11}$$

A model of this type, but using a different version of the dispersion parameter, has been suggested in Bruche & González-Aguado (2008). They employ the following log-linear model for the two parameters α and β,

$$\alpha = e^{c\zeta},\ \beta = e^{d\zeta}, \tag{4.12}$$

where, as usual, ζ is a vector of covariates and c and d are vector coefficients. This specification is chosen, however, only to ensure positivity of both α and β. Besides, α and β are both shape parameters, and an economic interpretation of such a model is very difficult. In this regard, we note that by substituting (4.12) into (4.2) we obtain

$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{e^{(c-d)\zeta}}{1+e^{(c-d)\zeta}},$$

which is simply a logit model with vector coefficient $c-d$. The variance is then given by

$$\sigma^2 = \frac{\mu(1-\mu)}{\alpha+\beta+1} = \frac{\mu^2(1-\mu)}{\alpha+\mu},$$

so that the following dispersion parameter is adopted,

$$\varphi = \alpha = e^{c\zeta}.$$

A second extension is that the mean parameter μ can be modeled by a Generalized Linear Mixed Model (GLMM). GLMM extends GLM by adding normally distributed random effects in the linear predictor η. The simplest mixed model is the *random intercept model*

$$g(\mu) = \eta = a\zeta + \nu, \tag{4.13}$$

where, in addition to the fixed effect $a\zeta$, η also has a single component of random effect ν that follows a univariate normal distribution $N(0, \sigma_\nu^2)$. In our setting ν can be thought of as a latent common factor for the LGD independent of the fixed effects and default as well.

Such a GLMM, along with the probit link (4.9), is employed to model the mean LGD in Hillebrand (2006). Other applications of GLMM for portfolio credit default and migration risk can be found in McNeil & Wendin (2006, 2007).

Note that the two extensions above can be readily combined to form a new model that jointly models the mean and dispersion by means of GLMMs, i.e., fixed and random effects can be included in the modeling of both mean and dispersion. Further extensions are possible, e.g., replacing the linear predictor by a Generalized Additive Model (GAM), see Hastie & Tibshirani (1990), or adding multi-level random effects in the GLMM.

4.4 Estimation

In this section, we discuss the parameter estimation in the GBR framework by

1. least squares,

2. maximum likelihood estimation (MLE).

The former requires only the knowledge of annual mean LGD and LGD volatility and can be used as a first approximation to the MLE.

Suppose we have a time series of LGD data for $\mathtt{T} \in \mathbb{N}$ years. Let $\mathtt{K_t}$ be the number of defaulted obligors in year $\mathtt{t}$ and $\lambda_{\mathtt{t,k}}$ be the observed LGD for defaulted obligor $\mathtt{k}$, $\mathtt{t} = 1, \ldots, \mathtt{T}$, $\mathtt{k} = 1, \ldots, \mathtt{K_t}$. Each year, a realization of the common factor $Y_\mathtt{t}$ can be inferred from the default model and historical default data. The value of $Y_\mathtt{t}, \mathtt{t} = 1, \ldots, \mathtt{T}$ should be considered a known fixed effect in the LGD model.

From now on we call the three models discussed above in the GBR framework GBR-GLM, GBR-JGLM and GBR-GLMM, respectively. The

parameters to be estimated are: $\{a, \varphi\}$ in GBR-GLM, $\{a, b\}$ in GBR-JGLM or $\{a, \varphi, \sigma_\nu\}$ in GBR-GLMM, where a represents the vector coefficients in the linear predictor (4.6), b the vector coefficients in the linear predictor (4.11), φ is the dispersion parameter and σ_ν^2 is the variance of the random effect ν in (4.13).

4.4.1 Least squares

The method of least squares we propose here only requires the knowledge of the yearly mean LGD and LGD volatility for parameter estimation. The estimates of the yearly mean LGD and LGD volatility for $\mathtt{t} = 1, \ldots, \mathtt{T}$ can be obtained by matching the first and second moments of the LGD realizations $\lambda_{\mathtt{t},\mathtt{k}}$ such that

$$m_\mathtt{t} = \frac{1}{\mathtt{K_t}} \sum_{\mathtt{k}=1}^{\mathtt{K_t}} \lambda_{\mathtt{t},\mathtt{k}}, \quad \sigma_t^2 = \frac{1}{\mathtt{K_t}} \sum_{\mathtt{k}=1}^{\mathtt{K_t}} \lambda_{\mathtt{t},\mathtt{k}}^2 - m_\mathtt{t}^2.$$

Estimation of a and μ

The estimate for parameter a can be obtained by employing a linear regression of the transformed mean LGD $g(m_\mathtt{t})$ on $Y_\mathtt{t}$ and the other covariates,

$$g(m_\mathtt{t}) = \hat{a}\zeta_\mathtt{t} + \nu_\mathtt{t}, \tag{4.14}$$

where $\nu_\mathtt{t}$ is the residual term. In the GBR-GLM and GBR-JGLM

$$\hat{\mu}_\mathtt{t} = g^{-1}(\hat{a}\zeta_\mathtt{t}). \tag{4.15}$$

And in the GBR-GLMM $\nu_\mathtt{t}$ is taken to be the realized random effect in year $\mathtt{t}$ so that

$$\hat{\mu}_\mathtt{t} = g^{-1}(\hat{a}\zeta_\mathtt{t} + \nu_\mathtt{t}) = m_\mathtt{t}. \tag{4.16}$$

Estimation of b or φ

The estimation of the parameters b or φ takes the prediction of $\hat{\mu}_\mathtt{t}$, produced by (4.15) or (4.16), as an input. From (4.3), we obtain

$$\varphi_\mathtt{t} = \frac{\hat{\mu}_\mathtt{t}(1 - \hat{\mu}_\mathtt{t})}{\sigma_\mathtt{t}^2} - 1.$$

In both the GBR-GLM and GBR-GLMM the dispersion parameter φ is treated as a nuisance parameter. Its method-of-moments estimator is simply

$$\hat{\varphi} = \frac{1}{\mathtt{T}} \sum_{\mathtt{t}=1}^{\mathtt{T}} \varphi_\mathtt{t}.$$

In the GBR-JGLM, the coefficient b can be calculated by a linear regression of the transformed dispersion $h(\varphi_{\mathtt{t}})$ on covariate vector ζ such that

$$h(\varphi_{\mathtt{t}}) = \hat{b}\zeta_{\mathtt{t}} + \epsilon_{\mathtt{t}}.$$

Estimation of σ_ν in GBR-GLMM

The moment based estimate for σ_ν^2 is given by

$$\hat{\sigma}_\nu^2 = \frac{1}{\mathtt{T}} \sum_{\mathtt{t}=1}^{\mathtt{T}} \nu_{\mathtt{t}}^2,$$

where $\nu_{\mathtt{t}}$ is the residual term in (4.14).

4.4.2 Maximum likelihood estimation

Parameter estimation by the method of maximum likelihood is also straightforward in the GBR framework. In the models without random effects, i.e., GBR-GLM and GBR-JGLM, the log-likelihood function to be maximized reads

$$\begin{aligned}\ell(\mu,\varphi) = \sum_{\mathtt{t}=1}^{\mathtt{T}} \sum_{\mathtt{k}=1}^{\mathtt{K_t}} \{&(\mu_{\mathtt{t}}\varphi_{\mathtt{t}} - 1)\log(\lambda_{\mathtt{t,k}}) + [(1-\mu_{\mathtt{t}})\varphi_{\mathtt{t}} - 1]\log(1-\lambda_{\mathtt{t,k}}) + \\ &+ \log\Gamma(\varphi_{\mathtt{t}}) - \log\Gamma(\mu_{\mathtt{t}}\varphi_{\mathtt{t}}) - \log\Gamma[(1-\mu_{\mathtt{t}})\varphi_{\mathtt{t}}]\} \qquad (4.17)\end{aligned}$$

The score function, the gradient of the log-likelihood function and the Fisher information matrix, i.e., the variance of the score, can be formulated explicitly in terms of polygamma functions. They are given in Appendix 4.A. Asymptotic standard errors of the maximum likelihood estimates of the parameters can be computed from the Fisher information matrix.

Since the corresponding estimating equations do not admit a closed form solution, numerical maximization of the log-likelihood is necessary. Estimates by the method of least squares may be used as the initial approximations to the solutions of the likelihood equations.

We remark that the maximum likelihood estimation in Ferrari & Cribari-Neto's Beta Regression Model is already implemented in the statistical computing software R (www.r-project.org) in package 'betareg' so that it can be used immediately.

Marginal likelihood in GBR-GLMM

With the presence of random effects, the samples are no longer independent. In the random intercept model (4.13), the LGDs in year $\mathtt{t}$ are only independent conditional on the random effect $\nu_{\mathtt{t}}$. Since we are only interested in inference of the variance of the random component ν, but not in its realizations, the random effect needs to be integrated out. Therefore we maximize the marginal log-likelihood,

$$\ell_m(a, \varphi, \sigma_\nu) = \sum_{\mathtt{t}=1}^{\mathtt{T}} \log \left(\int \prod_{\mathtt{k}=1}^{\mathtt{K_t}} L(a, \varphi, \zeta_{\mathtt{t}}, \nu_{\mathtt{t}}; \lambda_{\mathtt{t,k}}) p_{\sigma_\nu}(\nu_{\mathtt{t}}) \mathrm{d}\nu_{\mathtt{t}} \right)$$

where $p_{\sigma_\nu}(\cdot)$ is the $p.d.f.$ of a normal distribution with mean 0 and variance σ_ν^2, and $L(\ ; \lambda_{\mathtt{t,k}})$ is the likelihood of $\{\text{LGD} = \lambda_{\mathtt{t,k}}\}$ given $\nu_{\mathtt{t}}$. The integral can be efficiently evaluated by Gaussian quadrature. Alternatively, the marginal likelihood can be approximated analytically by the use of the Laplace approximation to the integral, such as the penalized quasi-likelihood (PQL) estimation (Breslow & Clayton 1993) and the h-likelihood (Lee & Nelder 2001), thus avoiding numerical integration.

Finally we note that the likelihood ratio test based on large sample inference can be employed for model selection. Information criteria such as Akaike's information criterion (AIC) or the Bayesian information criterion (BIC) can also be used.

4.4.3 A simulation study

We show in this section how the models in the GBR framework can be calibrated and how model selection can be dealt with. Our focus is not to identify possible covariates that influence the LGD, however. Our estimation is based on data from Bruche & González-Aguado (2008), which are extracted from the Altman-NYU Salomon Center Corporate Bond Default Master Database and give the annual default frequency, number of defaults, mean LGD and LGD volatility for a period of 24 years (1982 - 2005). For completeness the data are reproduced in Appendix 4.B.

Estimation results

First, we fit the Vasicek default model. We assume that, across the years, the number of obligors is sufficiently large and all obligors in the portfolio have the same probability of default p and asset correlation ρ. Denote by

p_t the annual default frequency. We take the MLE's for ρ and p according to Düllmann & Trapp (2004),

$$\rho = \frac{\mathbb{V}ar\left[\Phi^{-1}(p_t)\right]}{1+\mathbb{V}ar\left[\Phi^{-1}(p_t)\right]}, \quad p = \Phi\left(\frac{\sum_{t=1}^{T}\Phi^{-1}(p_t)}{T\sqrt{1+\mathbb{V}ar\left[\Phi^{-1}(p_t)\right]}}\right),$$

where $\mathbb{V}ar[\delta] = \frac{1}{T}\sum_{t=1}^{T}\delta_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\delta_t\right)^2$. This yields

$$\rho = 0.0569, \quad p = 0.0153. \tag{4.18}$$

The common factor Y_t for year t, assumed to be independence from year to year, can be estimated as follows,

$$Y_t = \frac{\Phi^{-1}(p) - \sqrt{1-\rho}\Phi^{-1}(p_t)}{\sqrt{\rho}}. \tag{4.19}$$

Before we move on to the LGD model, we run a brief preliminary graphical check. In Figure 4.1(a) we show the yearly average default rate and yearly mean LGD for the years 1982-2005, from which the correlation between PD and LGD is evident. Figure 4.1(b) presents a scatterplot of the common factor Y estimated by eq. (4.19) versus yearly mean LGD. This figure suggests that the common factor Y, which drives the default, may as well be an important risk factor for LGD.

Next we make inferences about the LGD in our GBR framework with both the least squares and MLE. The LGD models we consider include only one covariate, which is the common factor Y in the default model. In light of the observation from Figure 4.1(b), this may be a reasonable choice. The mean LGD is fitted using a logit link

$$\mu = \frac{e^{a_1+a_2Y}}{1+e^{a_1+a_2Y}} \tag{4.20}$$

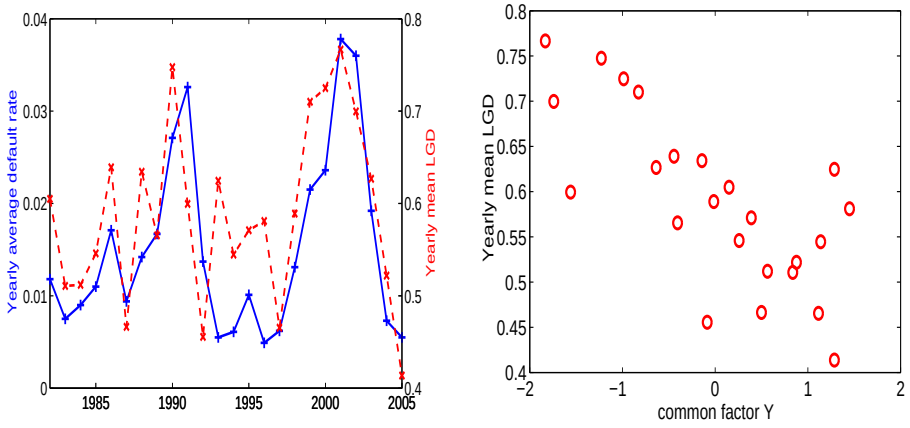
in the GBR-GLM and GBR-JGLM and

$$\mu = \frac{e^{a_1+a_2Y+\nu}}{1+e^{a_1+a_2Y+\nu}}$$

in the GBR-GLMM. In the GBR-JGLM model, the dispersion parameter is modeled to be

$$\varphi = e^{b_1+b_2Y}.$$

The estimates given by the method of least squares are presented in Table 4.1. These estimates are used as the first approximation to the MLE. We are already able to get a first impression of the characteristics of the LGD:



(a) (b)

FIGURE 4.1: (a) Yearly average default rate and yearly mean LGD (1982-2005); (b) The common factor Y estimated by eq. (4.19) versus yearly mean LGD (1982-2005).

1. the coefficient a_2 is negative, indicating a negative relation between Y and mean LGD, just as expected,

2. the coefficient b_2 is very close to zero, suggesting that Y may not be relevant for the estimation of dispersion φ.

It is important to keep in mind that these least squares estimates are only based on the annual mean LGD and LGD volatility. Consequently the above observations are not restricted to any particular sample of simulated LGD realizations, as opposed to estimates to be obtained from the MLE.

	GBR-GLM	GBR-JGLM	GBR-GLMM
a_1	0.3718	0.3718	0.3718
a_2	-0.3054	-0.3054	-0.3054
φ	4.1914	-	4.0907
b_1	-	1.3505	-
b_2	-	-0.0033	-
σ_ν	-	-	0.2686

TABLE 4.1: Estimates given by the method of least squares for different models.

To carry out the MLE we need a sample of LGD realizations. For each year, a realization of the LGD is simulated for each defaulted obligor from a beta distribution matching the empirical mean and variance. This gives in total $1,123$ LGD observations in $T = 24$ years. The Maximum Likelihood Estimates for the various parameters are given in Table 4.2. For the GBR-GLM model, we also report in parenthesis the asymptotic standard errors of the estimates. We find that the estimates given by MLE are very similar to those given by least squares. The Wald test confirms that both a_1 and a_2 are statistically significant (both p-values < 0.0001), which justifies our use of Y as a risk factor for the mean LGD. The log-likelihood ratio statistics of GBR-JGLM and GBR-GLMM to GBR-GLM are $-402.74-(-403.34) = 0.6$ and $-402.74-(-468.78) = 66.04$, respectively. They correspond to p-values 0.44 and < 0.0001 for the chi-square distribution with one degree of freedom. It is clear that GBR-GLMM provides a significant improvement over the basic GBR-GLM, whereas GBR-JGLM fails to do so. AIC and BIC lead to the same conclusion (see Table 4.2). Additional simulation tests show that the above estimation results are very robust. We remark that this however does not suggest that GBR-JGLM should be abandoned in general since the idea of jointly modeling mean and dispersion may be meaningful if we include other covariates, e.g., seniority and presence and quality of collateral.

	GBR-GLM	GBR-JGLM	GBR-GLMM
a_1	0.3459 (0.0359)	0.3471	0.3319
a_2	-0.3213 (0.0298)	-0.3246	-0.3307
φ	3.0276 (0.1149)	-	3.3240
b_1	-	1.0879	-
b_2	-	-0.0306	-
σ_ν	-	-	0.2943
-2ℓ	-402.74	-403.34	-468.78
AIC	-398.74	-395.34	-460.78
BIC	-381.67	-375.25	-440.69

TABLE 4.2: Maximum Likelihood Estimates of various models.

Additionally we also fit our GBR models to a second sample of LGD realizations, simulated from the probit model where the LGD is given by $\lambda_{t,k} = \Phi(c_t + d_t\epsilon_{t,k})$. The parameters c_t and d_t for all t can be conveniently

estimated by the method of moments since

$$E(\Lambda_t) = \Phi\left(\frac{c_t}{\sqrt{1+d_t^2}}\right), \quad E(\Lambda_t^2) = \Phi_2\left(\frac{c_t}{\sqrt{1+d_t^2}}, \frac{c_t}{\sqrt{1+d_t^2}}, \frac{d_t^2}{1+d_t^2}\right),$$

where $\Phi_2(\cdot,\cdot,\rho)$ denotes the bivariate cumulative Gaussian distribution function with correlation ρ. For a proof see Andersen & Sidenius (2004). The MLE procedure gives the estimates $a_1 = 0.3562$, $a_2 = -0.3247$, $\varphi = 3.0589$ for the GBR-GLM model, $a_1 = 0.3571$, $a_2 = -0.3275$, $b_1 = 1.1011$, $b_2 = -0.0260$ for GBR-JGLM and $a_1 = 0.3393$, $a_2 = -0.3144$, $\varphi = 3.4096$, $\sigma_\nu = 0.3261$ for GBR-GLMM. These estimates are broadly in agreement with those in Tables 4.1 and 4.2, which suggests that the parameters in the GBR models are robust to misspecification of the LGD distribution.

Moreover, for the second sample we also look at the quasi-likelihood (see Wedderburn 1974) of a model which assumes that the mean and variance of the LGD are given by (4.20) and $\mu(1-\mu)/(1+\varphi)$ respectively, but the distribution of the LGD is unknown. The quasi-likelihood is then given by

$$QL = \sum_{t=1}^{T}\sum_{k=1}^{K_t}\left[\lambda_{t,k}\log(\mu_t) + (1-\lambda_{t,k})\log(1-\mu_t)\right].$$

Maximization of QL gives $a_1 = 0.3799$, $a_2 = -0.3336$, $\varphi = 3.0979$, indicating that the assumption on the distribution of the LGD probably matters little.

Implication for portfolio risk

It is also interesting to see how much the choice of an LGD model can influence the VaR at the portfolio level. We consider a portfolio of 100 obligors with uniform PD p and correlation ρ as in (4.18) and exposures as follows

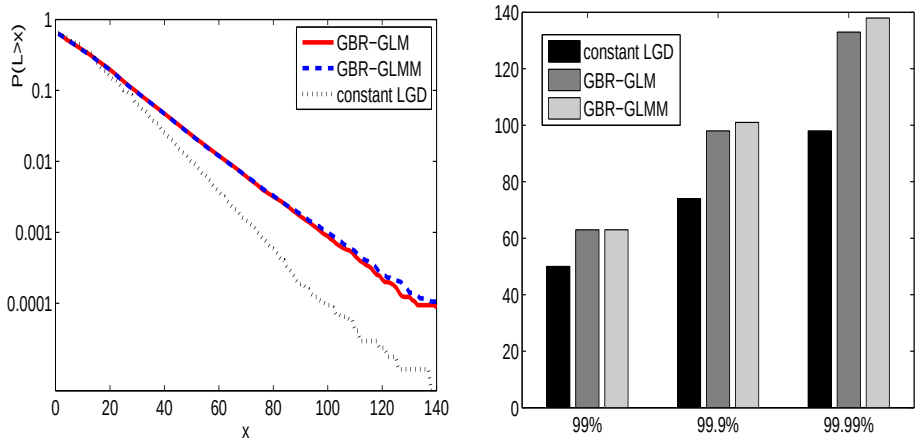
$$w_i = \begin{cases} 1, & k = 1,\ldots,20 \\ 4, & k = 21,\ldots,40 \\ 9, & k = 41,\ldots,60 \\ 16, & k = 61,\ldots,80 \\ 25, & k = 81,\ldots,100. \end{cases}$$

We compare three models for the LGD, (i) the GBR-GLM, (ii) the GBR-GLMM and (iii) the constant LGD model. For the GBR-GLM and GBR-GLMM, the LGD parameters are taken from Table 4.2. In the constant LGD model, we take for all obligors $\Lambda = 0.58$, matching the expected LGD

$E_Y[\mu(Y)]$ in the GBR-GLM model, where $E_Y(\cdot)$ denotes the expectation obtained by integrating over Y.

The portfolio loss distributions plotted in Figure 4.2(a) are based on Monte Carlo simulation with two hundred thousand scenarios. On the one hand the curves of GBR-GLMM and GBR-GLM are almost identical, with GBR-GLMM producing a slightly heavier tail. This is again an indication of the robustness of our GBR models. On the other hand the loss distribution under the constant LGD model deviates substantially from the other two models with random LGD.

We then look at the portfolio VaR at three particular confidence levels 99%, 99.9%, 99.99%, illustrated in Figure 4.2(b). Compared to the constant LGD model, the GBR-GLM (GBR-GLMM) increases the VaR at the three levels by a factor of 1.26 (1.26), 1.32 (1.36) and 1.36 (1.41), respectively. It is apparent that ignoring the systematic risk in the LGD significantly underestimates risk. Moreover, the further in the tail, the higher the degree of underestimation. These results are in line with those reported in Altman et al. (2005), Giese (2006).



(a) (b)

FIGURE 4.2: (a) The portfolio loss distributions and (b) the portfolio VaR at three confidence levels under the three LGD models. The results are based on Monte Carlo (MC) simulation of two hundred thousand scenarios. For GBR-GLM and GBR-GLMM, the LGD parameters are taken from Table 4.2. In the constant LGD model $\Lambda = 0.58$ for all obligors.

4.5 Loss distribution approximations

The calculation of portfolio loss distribution with random LGD is mostly based on Monte Carlo simulation in the literature. To our knowledge the only exception is Giese (2006), where the saddlepoint approximation was employed. An important advantage of adopting the Generalized Beta Regression framework for random LGD is that it allows both the normal approximation and the saddlepoint approximation to efficiently calculate the portfolio loss distribution, thus avoiding the need for time-consuming simulation. Both approximations apply to completely heterogeneous portfolios. For simplicity, we derive the formulas only for the basic GBR-GLM with a single covariate Y, or equivalently, a single-factor model, where the tail probability reads $\mathbb{P}(L \geq x) = \mathbb{E}_Y[\mathbb{P}(L \geq x|Y)]$. Generalization to more complex models is rather straightforward.

4.5.1 Normal approximation

First of all, in the case of a large homogeneous portfolio, the expected loss from obligor i conditional on Y reads

$$\mathbb{E}[L_i(Y)] = \omega_i \mathbb{E}[D_i(Y)]\mathbb{E}[\Lambda_i(Y)] = \omega_i p_i(Y)\mu_i(Y). \tag{4.21}$$

A version of the large homogeneous approximation (LHA) similar to that in the Vasicek model can also be obtained for random LGD:

$$\frac{L(Y)}{\sum_{i=1}^n \omega_i} \to \frac{\sum_{i=1}^n \omega_i p_i(Y)\mu_i(Y)}{\sum_{i=1}^n \omega_i} \quad a.s.$$

When the portfolio is however not sufficiently large or not very homogeneous, unsystematic risk arises. The normal approximation improves on the large homogeneous portfolio approximation by taking into account the variability of portfolio loss L conditional on the common factor Y. The conditional portfolio loss $L(Y)$ can be approximated by a normally distributed random variable with mean $M(Y)$ and variance $V^2(Y)$ such that

$$\begin{aligned} M(Y) &= \sum_{i=1}^n \omega_i p_i(Y)\mu_i(Y), \\ V^2(Y) &= \sum_{i=1}^n \mathbb{E}[L_i^2(Y)] - \sum_{i=1}^n \mathbb{E}[L_i(Y)]^2, \end{aligned}$$

where

$$\begin{aligned}\mathbb{E}[L_i^2(Y)] &= \omega_i^2\mathbb{E}[D_i(Y)]\mathbb{E}[\Lambda_i^2(Y)] = \omega_i^2 p_i(Y)\mathbb{E}[\Lambda_i^2(Y)] \\ &= \omega_i^2 p_i(Y)[\mu_i^2(Y) + \mathbb{V}ar(\Lambda|Y)] \\ &= \omega_i^2 p_i(Y)[\mu_i^2(Y) + \mu_i(Y)(1-\mu_i(Y))/(1+\varphi_i)].\end{aligned}$$

The conditional tail probability is $\mathbb{P}(L \geq x|Y) = \Phi\left(\frac{M(Y)-x}{V(Y)}\right)$ and it follows that the unconditional tail probability reads

$$\mathbb{P}(L \geq x) = \mathbb{E}_Y\left[\Phi\left(\frac{M(Y)-x}{V(Y)}\right)\right]. \tag{4.22}$$

4.5.2 Saddlepoint approximation

The only paper that applies the saddlepoint approximations to the calculation of portfolio credit risk in the presence of random LGD is Giese (2006). However, in his results the portfolio loss distributions obtained from the saddlepoint approximation deviate significantly from those given by Monte Carlo simulation for the majority of loss ranges (see Figures 4 and 5 therein). Following Martin et al. (2001*a*,*b*), Giese (2006) applied the saddlepoint approximation to the *unconditional* MGF of portfolio loss L.

In Chapter 2 we have shown that the saddlepoint approximation method, applied to the *conditional* MGF of L *given the common factor* Y, is an efficient tool to estimate the portfolio credit loss distribution in the Vasicek model. Here we extend Chapter 2 to models with random LGD and show by numerical examples that the saddlepoint approximation is able to produce accurate tail probability approximations to all loss levels and handles well heterogeneous portfolios with exposure concentration.

The use of the saddlepoint approximation only requires the existence of the moment generating function (MGF), which makes the beta distribution assumption for LGD in our framework very attractive. Recall that the MGF of a beta distributed random variable with parameters (α, β) is a confluent hypergeometric function as follows,

$$\mathcal{M}(t) = {}_1F_1(\alpha, \alpha+\beta; t).$$

By basic differentiation, we obtain the following first and second derivatives of the MGF

$$\begin{aligned}\mathcal{M}'(t) &= {}_1F_1(\alpha+1, \alpha+\beta+1; t)\frac{\alpha}{\alpha+\beta}, \\ \mathcal{M}''(t) &= {}_1F_1(\alpha+2, \alpha+\beta+2; t)\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.\end{aligned}$$

In our setting, the obligors are *independent conditional on the common factor* Y. For obligor i, (α_i, β_i) conditional on Y can be determined by (4.4). The conditional MGF and the cumulant generating function (CGF), denoted by $\mathcal{K}$, of the portfolio loss are then given by

$$\mathcal{M}(t, Y) = \prod_{i=1}^{n} \left[1 - p_i + p_i \, {}_1F_1(\alpha_i, \alpha_i + \beta_i; \omega_i t)\right],$$

$$\mathcal{K}(t, Y) = \log(\mathcal{M}(t, Y)) = \sum_{i=1}^{n} \log\left[1 - p_i + p_i \, {}_1F_1(\alpha_i, \alpha_i + \beta_i; \omega_i t)\right].$$

For simplicity of notation, we have suppressed the explicit dependence of p_i and (α_i, β_i) on the common factor Y.

The derivatives of the conditional CGF up to second order are

$$\mathcal{K}'(t, Y) = \sum_{i=1}^{n} \frac{\omega_i p_i \, {}_1F_1(\alpha_i + 1, \alpha_i + \beta_i + 1; \omega_i t)}{1 - p_i + p_i \, {}_1F_1(\alpha_i, \alpha_i + \beta_i; \omega_i t)} \frac{\alpha_i}{\alpha_i + \beta_i},$$

$$\begin{aligned}\mathcal{K}''(t, Y) = \sum_{i=1}^{n} \Bigg\{ & \frac{\omega_i^2 p_i \alpha_i (\alpha_i + 1) \, {}_1F_1(\alpha_i + 2, \alpha_i + \beta_i + 2; \omega_i t)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)[1 - p_i + p_i \, {}_1F_1(\alpha_i, \alpha_i + \beta_i; \omega_i t)]} \\ & - \frac{\omega_i^2 p_i^2 \alpha_i^2 \, {}_1F_1(\alpha_i + 1, \alpha_i + \beta_i + 1; \omega_i t)^2}{(\alpha_i + \beta_i)^2 [1 - p_i + p_i \, {}_1F_1(\alpha_i, \alpha_i + \beta_i; \omega_i t)]^2} \Bigg\}.\end{aligned}$$

After finding the saddlepoint T that solves $\mathcal{K}'(T, Y) = x$ for the loss level x, the tail probability conditional on Y can be approximated by the Lugannani-Rice formula (1.23).

Integrating over Y gives the unconditional tail probability $\mathbb{P}(L \geq x)$, from which the portfolio Value at Risk (VaR) can be derived. Formulas for the calculation of other risk measures like VaR contribution, Expected Shortfall (ES) and ES contribution can be found in Chapter 2.

4.5.3 Numerical results

We now illustrate the performance of the normal and saddlepoint approximations in loss distribution calculation. We first take a homogeneous portfolio with $n = 100$ obligors, each with

$$w = 1, \ p = 0.005, \ \rho = 0.18,$$

The parameters in the LGD are

$$a = [0.37, -0.32], \ \varphi = 3.16,$$

with a logit link for mean LGD. This leads to the following specification of the (conditional) mean LGD

$$\mu = \frac{1}{1 + e^{-0.37+0.32Y}}.$$

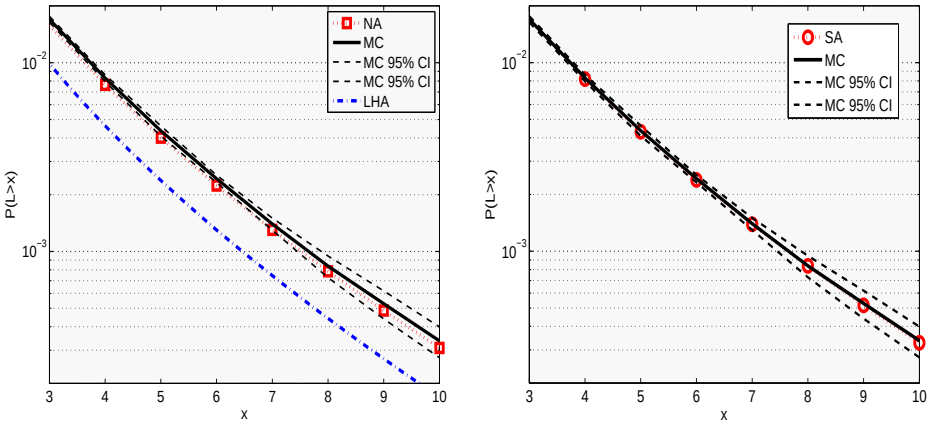
We compare the loss distributions obtained from various approximation methods to the results from a Monte Carlo (MC) simulation. Our benchmark is the sample mean and the accompanying 95% confidence intervals obtained by 10 subsamples of Monte Carlo simulation with 20 thousand replications each. The performance of the approximations is demonstrated in Figure 4.3(a)-(b).

The large homogeneous approximation (LHA) results deviate considerably from our benchmark. This is not surprising as the size of the portfolio is rather small. The normal approximation (NA) provides a significant improvement over the LHA and underestimates risk only slightly. Some of its tail probability estimates however fall out of the 95% confidence interval. By comparison, the saddlepoint approximation (SA) is able to give all tail probability estimates within the 95% confidence interval. The loss distribution given by the saddlepoint approximation is indistinguishable from the benchmark. A remark is that the calculation of the loss distribution in MATLAB costs roughly 4 seconds for the normal approximation and 4 minutes for the saddlepoint approximation on a Pentium 4 2.8 GHz desktop.

Finally we calculate the VaR for the portfolio considered in §4.4.3 with LGD modeled by the GBR-GLM. The results are given in Table 4.3. The MC results are based on two hundred thousand simulated scenarios and can be regarded as our benchmark. In this example the saddlepoint approximation is again very accurate. The normal approximation is however rather unsatisfactory: at all three levels relative errors are around 8%. This is certainly due to the existence of exposure concentration as the variation in the exposures is not negligible. More details on how robust the normal approximation and saddlepoint approximation are in terms of handling exposure concentration have been discussed in Chapter 3.

4.6 Conclusions

In this chapter we have proposed the Generalized Beta Regression framework for modeling systematic risk in loss given default (LGD) in the context of credit portfolio losses. The GBR framework provides great flexibility in random LGD modeling and accommodates well skewness and heteroscedastic errors. The quantities in the GBR models have simple economic inter-



(a) (b)

FIGURE 4.3: The loss distribution obtained from (a) the large homogeneous approximation (LHA), the normal approximation (NA) and (b) the saddlepoint approximation (SA) compared to results based on Monte Carlo (MC) simulation of two hundred thousand scenarios. The MC 95% confidence interval (CI) are based on the standard deviation calculated using 10 simulated sub-samples of 20 thousand scenarios each.

pretation. We have shown that parameter estimation and model selection are straightforward in this framework. Moreover, it has been demonstrated that the portfolio loss distribution can be efficiently evaluated by both the normal approximation and the saddlepoint approximation.

4.A Score function and Fisher information matrix

In this appendix we give details about the score function and the Fisher information matrix for the parameters appearing in the GBR-GLM and GBR-JGLM models. The score function may help to accelerate the convergence in the MLE procedure and the Fisher information matrix leads to the asymptotic standard errors of the maximum likelihood estimates of the parameters in the models. In the GBR-GLMM the corresponding formulas get more complicated and lengthy and therefore they are omitted here. We refer the interested reader to Pan & Thompson (2007) for an example.

The score function, i.e., the partial derivative of the log-likelihood func-

	$\text{VaR}_{99\%}$	$\text{VaR}_{99.9\%}$	$\text{VaR}_{99.99\%}$
MC	63	98	133
NA	58	90	123
SA	63	97	133

TABLE 4.3: Approximations to the portfolio VaR at three confidence levels. The LGD model adopted here is GBR-GLM. The MC results are based on two hundred thousand simulated scenarios and can be regarded as our benchmark.

tion with respect to parameters (μ, φ), reads

$$\frac{\partial \ell}{\partial \mu} = \varphi \left\{ \log\left(\frac{\lambda}{1-\lambda}\right) - \Psi(\mu\varphi) + \Psi[(1-\mu)\varphi] \right\}, \tag{4.23}$$

$$\frac{\partial \ell}{\partial \varphi} = \mu \log\lambda + (1-\mu)\log(1-\lambda) + \Psi(\varphi) - \mu\Psi(\mu\varphi) - (1-\mu)\Psi[(1-\mu)\varphi], \tag{4.24}$$

where λ is a realization of the LGD and $\Psi(\cdot)$ is the digamma function.

The second order partial derivatives of the log-likelihood function with respect to parameters (μ, φ) are

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\varphi^2\{\Psi'(\mu\varphi) + \Psi'[(1-\mu)\varphi]\}, \tag{4.25}$$

$$\frac{\partial^2 \ell}{\partial \varphi^2} = \Psi'(\varphi) - \mu^2\Psi'(\mu\varphi) - (1-\mu)^2\Psi'[(1-\mu)\varphi], \tag{4.26}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \varphi} = \frac{1}{\varphi}\frac{\partial \ell}{\partial \mu} - \varphi\{\mu\Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\}. \tag{4.27}$$

where $\Psi'(\cdot)$ is the trigamma function.

In the GBR-GLM, the parameters to be estimated are a and φ. The score function for φ is given by (4.24); the score function with respect to a_i, the i-th element of a, is given by

$$\frac{\partial \ell}{\partial a_i} = \frac{\partial \ell}{\partial \mu}\frac{\partial \mu}{\partial a_i} = \varphi \left\{ \log\left(\frac{\lambda}{1-\lambda}\right) - \Psi(\mu\varphi) + \Psi[(1-\mu)\varphi] \right\} \frac{\zeta_i}{g'(\mu)}. \tag{4.28}$$

The Fisher information matrix is the negative of the expectation of the second derivative of the log-likelihood with respect to the parameters. The

entries in the Fisher information matrix are

$$-\mathbb{E}\left(\frac{\partial^2\ell}{\partial\varphi^2}\right) = -\frac{\partial^2\ell}{\partial\varphi^2}, \tag{4.29}$$

$$-\mathbb{E}\left(\frac{\partial^2\ell}{\partial a_i\partial a_j}\right) = -\frac{\partial^2\ell}{\partial\mu^2}\frac{\zeta_i\zeta_j}{(g'(\mu))^2}, \tag{4.30}$$

$$-\mathbb{E}\left(\frac{\partial^2\ell}{\partial a_i\partial\varphi}\right) = \varphi\{\mu\Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\}\frac{\zeta_i}{g'(\mu)}. \tag{4.31}$$

In the GBR-JGLM, the parameters to be estimated are a and b. The score function for the coefficient a is given by (4.28) and that for b_i, the i-th element of b, is as follows

$$\begin{aligned}\frac{\partial\ell}{\partial b_i} = \frac{\partial\ell}{\partial\varphi}\frac{\partial\varphi}{\partial b_i} = \{&\mu\log\lambda + (1-\mu)\log(1-\lambda) + \\ &+\Psi(\varphi) - \mu\Psi(\mu\varphi) - (1-\mu)\Psi[(1-\mu)\varphi]\}\,\frac{\zeta_i}{h'(\varphi)}.\end{aligned} \tag{4.32}$$

The Fisher information matrix contains $-E\left(\frac{\partial^2\ell}{\partial a_i\partial a_j}\right)$ given by (4.30) and

$$-\mathbb{E}\left(\frac{\partial^2\ell}{\partial b_i\partial b_j}\right) = -\frac{\partial^2\ell}{\partial\varphi^2}\frac{\zeta_i\zeta_j}{(h'(\varphi))^2}, \tag{4.33}$$

$$-\mathbb{E}\left(\frac{\partial^2\ell}{\partial a_i\partial b_j}\right) = \varphi\{\mu\Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\}\frac{\zeta_i\zeta_j}{g'(\mu)h'(\varphi)}. \tag{4.34}$$

4.B LGD Statistics by Year

In this section we present a table of the LGD statistics by year, from 1982 until 2005. This table is taken from Bruche & González-Aguado (2008), where mean recovery rate (RR) is reported instead of LGD. The column of mean LGD here is calculated to be 1 minus RR, i.e., LGD=1-RR.

Year	PD	# of defaults	mean LGD	LGD volatility
1982	1.18%	12	60.49%	14.9%
1983	0.75%	5	51.07%	23.53%
1984	0.9%	11	51.19%	17.38%
1985	1.1%	16	54.59%	21.87%
1986	1.71%	24	63.91%	18.82%
1987	0.94%	20	46.64%	26.94%
1988	1.42%	30	63.43%	17.97%
1989	1.67%	41	56.54%	28.78%
1990	2.71%	76	74.76%	22.28%
1991	3.26%	95	59.95%	26.09%
1992	1.37%	35	45.55%	23.38%
1993	0.55%	21	62.46%	20.11%
1994	0.61%	14	54.46%	20.46%
1995	1.01%	25	57.1%	25.25%
1996	0.49%	19	58.1%	24.68%
1997	0.62%	25	46.54%	25.53%
1998	1.31%	34	58.9%	24.56%
1999	2.15%	102	71.01%	20.4%
2000	2.36%	120	72.49%	23.36%
2001	3.78%	157	76.66%	17.87%
2002	3.6%	112	69.97%	17.18%
2003	1.92%	57	62.67%	23.98%
2004	0.73%	39	52.19%	24.1%
2005	0.55%	33	41.37%	23.46%

Chapter 5

Adaptive Integration for Multi-factor Portfolio Credit Loss Models

5.1 Introduction

In this chapter we switch from the one-factor portfolio loss model to multi-factor models. Consider the computation of the tail probability of credit portfolio loss L in the two-level factor model in the form of (1.15). In such latent factor models the obligors are independent conditional on some d factors, denoted by $\mathbf{Y}$. We are interested in the estimation of the tail probability

$$\mathbb{P}(L > x) = \int \mathbb{P}\left(L > x \,|\, \mathbf{Y} = \mathbf{y}\right) \mathrm{d}F_{\mathbf{Y}}(\mathbf{y}), \tag{5.1}$$

especially for extreme losses x. We shall focus on the Gaussian factor model where $\mathbf{Y}$ follows a d-dimensional joint normal distribution, although more general distributions can be handled as well in the present context.

The integrand $\mathbb{P}\left(L > x \,|\, \mathbf{Y}\right)$ can be approximated by a variety of methods (see Chapter 3) since conditional on $\mathbf{Y}$, the portfolio loss L reduces to a sum of independent random variables. In a one-factor model ($d = 1$), the calculation of the integral can be handled efficiently by adaptive or non-adaptive Gaussian quadratures. The computation of the tail probability $\mathbb{P}(L > x)$ in a multi-factor model is much more involved. The product quadrature rule becomes impractical because the number of function evaluations grows exponentially with d and the so-called *curse of dimensionality* arises.

In this chapter we deal with the high-dimensionality and show that *globally adaptive algorithms* are very well suited for the calculation of the tail probability. A high-dimensional globally adaptive integration algorithm successively divides the integration region into subregions, detects the subregions where the integrand is most irregular, and places more points in those subregions. We first recall the Genz-Malik (Genz & Malik 1980) rule, a deterministic multiple integration rule, and adapt it so that it takes advantage of the specific properties of the integral of interest and is suitable for portfolio credit models with a number of factors less than, approximately, 8. Later on we arrive at the adaptive Monte Carlo integration, which essentially replaces the deterministic integration rule by antithetic random numbers. Being a globally adaptive algorithm, our approach is distinct from the well-known recursive Monte Carlo algorithm MISER (see Press & Farrar 1990), although both methods use stratified sampling.

The rest of the chapter is organized as follows. We give in §5.2 an introduction into a multi-factor portfolio credit loss model and point out the important monotonicity property of the conditional tail probability as a function of the common factors. In §5.3 we briefly review the globally adaptive integration algorithm and present a tailor-made Genz-Malik rule for the computation of tail probability in the context of portfolio credit loss, followed by some numerical results. We then discuss the adaptive Monte Carlo integration, and, in particular, the adaptivity criterion and the probabilistic error bounds in this context, in §5.4. §5.5 concludes.

5.2 Multi-factor portfolio credit loss model

We are interested in the estimation of tail probability (5.1), especially for extreme losses. Our starting point is the widely used Gaussian factor model in the form of (1.15). A decomposition of the standardized log asset value X_i is carried out as follows,

$$X_i = a_{i1}Y_1 + \cdots + a_{id}Y_d + b_i\varepsilon_i, \tag{5.2}$$

where $\mathbf{Y} = (Y_1 \dots Y_d)$ can be seen as systematic factors that affect more than one obligor and ε_i is an idiosyncratic part that only affects an obligor itself. $Y_1, \dots, Y_d$ and ε_i are all independent univariate standard Gaussian random variables and $a_{i1}^2 + \cdots + a_{id}^2 + b_i^2 = 1$ so that the X_i are also standard normally distributed.

We further make the following assumption.

Assumption 5.1. *The coefficients a_{ik} and b_i in (5.2) are nonnegative for all i, k.*

A dedicated algorithm to ensure that this assumption is satisfied is proposed in the next chapter.

Write $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{id})$. For an obligor i with default probability p_i and default threshold $c_i = \Phi^{-1}(p_i)$, its probability of default conditional on the common factor $\mathbf{Y}$ is given by

$$p_i(\mathbf{Y}) = \mathbb{P}(D_i = 1|\mathbf{Y}) = \mathbb{P}(X_i < c_i|\mathbf{Y}) = \Phi\left(\frac{\Phi^{-1}(p_i) - \boldsymbol{a}_i \cdot \mathbf{Y}}{b_i}\right). \quad (5.3)$$

Eq. (5.3) shows that the individual conditional default probability is non-increasing in $\mathbf{Y}$. An important consequence is that the conditional tail probability of portfolio loss $\mathbb{P}(L > x \,|\, \mathbf{Y})$ is also non-decreasing in $\mathbf{Y}$. Without loss of generality, we prove the following proposition in the case of constant LGD.

Proposition 5.2. *The function*

$$f(y_1, y_2, ..., y_m) = \mathbb{P}(L > x|Y_1 = y_1, Y_2 = y_2, ..., Y_m = y_m),$$

is non-increasing in all its variables y_k.

Proof. Let us write

$$L = \sum_{i=1}^{n} \omega_i \mathbf{1}_{\{X_i < c_i\}} = \sum_{i=1}^{n} \omega_i \mathbf{1}_{\{a_{i1}y_1 + \cdots + a_{id}y_d + b_i\varepsilon_i < c_i\}}.$$

The conditional tail probability can be reformulated to be

$$\mathbb{P}(L > x|Y_1 = y_1, \ldots, Y_d = y_d) = \mathbb{P}\left(\sum_{i=1}^{n} \omega_i \mathbf{1}_{\{a_{i1}y_1 + \cdots + a_{id}y_d + b_i\varepsilon_i < c_i\}} > x\right).$$

The indicator function

$$\mathbf{1}_{\{a_{i1}y_1 + \cdots + a_{id}y_d + b_i\varepsilon_i < c_i\}} = \mathbf{1}_{\{\varepsilon_i < \frac{1}{b_i}(c_i - a_{i1}y_1 - \cdots - a_{id}y_d)\}}$$

is non-increasing in y_k for all k when a_{ik} and b_i are nonnegative for all i. It follows that

$$\sum_{i=1}^{n} \omega_i \mathbf{1}_{\{a_{i1}y_1 + \cdots + a_{id}y_d + b_i\varepsilon_i < c_i\}}$$

is also non-increasing in y_k for all k. This immediately leads to the assertion. □

In addition it is not difficult to derive that $\mathbb{P}(L > x| - \infty, \ldots, -\infty) = 1$ and $\mathbb{P}(L > x| + \infty, \ldots, +\infty) = 0$.

The rest of this chapter hinges strongly on the validity of Prop. 5.2. Note that Prop. 5.2 is quite a general result. Its proof is not contingent on the assumption that $Y_1, \ldots Y_d$ are independent. The distributions of $\mathbf{Y}$ and ε_i, $i = 1, \ldots, n$ are not relevant either. The monotonicity holds more generally for models in which

1. $\mathbf{Y}$ and ε_i, $i = 1 \ldots n$ are independent, and
2. the factor loadings, a_{ik}, $i = 1, \cdots, n$, $k = 1, \cdots, d$ are all nonnegative.

Therefore there is absolutely no problem to apply our adaptive integration methods if the Gaussian model is replaced by Lévy models that are able to produce more heavy-tailed loss distributions and provide a better fit to the present day financial market data.

Proposition 5.3. *$\mathbb{P}(L \geq x|Y_1, Y_2, \ldots, Y_d)$ is continuous and differentiable with respect to Y_k, $k = 1, \cdots, d$.*

Proof. Denote by $\theta = (\theta_1, \ldots, \theta_n) = \{0, 1\}^n$ a realization of $(D_1, \ldots, D_n)$ and write $\omega = (\omega_1, \ldots, \omega_n)$. The conditional tail probability is given by

$$\mathbb{P}(L > x|\mathbf{Y}) = \sum_{\theta:\omega\cdot\theta>x} \mathbb{P}\left(D_i = \theta_i, i = 1, \ldots, n|\mathbf{Y}\right).$$

As D_i and D_j are independent conditional on $\mathbf{Y}$, we get

$$\mathbb{P}\left(D_i = \theta_i, i = 1, \ldots, n|\mathbf{Y}\right) = \prod_{i=1}^{n} \left[p_i\left(\mathbf{Y}\right)\right]^{\theta_i} \left[1 - p_i\left(\mathbf{Y}\right)\right]^{1-\theta_i}.$$

Since $p_i\left(Y_1, Y_2, \ldots, Y_d\right)$ is continuous and differentiable in Y_k for all k, so is the tail probability $\mathbb{P}(L > x|Y_1, Y_2, \ldots, Y_d)$. □

Remark 5.4. More generally, Prop 5.2 also holds if LGD is stochastically decreasing in all Y_k, $k = 1, \cdots, d$ and Prop 5.3 also holds if LGD is continuous and differentiable with respect to Y_k, $k = 1, \cdots, d$.

5.3 Globally adaptive algorithms for numerical integration

5.3.1 Globally adaptive algorithms

Consider now a general integral over a d-dimensional rectangular region $\mathcal{C}^d$

$$I(f) = \int \cdots \int_{\mathcal{C}^d} f(\mathbf{x}) g(\mathbf{x}) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_d, \tag{5.4}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $g(\cdot)$ is a weight function.

Monte Carlo (MC) simulation and quasi-Monte Carlo (QMC) methods are the prevailing methods to solve the multi-dimensional problems appearing in finance. Both methods do not suffer from the dimensionality issue. Recall that integration with both MC and QMC methods requires a transformation of integration region into the unit hypercube $[0,1]^d$. Pseudo-random numbers or quasi-random sequences are then generated uniformly in the $[0,1]^d$ hypercube. QMC methods use deterministic sequences that have better uniform properties measured by discrepancy. Both methods become inefficient if most of the points fall outside the regions which are significant for the evaluation of the integral. In this case the better uniform properties of QMC sequences over MC can be meaningless.

We remark that importance sampling (IS) provides effective variance reduction to plain MC methods in the case of rare event simulation. However, as shown in Glasserman et al. (2008), IS in multi-factor credit models requires the use of a mixture of IS distributions and its implementation is highly non-trivial. Therefore we do not include this method in our comparison.

An adaptive integration algorithm differs fundamentally from Monte Carlo and quasi-Monte Carlo methods in that it successively divides the integration region into subregions, detects the subregions where the integrand is most irregular, and places more points in those subregions.

We will restrict ourselves to the globally adaptive algorithms for multi-dimensional integration, which typically have a structure that consists of the following steps:

1. Choose the subregion with largest estimated error from a collection of subregions.

2. Subdivide the chosen subregion.

3. Apply an integration rule to the resulting new subregions; update the collection of subregions.

4. Update the global integral and error estimate; check whether a predefined termination criterion is met; if not, go back to step 1 .

The important ingredients of an adaptive algorithm are

1. an integration rule for estimating the integral in each subregion.

2. an error estimator for each subregion.

3. a subdivision rule for dividing the chosen region(s) into subregions.

4. a stop rule to check whether the termination criteria are met.

5.3.2 The Genz-Malik rule

The Genz-Malik (Genz & Malik 1980) rule represents an integration rule in the square $[-1,1]^d$ that can be readily generalized to any rectangular region by an affine transformation. It is a fully symmetric degree 7 rule that is given as follows

$$\begin{aligned} I_7(f) =& u_1 f(0,0,\ldots,0) + u_2 \sum_{FS} f(\lambda_2,0,0,\ldots,0) + u_3 \sum_{FS} f(\lambda_3,0,0,\ldots,0) + \\ &+ u_4 \sum_{FS} f(\lambda_4,\lambda_4,0,0,\ldots,0) + u_5 \sum_{FS} f(\lambda_5,\lambda_5,\ldots,\lambda_5), \end{aligned} \tag{5.5}$$

where $\sum_{FS}$ denotes a fully symmetric summation over all permutations of coordinates including sign changes and

$$\lambda_2 = \frac{3}{\sqrt{70}},\ \lambda_3 = \lambda_4 = \frac{3}{\sqrt{10}},\ \lambda_5 = \frac{3}{\sqrt{19}},$$

$$u_1 = 2^d(12824 - 9120d + 400d^2)/19683, u_2 = 2^d(980/6561),$$

$$u_3 = 2^d(1820 - 400d)/19683, u_4 = 2^d(200/19683), u_5 = 6859/19683.$$

All integration nodes are inside the integration domain. The degree 7 integration rule requires $2^d + 2d^2 + 2d + 1$ integrand evaluations for a function of d variables and is thus known to be most advantageous for problems with $d \leq 8$ (see Genz 1984). We remark that, by contrast, a Gauss-Legendre quadrature rule of degree 7 would require 4^d integration evaluations, which is significantly larger for $d \geq 3$.

The Genz-Malik rule distinguishes itself from other multiple integration rules in that it has an embedded degree 5 rule for error estimation. The degree 5 rule uses a subset of points of the degree 7 rule, which means that no additional integrand evaluations are necessary. This is highly desirable for multi-dimensional problems. The embedded degree 5 rule is given by

$$\begin{aligned} I_5(f) =& u'_1 f(0,0,\ldots,0) + u'_2 \sum_{FS} f(\lambda_2,0,0,\ldots,0) + u'_3 \sum_{FS} f(\lambda_3,0,0,\ldots,0) + \\ &+ u'_4 \sum_{FS} f(\lambda_4,\lambda_4,0,0,\ldots,0), \end{aligned}$$

with $u'_1 = 2^d(729 - 950d + 50d^2)/729$, $u'_2 = 2^d(245/486)$, $u'_3 = 2^d(265 - 100d)/1458$, $u'_4 = 2^d(25/729)$. The error approximation for each subregion is simply the difference of these two rules, i.e.,

$$\epsilon = I_7 - I_5. \tag{5.6}$$

Starting from the whole integration region, in every step the (sub)region with the largest error estimate in absolute value will be chosen for subdivision. To avoid an exponential explosion in the number of subregions, the chosen region is not divided into 2^d subregions but only into two. The subdivision rule used to determine along which direction to divide is due to van Dooren & de Ridder (1976). In particular, the direction that has the largest fourth divided difference is halved. Five points are used in the direction $i = 1, \ldots, d$,

$$x_i = -\lambda_3, -\lambda_2, 0, \lambda_2, \lambda_3, \quad \text{and } x_j = 0 \text{ for } j \neq i$$

and the fourth divided differences are given by

$$Dif_i = [f(-\lambda_3) - 2f(0) + f(\lambda_3)] - \frac{\lambda_2^2}{\lambda_3^2}[f(-\lambda_2) - 2f(0) + f(\lambda_2)]. \tag{5.7}$$

Note that no additional integrand evaluations are required here.

It follows that after $K - 1$ subdivisions, the integration region $\mathcal{C}^d$ is divided into K non-overlapping rectangular subregions. For any subregion k, the Genz-Malik rule gives a local integral estimate $I_7^{(k)}$, a local error estimate $\epsilon^{(k)}$ and a direction $s^{(k)}$ that has the largest fourth divided difference given by (5.7), which is then chosen for the next subdivision. Aggregating the local information over $\mathcal{C}^d$ we obtain a global integral estimate to $I(f)$ as follows,

$$I_7(f) = \sum_{k=1}^{K} I_7^{(k)}(f), \tag{5.8}$$

where $I_7^{(k)}(f)$ is calculated by (5.5) with a suitable affine transformation. Meanwhile the K local error estimates sum to a global error estimate i.e., $\epsilon = \sum_{k=1}^{K} \epsilon^{(k)}$. Typically the error estimate is used to check whether the termination criteria are met.

A remark is that when a region is subdivided, integrand values previously evaluated in this region are discarded and the integration rule is applied in both new subregions. Roughly this means that in the long run only half of the integrand evaluations is used for the calculation of the integral, the other half is abandoned in the process of subdivision.

5.3.3 A tailor-made adaptive Genz-Malik rule

We restate our problem as calculating

$$I(f) = \int \cdots \int_{\mathcal{C}^d} f(\mathbf{Y})\phi(\mathbf{Y})\mathrm{d}Y_1 \cdots \mathrm{d}Y_d, \tag{5.9}$$

where $f(\mathbf{Y}) = \mathbb{P}(L > x|\mathbf{Y})$ is monotonic along all dimensions and bounded in $[0, 1]$ and $\phi(\mathbf{Y})$ is the probability density function of d-dimensional normal distribution with zero mean and identity covariance matrix.

A significant problem with the Genz-Malik rule is that the weights u_i can be negative. Consequently even though our integrand is always positive in some subregions a straightforward Genz-Malik rule may give negative results for the integral. This however can be rather easily dealt with in our context. Recall from Prop. 5.2 that $f(\mathbf{Y})$ should be bounded in any rectangular (sub)region $[\mathtt{a}_1, \mathtt{b}_1] \times [\mathtt{a}_2, \mathtt{b}_2] \ldots \times [\mathtt{a}_d, \mathtt{b}_d]$, more specifically, $f(\mathtt{b}) \leq f(\mathbf{Y}) \leq f(\mathtt{a})$, where $\mathtt{a} = (\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_d)$ and $\mathtt{b} = (\mathtt{b}_1, \mathtt{b}_2, \ldots, \mathtt{b}_d)$. As a result we have for $I^{(k)}(f)$ both an upper bound and a lower bound, i.e.,

$$f(\mathtt{b}^{(k)}) \prod_{i=1}^{d} \left(\Phi(\mathtt{b}_i^{(k)}) - \Phi(\mathtt{a}_i^{(k)}) \right) \leq I^{(k)}(f) \leq f(\mathtt{a}^{(k)}) \prod_{i=1}^{d} \left(\Phi(\mathtt{b}_i^{(k)}) - \Phi(\mathtt{a}_i^{(k)}) \right). \tag{5.10}$$

Denote by $\mathcal{U}^{(k)}$, $\mathcal{L}^{(k)}$ the upper bound and lower bound respectively for subregion k. Positivity of the integrand can then be preserved by the following correction,

$$I_7^{(k)}(f) = I_7^{(k)}(f) 1_{\{\mathcal{L}^{(k)} \leq I_7^{(k)}(f) \leq \mathcal{U}^{(k)}\}} + \mathcal{L}^{(k)} 1_{\{I_7^{(k)}(f) < \mathcal{L}^{(k)}\}} + \mathcal{U}^{(k)} 1_{\{I_7^{(k)}(f) > \mathcal{U}^{(k)}\}}. \tag{5.11}$$

The last term in Eq. (5.11) corrects in addition, to some extent, possible overshooting of the integration rule. More importantly, the local bounds over all subregions can be aggregated to a global upper bound and a global lower bound for the whole integration region $\mathcal{C}^d$. It follows that the estimate to the integral should asymptotically converge to $I(f)$ if we continue the subdivision until the global upper and lower bounds coincide.

It is also important to recognize that the integral can be calculated exactly for subregions where the integrand is constantly 0 or 1. These subregions can be identified by simply evaluating the integrand at the end points $\mathtt{a}^{(k)}$ and $\mathtt{b}^{(k)}$. By bounded monotonicity we have

$$I^{(k)}(f) = \begin{cases} 0 & \text{if } f(\mathtt{a}^{(k)}) = 0, \\ \prod_{i=1}^{d} \left(\Phi(\mathtt{b}_i^{(k)}) - \Phi(\mathtt{a}_i^{(k)}) \right) & \text{if } f(\mathtt{b}^{(k)}) = 1. \end{cases} \tag{5.12}$$

In these subregions we should set $\epsilon^{(k)} = 0$.

We are now in a position to present our adaptive integration algorithm based on a tailor-made Genz-Malik rule. It is presented as Algorithm 5.1. For clarity in notation we use superscript l for local estimates in any subregion. s denotes the subdivision direction of a subregion.

Algorithm 5.1 adaptive integration based on the Genz-Malik rule

Apply the GM rule over the integration region,
return I_7^l, ϵ^l and subdivision direction s, impose (5.11)
while termination criteria not met **do**
  Choose the (sub)region with largest ϵ^l and divide along direction s.
  Compute $f(\mathtt{a})$ and $f(\mathtt{b})$ for the resulting two subregions.
  if $f(\mathtt{a}) = 0$ or $f(\mathtt{b}) = 1$ **then**
    Apply (5.12), let $\epsilon^l = 0$.
  else
    Apply the GM rule to both subregions, return I_7^l, ϵ^l and s, impose (5.11).
  end if
  Update I_7, ϵ and the subregion collection.
end while

Remark 5.5. The above algorithm 5.1 can be easily generalized to accommodate non-Gaussian factor models. In the case of orthogonal common factors, it is only necessary to replace ϕ and Φ by the joint *p.d.f.* and *c.d.f.* of the common factors.

The error estimate ϵ deserves further investigation. According to Lyness & Kaganove (1976), Berntsen (1989), error estimates based on differences of two rules can be unreliable. Various ways of improving the reliability of error estimates can be found in Berntsen (1989), Berntsen et al. (1991), among which a simple approach is to use more than two integration rules for error estimation. Following this we take a parsimonious change by including the degree 1 midpoint rule for the square $[-1, 1]^d$,

$$I_1 = f(0, 0, \ldots, 0),$$

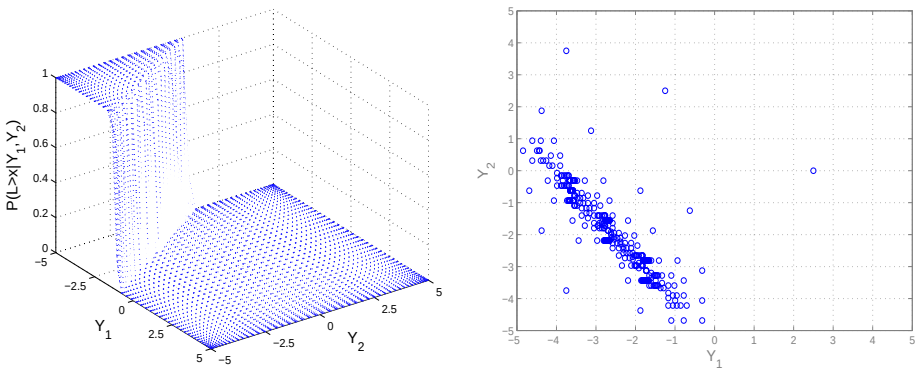
which is also embedded in the degree 7 rule, as a second check on error. Thus the error estimate is defined to be

$$\epsilon = (I_7 - I_5)\mathbf{1}_{\{|I_7 - I_5| \geq |I_7 - I_1|\}} + (I_7 - I_1)\mathbf{1}_{\{|I_7 - I_5| < |I_7 - I_1|\}}. \tag{5.13}$$

The new error estimate is more reliable but also conservative. A stop rule based on such absolute or relative errors can consequently be ineffective. It may well happen that while the integration rule is giving accurate results, the error estimate remains above a given precision level and the subdivision carries on more than necessary, see e.g. Genz & Kass (1997). Therefore we adopt a simple termination criterion that does not rely on ϵ: we prescribe a maximum number of integrand evaluations or similarly, a maximum number of subdivisions.

5.3.4 Numerical results

Here we first illustrate by a two-factor model example how the adaptive integration algorithm works. For some arbitrary portfolio and suitable loss level x, Figure 5.1(a) gives the conditional tail probability $\mathbb{P}(L > x|Y_1, Y_2)$ for (Y_1, Y_2) truncated to the square $[-5, 5]^2$. The integrand turns out to contribute nothing to the integral value in almost 7/8 of the area, which suggests that an adaptive algorithm should be favored. Figure 5.1(b) shows a scatterplot of the subregion centers generated by the adaptive algorithm. It is clearly seen that the adaptive algorithm does focus its integrand evaluation in those subregions in which the integrand values vary rapidly.



(a) integrand $\mathbb{P}(L > x|Y_1, Y_2)$ (b) scatterplot subregion centers

FIGURE 5.1: Adaptive integration for a two-factor model. (a) integrand $\mathbb{P}(L > x|Y_1, Y_2)$; (b) centers of the subregions generated by adaptive integration.

Let us consider a credit portfolio of 1000 obligors with $\omega_i = 1$, $p_i = 0.0033$, $i = 1, \ldots, 1000$. We move to a five-factor model such that the obligors are grouped into 5 buckets of 200 obligors. Within each bucket, the

obligors have identical factor loadings

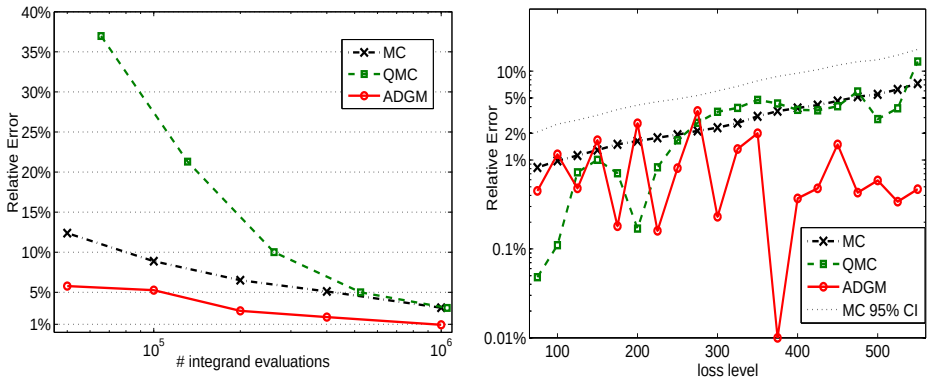
$$\boldsymbol{a}_i = \begin{cases} \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right), i = 1, \ldots, 200, \\ \left(\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right), i = 201, \ldots, 400, \\ \left(\frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, 0, 0\right), i = 401, \ldots, 600, \\ \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, 0, 0\right), i = 601, \ldots, 800, \\ \left(\frac{1}{\sqrt{2}}, 0, 0, 0, 0\right), i = 800, \ldots, 1000. \end{cases}$$

We compute the tail probabilities over a wide range of 20 loss levels from 75 to 550, with an increment of 25. These losses correspond to quantiles of the portfolio loss distribution roughly from 99% to 99.99%. As a benchmark we use simulation with a tremendous amount of scenarios. Integrand evaluation is accomplished by the normal approximation (see §3.2.1) and is considered to be exact. We compare the results obtained by the adaptive Genz-Malik rule (ADGM), the MC and QMC methods with a similar number of integrand evaluations, denoted by N. For the QMC method we choose the SOBOL sequence. The sequence is generated by the GSL library, which is based on Antonov & Saleev (1979).

We control the number of integrand evaluations rather than computation time in the course of subdivision because the latter can vary substantially for different portfolios, different methods for integrand evaluation and different data structures of the subregion collection. The approximation error is measured by the absolute relative error (RE) defined as $|\hat{I}(f) - I(f)|/I(f)$, where $I(f)$ is the result given by the benchmark and $\hat{I}(f)$ denotes any estimate to $I(f)$. The absolute relative errors reported for the Monte Carlo method are averaged over 100 different runs. Alongside the mean absolute relative error we also report 1.96 times the normalized standard deviation of MC, which gives roughly the absolute value of both end points of the 95% confidence interval.

We first show in Figure 5.2(a) the average performance of each method over all 20 loss levels with different numbers of integrand evaluations N ranging from $50,000$ to 2^{20}. Note that for the adaptive integration, these correspond roughly to K, the number of subregions, from 250 to $5,000$ because the Genz-Malik rule samples in five dimensions around 100 points in each subregion. Apparently the adaptive integration consistently outperforms both Monte Carlo and the quasi-Monte Carlo methods for all levels of N.

With around $N = 10^6$ evaluations, it seems that all three methods produce satisfactory results. Relative errors are, respectively, 0.9% (ADGM),

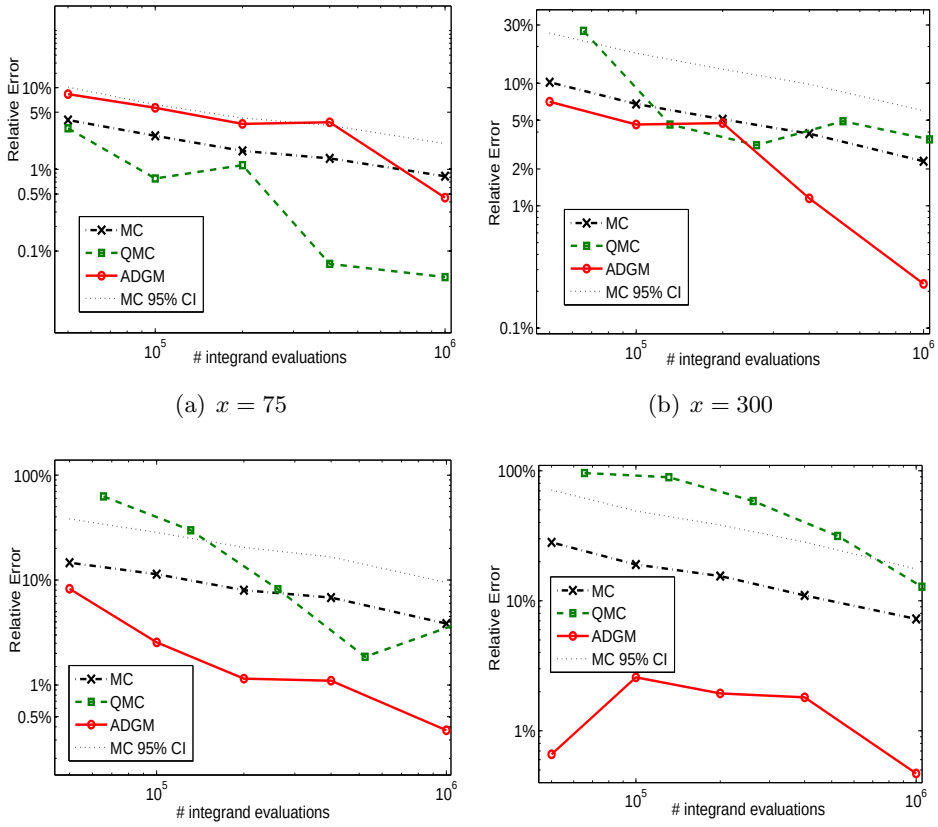


(a) (b)

FIGURE 5.2: (a) Estimation of relative errors with the adaptive Genz-Malik rule (ADGM), Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods over 20 loss levels. The number of total integrand evaluations N ranges from $50,000$ to 2^{20}. (b) Estimation of relative errors with the adaptive Genz-Malik rule (ADGM), Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods with around $N = 10^6$ evaluations for various loss levels.

3.0% (QMC) and 3.1% (MC). Figure 5.2(b) further compares the performance of the different methods with around 10^6 evaluations for various loss levels. Monte Carlo and quasi-Monte Carlo methods are quite accurate for low loss levels but deteriorate notably as the loss level increases. An upward trend in the relative error is conspicuous for both methods. In particular, for the loss level $x = 550$, Monte Carlo has an error of 8.8% and quasi-Monte Carlo gives 12.8%. By contrast, the relative error of the adaptive integration for the same loss level is merely 0.5%. Even though at some low loss levels adaptive integration is not superior to the other two methods, it dominates its two opponents for loss levels larger than 300.

A close-up look at the three methods for different loss levels is presented in Figure 5.3. We show results for four loss levels, $x = 75$, 300, 400, 550, which correspond roughly to quantiles 99%, 99.9%, 99.95% and 99.99%, respectively. The adaptive integration is remarkably distinct from Monte Carlo and quasi-Monte Carlo methods in that it is not sensitive to the portfolio loss level of interest. As a consequence, the adaptive integration becomes more and more advantageous compared to Monte Carlo and quasi-Monte Carlo methods for increasing loss levels. This is especially attractive for the purpose of determining the portfolio VaR, which always involves large loss levels.



(a) $x = 75$ (b) $x = 300$

(c) $x = 400$ (d) $x = 550$

FIGURE 5.3: Relative estimation error of $\mathbb{P}(L > x)$ by all methods for four different loss levels x. PD= 0.0033, $\rho = 0.2$, $d = 5$.

5.4 Adaptive Monte Carlo integration

Adaptive integration based on the Genz-Malik rule thus provides an efficient tool for calculating credit portfolio loss distribution in a multi-factor framework. It is particularly advantageous in the tail of the loss distribution. However the adaptive Genz-Malik rule suffers from two problems. First, the integration rule is only able to handle models with relatively low dimension, say $d \leq 8$. This is due to the fact that the number of integrand evaluations is fully determined by d and grows exponentially. Second, no practical error bounds are available for the estimates.

A natural alternative that does not suffer from the above two problems is Monte Carlo integration. A Monte Carlo integration *embedded in a globally*

adaptive algorithm is able to provide an unbiased estimate of the integral and also probabilistic error bounds for the estimate. In the mean-time it has higher accuracy and faster convergence than the plain Monte Carlo integration. The idea of adaptive Monte Carlo integration is not new. Two well-known algorithms can be found in Press & Farrar (1990) and Lepage (1978, 1980). Our approach resembles that of Press & Farrar (1990) in the sense that both methods use stratified sampling. However the two algorithms are distinct in terms of error estimation, subdivision rule, stop rule, etc. From a more general perspective, the method in Press & Farrar (1990) is *not* a globally adaptive algorithm.

Our adaptive Monte Carlo integration replaces the degree 7 Genz-Malik rule with uniform random numbers as the integration rule. Let us go back to Eq. (5.9) and write $\xi = f \cdot \phi$. The tail probability as in Eq. (5.8) can then be approximated by

$$\hat{I}(\xi) = \sum_{k=1}^{K} \hat{I}^{(k)}(\xi) = \sum_{k=1}^{K} v^{(k)} \sum_{j=1}^{M} \frac{\xi_j^{(k)}}{M}, \tag{5.14}$$

where K is the number of subregions, M is a fixed constant that gives the number of points in each subregion and $v^{(k)}$ denotes the volume of subregion k. This estimate $\hat{I}(\xi)$ is unbiased since it is a sum of unbiased Monte Carlo estimates. The variance of $\hat{I}(\xi)$ is given by

$$\mathbb{V}ar\left(\hat{I}(\xi)\right) = \sum_{k=1}^{K} \mathbb{V}ar\left(v^{(k)} \sum_{j=1}^{M} \frac{\xi_j^{(k)}}{M}\right) = \sum_{k=1}^{K} \frac{\left(v^{(k)}\right)^2}{M} \mathbb{V}ar\left(\xi^{(k)}\right), \tag{5.15}$$

where $\mathbb{V}ar\left(\xi^{(k)}\right)$ can be estimated from the simulated sample. If we use the unbiased version of sample variance for each subregion, Eq. (5.15) gives an unbiased estimate as well.

Additionally, an upper bound for the variance can be derived. Recall that for any subregion both an upper bound and a lower bound for the integral are available. We denote them by $\mathcal{U}^{(k)}$, $\mathcal{L}^{(k)}$ respectively for subregion k and let $\delta^{(k)} = \mathcal{U}^{(k)} - \mathcal{L}^{(k)}$. It is immediate to see that

$$\mathbb{V}ar\left(\xi^{(k)}\right) = \mathbb{E}\left(\xi^{(k)} - \mathbb{E}\left(\xi^{(k)}\right)\right)^2 \leq \mathbb{E}\left(\mathcal{U}^{(k)} - \mathcal{L}^{(k)}\right)^2 = \left(\delta^{(k)}\right)^2.$$

It follows that the upper bound for the variance is

$$\mathbb{V}ar\left(\hat{I}(\xi)\right) \leq \sum_{k=1}^{K} \frac{\left(v^{(k)}\delta^{(k)}\right)^2}{M}, \tag{5.16}$$

which, as $K \to \infty$, approaches zero, the weighted quadratic variation of the continuously differentiable function f.

To reduce the variance we minimize its upper bound. This is achieved by choosing in each step the subregion with the largest $v^{(k)}\delta^{(k)}$ for subdivision. We find empirically it is more robust to rely on $v^{(k)}\delta^{(k)}$ than on the estimated variance, since a large $v^{(k)}\delta^{(k)}$ generally implies a large variance, but the converse does not hold due to simulation noise in the sample variance, esp. for small M. In particular, given any collection of subregions, the choice of subregion for the next subdivision is deterministic and requires no simulation at all. Moreover, the upper bound of variance given by (5.16) is strictly decreasing in the process of subdivision but this is not necessarily the case for the estimated variance.

We furthermore require a subdivision rule replacing the fourth divided differences as in (5.7), since simulated samples cannot be fully symmetric. Suppose subregion k is divided into two subregions k_1 and k_2 in direction i. Its new variance becomes

$$\begin{aligned}
\mathbb{V}ar\left(\hat{I}^{(k)}(\xi)\right) &= \frac{\left(v^{(k_1)}\right)^2}{M}\mathbb{V}ar\left(\xi^{(k_1)}\right) + \frac{\left(v^{(k_2)}\right)^2}{M}\mathbb{V}ar\left(\xi^{(k_2)}\right) \\
&= \frac{\left(v^{(k)}\right)^2}{4M}\left[\mathbb{V}ar\left(\xi^{(k_1)}\right) + \mathbb{V}ar\left(\xi^{(k_2)}\right)\right] \\
&= \frac{\left(v^{(k)}\right)^2}{2M}\left[\mathbb{V}ar\left(\xi^{(k)}\right) - \frac{1}{4}\left(\mathbb{E}\xi^{(k_1)} - \mathbb{E}\xi^{(k_2)}\right)^2\right]. \qquad (5.17)
\end{aligned}$$

To minimize the variance is equivalent to finding the direction i that *maximizes* $\left(\mathbb{E}\xi^{(k_1)} - \mathbb{E}\xi^{(k_2)}\right)^2$. For any simulated sample, we have

$$\left(\mathbb{E}\xi^{(k_1)} - \mathbb{E}\xi^{(k_2)}\right)^2 \approx \frac{4}{M^2}\left(\sum_{y\in k_1}\xi(y) - \sum_{y\in k_2}\xi(y)\right)^2 \qquad (5.18)$$

if exactly $M/2$ points fall in each subregion. To this end we generate random numbers *antithetically*. Since antithetic variates are no longer independent, the variance estimated needs a slight modification. Suppose that ξ and $\bar{\xi}$ are obtained from antithetic pairs, then the variance should be estimated by $M/2$ pairs of averaged antithetic pairs $(\xi + \bar{\xi})/2$, i.e.,

$$\mathbb{V}ar\left(\hat{I}(\xi)\right) = \sum_{k=1}^{K}\frac{\left(v^{(k)}\right)^2}{M/2}\mathbb{V}ar\left(\frac{\xi^{(k)} + \bar{\xi}^{(k)}}{2}\right). \qquad (5.19)$$

We are now able to summarize the algorithm of adaptive Monte Carlo integration for the calculation of tail probability in a multi-factor credit

portfolio loss model. This is presented as Algorithm 5.2. Note that constraint (5.10) used in Algorithm 5.1 is dropped to ensure that (5.14) gives an unbiased estimate.

Algorithm 5.2 adaptive Monte Carlo integration

Generate M antithetic uniform random variables over the integration region,
return I^l, $\mathbb{V}ar(I^l)$, v^l, δ^l and subdivision direction i_s
while termination criteria not met **do**
  Choose the (sub)region with largest $v^l\delta^l$ and divide along direction i_s.
  Compute $f(\mathtt{a})$ and $f(\mathtt{b})$ for the resulting two subregions.
  if $f(\mathtt{a}) = 0$ or $f(\mathtt{b}) = 1$ **then**
    Apply (5.12), let $\mathbb{V}ar(I^l) = \delta^l = 0$.
  else
    Generate M antithetic uniform random variables in both subregions, return I^l, v^l, δ^l and i_s.
  end if
  Update I, $\mathbb{V}ar(I)$ and the subregion collection.
end while

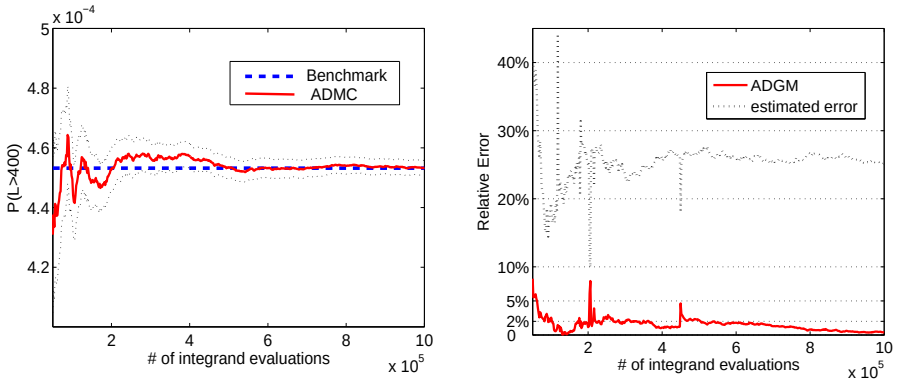
We should finally remark that the adaptive Monte Carlo integration allows flexibility in the choice of M, the number of sample points in each subregion. In terms of accuracy, it is not necessarily inferior to the adaptive algorithm based on fully symmetric interpolation rules like the Genz-Malik rule, although the latter is supposed to provide more accurate approximation for smooth integrands. With a fixed number of samples $N = MK$, the adaptive Monte Carlo integration may choose an M much less than the samples required for the Genz-Malik rule and may therefore obtain many more subregions K.

Numerical results

We continue our numerical experiments with the five-factor model for the portfolio from section 5.3.4 and compare adaptive Monte Carlo integration to plain Monte Carlo integration. Rather than the relative error, we report the standard deviation normalized by benchmark, i.e., $\sqrt{\mathbb{V}ar(\hat{I}(\xi))}/I$.

Figure 5.4(a) shows the estimated tail probability for the loss level $x = 400$ by adaptive Monte Carlo integration along with the corresponding 95% confidence interval. It is evident that the adaptive Monte Carlo integration

indeed gives a convergent estimate with reliable error bounds. By contrast, the error estimate given by the adaptive Genz-Malik rule (based on Eq. (5.6)) can be less reliable. It is shown by Figure 5.4(b) that for the same loss level, although the relative error of the tail probability estimate given by the adaptive Genz-Malik rule is only around 2%, the estimated error by Eq. (5.6) is more than 20%.



(a) (b)

FIGURE 5.4: (a) Tail probability $\mathbb{P}(L > 400)$ computed by adaptive Monte Carlo integration and their corresponding 95% confidence intervals (dotted lines). The dashed line is our Benchmark. (b) Relative errors of the adaptive Genz-Malik rule for $\mathbb{P}(L > 400)$ compared to its associated error estimates (dotted lines) based on Eq. (5.6). The number of integrand evaluations ranges from $50,000$ to 10^6.

We further demonstrate in Figure 5.5 the performance of the adaptive Monte Carlo integration with $M = 10$ for four different loss levels as in section 5.3.4. It comes with no surprise that, just like the adaptive Genz-Malik rule, the adaptive Monte Carlo integration is not sensitive to the portfolio loss level. At the loss level $x = 300$, the accuracy of the adaptive Monte Carlo integration with around 50 thousand integrand evaluations is already comparable to that of the plain Monte Carlo integration with 1 million integrand evaluations, which is a reduction of a factor of 20.

Finally, we would like to point out that the grid generated by the adaptive Monte Carlo integration may also provide a good basis for the calculation of the marginal VaR contributions (VaRC), i.e., $\mathbb{E}(L_i|L = x)$. As an example we present in Table 5.1 the VaRC of the obligors in different buckets for the loss level $x = 300$. The estimates obtained from the adaptive Monte Carlo integration are based on 50 thousand integrand evaluations. The standard deviations (std) are calculated with 20 independent trials and

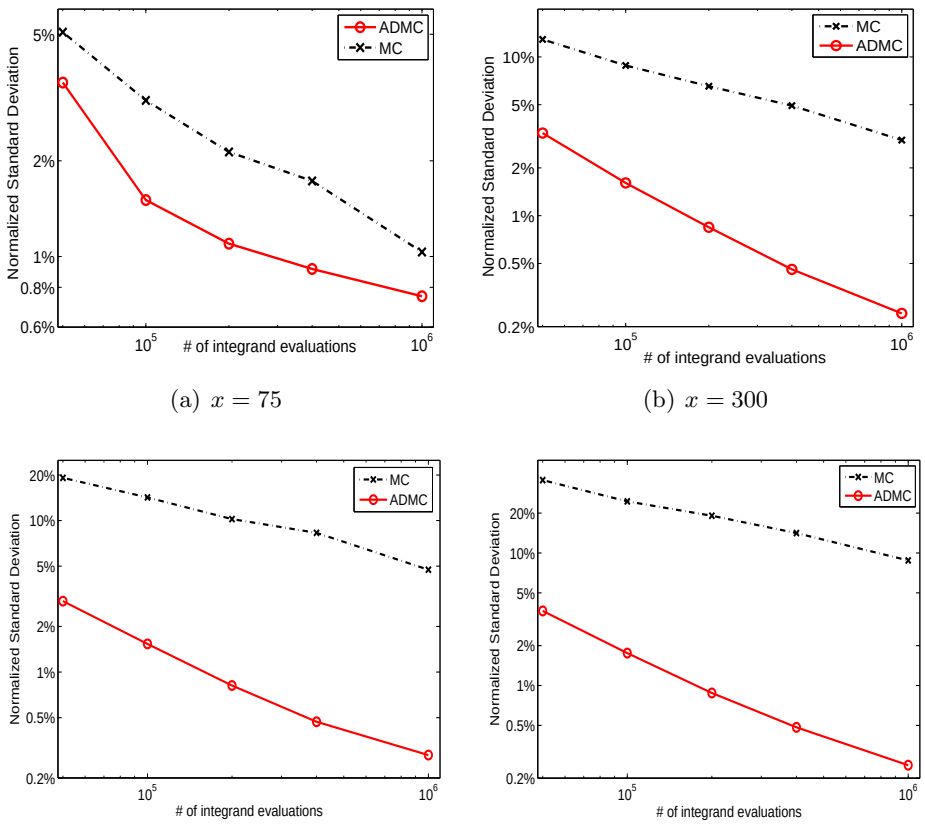
in parentheses are the standard deviations as a percentage of their corresponding benchmark. Both the VaRC estimates and standard deviations are similar to those given by plain Monte Carlo integration with 1 million integrand evaluations. This is in line with the performance regarding the tail probability.

bucket	BM	MC	std	ADMC	std
1	0.4331	0.4258	0.0239 (5.5%)	0.4293	0.0163 (3.8%)
2	0.4498	0.4504	0.0141 (3.1%)	0.4489	0.0127 (2.8%)
3	0.3467	0.3526	0.0167 (4.8%)	0.3475	0.0120 (3.5%)
4	0.2022	0.2037	0.0129 (6.4%)	0.2076	0.0157 (7.8%)
5	0.0683	0.0676	0.0089 (13.0%)	0.0667	0.0069 (10.1%)

TABLE 5.1: The VaR contributions of the obligors in different buckets for the loss level $x = 300$. The adaptive Monte Carlo (ADMC) integration uses 50 thousand integrand evaluations and plain Monte Carlo (MC) integration uses 1 million integrand evaluations. The standard deviations (std) are calculated with 20 independent trials and in parentheses are the standard deviations normalized by benchmark.

5.5 Conclusions

In this chapter we propose algorithms of adaptive integration for the calculation of the tail probability in multi-factor credit portfolio loss models. We showed that under mild conditions, the conditional tail probability, as a function of the common factors, is monotone and differentiable. The algorithms devised heavily rely on this. We modify the adaptive Genz-Malik rule so that it becomes suitable for portfolio credit models with a number of factors $2 \leq d \leq 8$. The algorithm based on the Genz-Malik rule is asymptotically convergent and particularly attractive for large loss levels. It consistently outperforms the plain Monte Carlo and quasi-Monte Carlo methods in terms of approximation error. Finally we arrive at the adaptive Monte Carlo integration, which essentially replaces the Genz-Malik rule by antithetic random numbers. The algorithm is advantageous in that it can handle higher-dimensional models and is able to provide reliable probabilistic error bounds. In summary, especially for higher-dimensional problems the adaptive Monte Carlo method seems the clear favorite, whereas for lower-dimensional problems both adaptive methods, the deterministic and the Monte Carlo version, work very well.



(a) $x = 75$ (b) $x = 300$

(c) $x = 400$ (d) $x = 550$

FIGURE 5.5: Standard deviations of the tail probability estimates given by plain Monte Carlo (MC) and Adaptive Monte Carlo (ADMC) for four loss levels. Standard deviations are reported as a percentage of the respective tail probabilities. For plain MC standard deviations are computed based on 100 independent runs of simulation and for ADMC, standard deviations are estimated by Eq. (5.15.)

Chapter 6

Nonnegative Matrix Factorization of a Correlation Matrix

6.1 Introduction

An assumption on nonnegative coefficients in the multi-factor model is made in Chapter 5. The non-negativity requirement is beneficial since it greatly facilitates the computation of the tail probability of portfolio loss (5.1) in the multi-factor portfolio loss models. In this chapter we present a dedicated algorithm that takes care of this assumption.

Suppose we are given the correlation matrix $\boldsymbol{C}$ of $\Psi_1, \ldots, \Psi_M$ in the first level factor model given by (1.12). The Assumption 5.1 can be achieved if the coefficients a_{ij} in (1.14) are all nonnegative. Usual dimension reduction techniques such as principle component analysis (PCA) do not guarantee to lead to nonnegative coefficients. Moreover, since the correlation of Ψ_i and Ψ_j is given by $\varrho(\Psi_i, \Psi_j) = \boldsymbol{a}_i(\boldsymbol{a}_j)^T$ and $\varrho(\Psi_i, \Psi_i)$ must be 1, we have the hard constraint that

$$\boldsymbol{a}_i(\boldsymbol{a}_i)^T = 1. \tag{6.1}$$

This cannot be expected either by employing a straightforward PCA. As a result a different matrix decomposition, which takes care of both (6.1) and nonnegativity of coefficients $\boldsymbol{a}_i$ for all i, needs to be developed. This is the contents of the present chapter.

Let us restate the problem of interest in the present chapter in matrix form. For a given correlation matrix $\boldsymbol{C}_{M\times M}$ with $C_{ij} \geq 0$ and given $m \leq M$,

find a matrix $\boldsymbol{A}_{M\times m}$ that minimizes the Frobenius norm of the matrix

$$\boldsymbol{C}-\boldsymbol{A}\boldsymbol{A}^T,$$

subject to $\boldsymbol{A}_{ij} \geq 0$ and $(\boldsymbol{A}\boldsymbol{A}^T)_{ii} = 1$. The Frobenius norm is the norm usually used for similar problems, see eg. Lee & Seung (1999), Zhang & Wu (2003). Two alternatives are the weighted Frobenius norm or the Kullback-Leibler divergence. More choices can be found in Berry et al. (2007).

Our problem falls in the class of *nonnegative matrix factorization* $\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}$ and further requires that $\boldsymbol{W}^T = \boldsymbol{H}$. Nonnegative Matrix Factorization (NMF) algorithms aim to find for a matrix $\boldsymbol{V}$ two matrix factors such that $\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}$, where $\boldsymbol{W}$ and $\boldsymbol{H}$ are both nonnegative matrices, i.e., all elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ are equal to or greater than zero. The non-negativity constraint arises often naturally in applications in physics and engineering.

The notion of Nonnegative Matrix Factorization originates from Lee & Seung (1999), where simple multiplicative update rules were introduced to solve the approximation problem. Since then, different aspects of NMF, such as its analysis or the extension of the algorithms to various applications have been extensively investigated. For a recent review see Berry et al. (2007).

NMF with the additional constraint $\boldsymbol{W}^T = \boldsymbol{H}$, called *symmetric nonnegative factorization*, has already been treated in Vandendorpe et al. (2008). They discussed another application to portfolio credit risk and in particular, in the framework of CreditRisk$^+$. The problem considered in their article is very similar to ours but the diagonal elements of $\boldsymbol{V}$ are not constrained to be 1.

Viewed from a different perspective our problem of interest is essentially finding the nearest low-rank correlation matrix, but with an additional nonnegativity constraint. The nearest low-rank correlation matrix question has drawn broad attention in the financial community. A variety of methods have been proposed. To name a few, the geometric programming approach (Groenen & Pietersz 2007), the Lagrange multiplier method (Zhang & Wu 2003, Wu 2003), and majorization (Pietersz & Groenen 2004) have been introduced for this purpose. However, none of these approaches accommodate an additional non-negativity constraint.

6.2 An algorithm for nonnegative factorization

Now we return to the problem of finding the best nonnegative factor with respect to the Frobenius norm. So, with $\boldsymbol{E} = \boldsymbol{C}-\boldsymbol{A}\boldsymbol{A}^T$, we need to minimize

$$\|\boldsymbol{E}\|^2_{\text{frob}} = \sum_{k,l} E^2_{k,l} \text{ with } \text{diag}(\boldsymbol{E}) = \boldsymbol{O}.$$

We aim to produce an approximated nonnegative factorization of $\boldsymbol{C}$. Attempts with standard constrained optimization programs, such as Rosen's *gradient projection method* (Rosen 1960) did not converge at all, most likely because of the non-convexity of the problem. An interesting algorithm for the problem without the unit norm condition can be found in Pauca et al. (2006); interesting theory is then available in Catral et al. (2004). However, the unit norm condition is essential in our case, and we have to develop an algorithm to deal with it. Therefore, we approach the problem here by a relaxation technique.

Suppose we have an approximation for $\boldsymbol{A}$, satisfying the constraints, then we aim at improving the rows of $\boldsymbol{A}$, one by one. Let $\boldsymbol{c}^j$ denote the j-th column of $\boldsymbol{C}$, we solve:

$$\text{Minimize } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|_2, \quad \text{with} \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \|\boldsymbol{x}\| = 1. \tag{6.2}$$

Vector $\boldsymbol{x}$ is meant to be an improved version of $\boldsymbol{a}_j$. Replacement of the row $\boldsymbol{a}_j$ by $\boldsymbol{x}$, should result in an improved approximation.

Also for this subproblem, the application of standard constrained optimization methods did not show any success (although we did not experiment too extensively). We expect a fundamental reason behind the failure of these techniques.

The requirement $\boldsymbol{x} \geq \boldsymbol{0}$ typically gives rise to a linear programming problem, whereas the constraint $\|\boldsymbol{x}\| = 1$ asks for an analytic approach, such as a procedure with Lagrange multipliers. These two techniques appear to be not on speaking terms, so we split the treatment of the two constraints, by setting up a new approach, based on relatively basic numerical tools.

6.2.1 The non-negativity conditions

Let the object function, Θ, and its gradient, $\boldsymbol{g}$, be defined by:

$$\Theta(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|^2, \ \boldsymbol{g} = \frac{\partial \Theta}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j). \tag{6.3}$$

The non-negativity condition usually leads to an approximate solution for which the elements x_k can be divided in a so-called feasible set, $\mathcal{H}$, and its complement, the non-feasible set. The solution satisfies

$$\begin{aligned} x_k > 0,\ g_k = 0, \qquad & k \in \mathcal{H}, \\ x_k = 0,\ g_k \geq 0, \qquad & k \notin \mathcal{H}. \end{aligned}$$

If $\boldsymbol{x}$ and $\boldsymbol{g}$ satisfy these complementary relations, we obviously have reached a *local* minimum. By a convexity argument, it is easily proved that this minimum is in fact an absolute minimum.

A well-established algorithm for the nonegative least squares problem (i.e. (6.2) without the unit norm constraint) is given by Lawson & Hanson (1974). The algorithm is implemented in Matlab by the name of **lsqnonneg**. This routine failed rather often, uttering protesting remarks about tolerances that were too severe, or not severe enough.

In the Matlab routine **lsqnonneg** the feasible set is determined in a trial-and-error procedure similar to the classical Simplex method. This requires the least squares solution of (restricted variants of) the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}$. For reasons of numerical stability, these systems are solved using the pseudo-inverse of the restricted matrix. Because our problems have moderate size ($M \approx 50$), this rather expensive way of operating does not seem necessary. Therefore we replaced the pseudo-inverse approach in **lsqnonneg** by a basic direct solution, using Matlab's formal solution method `x=A\c`.

Also we searched for a shorter path to the solution, by choosing a significantly more efficient strategy for updating the feasible set. Suppose we have a temporary feasible set $\mathcal{T}$. The most successful variant of the nonnegative least squares algorithm proceed as follows:

Algorithm 6.1 fast nonnegative least squares

repeat
  Solve the system restricted to $\mathcal{T}$,
  Set $x_k = 0$ outside this set,
  Remove the indices $k \in \mathcal{T}$ for which x_k is negative,
  Add indices $l \notin \mathcal{T}$ to set $\mathcal{T}$ if $g_l < 0$,
until No further updates can be made.

When this process terminates, we should have found the nonnegative least squares solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}$. The algorithm is usually substantially faster than **lsqnonneg**. Ocassionally the algorithm stalls. In such cases we replace it by **lsqnonneg**.

6.2.2 Unit norm condition

Let $\boldsymbol{e}^1, \boldsymbol{e}^2, \ldots, \boldsymbol{e}^p$ be the Cartesian unit vectors in $\mathbb{R}^p$, and let $\boldsymbol{A}$ be a real nonnegative $M \times p$ matrix. Suppose we have found a solution $\boldsymbol{x}$ to the nonnegative least squares problem of minimizing $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ under the condition $\boldsymbol{x} \geq \boldsymbol{0}$. To this solution corresponds a feasible subset $\mathcal{H}$ and a corresponding 'feasible subspace' $\mathcal{E} = \text{span}(\cup_{k \in \mathcal{H}} \{\boldsymbol{e}^k\})$.

Let $\widetilde{\boldsymbol{A}}$ be the matrix consisting of the columns of $\boldsymbol{A}$ with indices in $\mathcal{H}$,

and let $\widetilde{\boldsymbol{x}}$ be the feasible, i.e. the nonzero, part of $\boldsymbol{x}$, then $\widetilde{\boldsymbol{x}}$ is the ordinary least squares solution to $\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$.

The intersection of the unit sphere in $\mathbb{R}^p$ with the subspace $\mathcal{E}$ is the unit sphere in $\mathcal{E}$. If we succeed in finding a solution $\widehat{\boldsymbol{x}} \in \mathcal{E}$, subject to the condition $\|\widehat{\boldsymbol{x}}\| = 1$, the extension of $\widehat{\boldsymbol{x}}$ to $\mathbb{R}^p$ (by choosing $x_k = 0$ for the non-feasible components) will approximate a solution of problem (6.2) well. Only a few issues may hamper the convergence of this algorithm:

1. A feasible variation could exist in which Θ decreases.
2. The solution $\widetilde{\boldsymbol{x}}$ in $\mathcal{E}$ may have one or more negative components. This may happen if the hyper-ellipsoids $\Theta(\widetilde{\boldsymbol{x}}) = \text{const} = \overline{C}$ have very different axes, and very skew orientations.

A robust remedy for these issues requires some more attention in a future variant. For now we neglect these.

Analysis of the unit-norm problem.

We now concentrate on the problem of minimizing $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|$ under the condition $\|\boldsymbol{x}\| = 1$. In this analysis the nonnegativity does not play a role, since everything happens in a feasible subspace $\mathcal{E}$ that is already obtained. For convenience, we drop the tilde signs etc.

In a more familiar setting, in which the second constraint reads $\|\boldsymbol{x}\| \leq 1$, the domain is clearly convex. In our case, however, we have to satisfy $\|\boldsymbol{x}\| = 1$ and this domain is not convex. Many local minima and even saddlepoint solutions could be found.

Let $\widehat{\boldsymbol{x}}$ be the solution of the unconstrained least squares problem. Define for convenience $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A}$. The object function, Θ, can then be defined by:

$$\begin{aligned} \Theta(\boldsymbol{x}) &= \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \|\boldsymbol{b}\|^2 \\ &= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^T\boldsymbol{B}(\boldsymbol{x} - \widehat{\boldsymbol{x}}) + \Theta(\widehat{\boldsymbol{x}}). \end{aligned} \tag{6.4}$$

The usual least squares solution is determined by the normal equations $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$, which is equivalent to setting the gradient of Θ to zero:

$$\boldsymbol{g} = \frac{\partial\Theta}{\partial\boldsymbol{x}} = 2(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{b}) = \boldsymbol{0}$$

The hyper-surface $\Theta(\boldsymbol{x}) = \overline{C}$ is an ellipsoid in a p-dimensional space, centered around $\widehat{\boldsymbol{x}}$. If $\overline{C}$ is very small, this surface is either completely in the interior of the unit sphere, or completely outside. If we let $\overline{C}$ grow, then

eventually the surface will have contact to the sphere $\|\boldsymbol{x}\| = 1$, and this contact will typically be a one point contact. Since both surfaces are smooth, the normal directions have to be the same in the inside situation, or opposite in the outside case. Therefore the gradient of Θ should be proportional to $\boldsymbol{x}$: So, we need to have

$$\boldsymbol{g} = 2(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{b}) = 2\alpha\boldsymbol{x}, \tag{6.5}$$

for some scalar α. Solution of (6.5) for a given α formally yields

$$\boldsymbol{x}(\alpha) = (\boldsymbol{B} - \alpha\boldsymbol{I})^{-1}\boldsymbol{c},$$

with $\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{b}$. Then the requirement $\|\boldsymbol{x}\| = 1$ reads

$$\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-2}\boldsymbol{c} = 1, \tag{6.6}$$

which is a nonlinear equation in α. This equation can be solved by several techniques, provided a good initial estimate is available.

Probably several solutions can be found, some of which will obviously not make sense, but certainly a choice should be made. First we obtain insight in the proper choice of solution from a geometric point of view.

Geometric consideration

Let's assume that $\boldsymbol{x}$ is the point of first contact between $\Theta(\boldsymbol{x}) = \overline{C}$ and $\|\boldsymbol{x}\| = 1$, then we must have $\Theta(\widetilde{\boldsymbol{x}}) \geq \Theta(\boldsymbol{x})$, for all $\widetilde{\boldsymbol{x}}$ with $\|\widetilde{\boldsymbol{x}}\| = 1$. Now, let $\widetilde{\boldsymbol{x}} = \tau\boldsymbol{x} + \sigma\boldsymbol{t}$ be a vector with $\|\widetilde{\boldsymbol{x}}\| = 1$, with

$$\tau = \cos(\varphi),\ \sigma = \sin(\varphi),\ \boldsymbol{t}^T\boldsymbol{x} = 0,\ \|\boldsymbol{t}\| = 1. \tag{6.7}$$

Then, we should have

$$\begin{aligned}&\Theta(\boldsymbol{x} + (\tau - 1)\boldsymbol{x} + \sigma\boldsymbol{t}) - \Theta(\boldsymbol{x})\\ &\quad = \boldsymbol{g}^T((\tau - 1)\boldsymbol{x} + \sigma\boldsymbol{t}) + \big((\tau - 1)^2\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} + 2(\tau - 1)\sigma\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{x} + \sigma^2\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t}\big) \geq 0,\end{aligned}$$

for all $\boldsymbol{t} \perp \boldsymbol{x}$, with $|\tau|$ and $|\sigma|$ sufficiently small.

Now for small σ, we have $\tau - 1 = -\frac{1}{2}\sigma^2 + O(\sigma^4)$ from (6.7), and using $\boldsymbol{g}^T\boldsymbol{t} = 2\alpha\boldsymbol{x}^T\boldsymbol{t} = 0$, we get

$$-\frac{1}{2}\sigma^2\boldsymbol{g}^T\boldsymbol{x} + \sigma^2\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t} + O(\sigma^3) \geq 0.$$

Dividing by σ^2, letting $\sigma \to 0$, and using $\boldsymbol{g}^T\boldsymbol{x} = 2\alpha\|\boldsymbol{x}\|^2 = 2\alpha$, we finally obtain

$$-\alpha + \boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t} \geq 0$$

for all $\boldsymbol{t} \perp \boldsymbol{x}$, with $\|\boldsymbol{t}\| = 1$. This is equivalent to $\alpha \leq \mathcal{R}(\boldsymbol{B}, \boldsymbol{y})$, for each $\boldsymbol{y} \perp \boldsymbol{x}$, where $\mathcal{R}$ denotes the Rayleigh quotient function corresponding to $\boldsymbol{B}$.

Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of $\boldsymbol{B}$, then according to a theorem by Rayleigh, a $\theta \in (0, 1)$ exists such that:

$$\min_{\boldsymbol{t} \perp \boldsymbol{x}} \frac{\boldsymbol{t}^T \boldsymbol{B} \boldsymbol{t}}{\boldsymbol{t}^T \boldsymbol{t}} = \lambda_1 + \theta(\lambda_2 - \lambda_1).$$

Hence we have that $\alpha < \lambda_1 + \theta(\lambda_2 - \lambda_1)$ for some $\theta \in (0, 1)$. It follows that for our first contact, $\alpha < \lambda_2$ in any case. Next, it will be shown that the optimal value for α is the *smallest solution* of equation (6.6), and satisfies $\alpha < \lambda_1$.

Algebraic consideration

Let $\boldsymbol{x}(\alpha)$ be the solution of (6.5), for a given α. We should find α such that $\|\boldsymbol{x}(\alpha)\|^2 = 1$. From all values of α for which $\|\boldsymbol{x}(\alpha)\|^2 = 1$, we should select the value for which $\Theta(\boldsymbol{x}(\alpha))$ is minimal. Define $\boldsymbol{x}(\alpha)$, $F(\alpha)$ and $N(\alpha)$ by

$$\begin{aligned} \boldsymbol{B}\boldsymbol{x}(\alpha) &= \alpha \boldsymbol{x}(\alpha) + \boldsymbol{c}, && (6.8) \\ N(\alpha) &= \boldsymbol{x}(\alpha)^T \boldsymbol{x}(\alpha), && (6.9) \\ F(\alpha) &= \boldsymbol{x}(\alpha)^T \boldsymbol{B} \boldsymbol{x}(\alpha) - 2\boldsymbol{x}(\alpha)^T \boldsymbol{c} + \boldsymbol{b}^T \boldsymbol{b}. && (6.10) \end{aligned}$$

then we must determine

$$\min_{N(\alpha)=1} F(\alpha).$$

The following theorem provides us with a simple choice for the α-value for which F is minimal.

Theorem 6.1. *Let $\boldsymbol{B}$ be a symmetric positive definite matrix. Let $\boldsymbol{x}(\alpha)$, $F(\alpha)$ and $N(\alpha)$ be defined by (6.8), (6.9), (6.10), respectively. Then, for each pair α and β with $\alpha \neq \beta$ and $N(\alpha) = N(\beta)$, the following inequality holds:*

$$\frac{F(\alpha) - F(\beta)}{\alpha - \beta} > 0. \tag{6.11}$$

Proof. The expression for $\boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}$ can be simplified by left multiplication of (6.8) by $\boldsymbol{x}(\alpha)^T$:

$$\boldsymbol{x}(\alpha)^T \boldsymbol{B} \boldsymbol{x}(\alpha) = \alpha \boldsymbol{x}(\alpha)^T \boldsymbol{x}(\alpha) + \boldsymbol{x}(\alpha)^T \boldsymbol{c} = \alpha N(\alpha) + \boldsymbol{x}(\alpha)^T \boldsymbol{c},$$

and therefore $F(\alpha)$ can be written as $F(\alpha) = \alpha N(\alpha) - \boldsymbol{x}(\alpha)^T\boldsymbol{c} + \boldsymbol{b}^T\boldsymbol{b}$. Then, we may write

$$F(\alpha) - F(\beta) = \alpha N(\alpha) - \beta N(\beta) - [\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}.$$

If $N(\alpha) = N(\beta)$ this is equivalent to

$$F(\alpha) - F(\beta) = (\alpha - \beta)\left[\frac{1}{2}N(\alpha) + \frac{1}{2}N(\beta)\right] - [\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}.$$

An expression for $[\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}$ can be obtained by left multiplication of (6.8) by $\boldsymbol{x}(\beta)$, yielding

$$\boldsymbol{x}(\beta)^T\boldsymbol{B}\boldsymbol{x}(\alpha) = \alpha\boldsymbol{x}(\beta)^T\boldsymbol{x}(\alpha) + \boldsymbol{x}(\beta)^T\boldsymbol{c}. \tag{6.12}$$

Interchanging α and β in this equation leaves the inner product $\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)$ unchanged. Also the left-hand side does not change, since $\boldsymbol{B}$ is symmetric. By subtracting the interchanged variant of (6.12) from (6.12) itself, we therefore get:

$$0 = (\alpha - \beta)\boldsymbol{x}(\beta)^T\boldsymbol{x}(\alpha) + [\boldsymbol{x}(\beta) - \boldsymbol{x}(\alpha)]^T\boldsymbol{c},$$

from which it follows that

$$[\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c} = (\alpha - \beta)\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta).$$

So, the *F-difference* can be written as

$$F(\alpha) - F(\beta) = (\alpha - \beta)\left[\frac{1}{2}N(\alpha) + \frac{1}{2}N(\beta) - \boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)^T\right].$$

Substituting $\|\boldsymbol{x}(\alpha)\|^2$ for $N(\alpha)$, etc, we finally arrive at

$$\begin{aligned} F(\alpha) - F(\beta) &= \frac{1}{2}(\alpha - \beta)[\|\boldsymbol{x}(\alpha)\|^2 + \|\boldsymbol{x}(\beta)\|^2 - 2\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)] \\ &= \frac{1}{2}(\alpha - \beta)\|\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)\|^2, \end{aligned}$$

which implies (6.11). □

According to this theorem, we should search for the smallest solution of the equation $N(\alpha) = 1$.

Now consider the explicit formula (6.6) for $\|\boldsymbol{x}(\alpha)\|^2$,

$$N(\alpha) = \|\boldsymbol{x}(\alpha)\|^2 = \boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-2}\boldsymbol{c}.$$

Differentiating $N(\alpha)$ with respect to α gives $N'(\alpha) = 2\boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-3}\boldsymbol{c}$. If $\alpha < \lambda_1$, $\boldsymbol{B}-\alpha\boldsymbol{I}$ is a positive definite matrix, and so is $(\boldsymbol{B}-\alpha\boldsymbol{I})^{-3}$. Therefore N is monotonically increasing for $\alpha < \lambda_1$. Now,

$$\begin{aligned} N(\alpha) &\to 0 \qquad \text{as} \quad \alpha \to -\infty, \\ N(\alpha) &\to +\infty \quad \text{as} \quad \alpha \uparrow \lambda_1. \end{aligned}$$

Hence, there is precisely one value $\alpha < \lambda_1$ for which $N(\alpha) = 1$, and this value is the smallest solution of equation (6.6).

This makes it relatively easy to determine an initial guess for the iterative solution procedure.

6.2.3 Algorithm

Here is a global description of the resulting algorithm for the nonnegative matrix factorization of $\boldsymbol{C}$ into $\boldsymbol{A}$ and $\boldsymbol{A}^T$ subject to $\boldsymbol{a}_i(\boldsymbol{a}_i)^T = 1$:

- If $\boldsymbol{A}$ is an approximate nonnegative factor, with $\text{diag}(\boldsymbol{A}\boldsymbol{A}^T) = \boldsymbol{I}$, then the main step in the algorithm is the replacement of all rows of $\boldsymbol{A}$ by improved versions.
- The *replacement of a row*, $\boldsymbol{a}_j$, by $\boldsymbol{x}$ requires the solution of the following expression:

$$\text{Minimize } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|^2, \text{ subject to: } \boldsymbol{x} \geq \boldsymbol{0}, \ \|\boldsymbol{x}\| = 1.$$

  The solution of this norm-restricted, nonnegative least squares problem for row replacement is done in two steps:

  1. Find a nonnegative least squares solution, $\boldsymbol{x}_c$, with algorithm 6.1. If this process is not finite (and thus periodic), choose the Matlab procedure **lsqnonneg** instead.
     The indices of the nonzero entries of $\boldsymbol{x}_c$ build the feasible set $\mathcal{H}$. Suppose that this set has p elements.
  2. Let $\widehat{\boldsymbol{B}}$ the $p \times p$ sub-matrix of $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A}$, obtained by removing the rows and columns with indices that are not in $\mathcal{H}$. Let $\widehat{\boldsymbol{c}}$ be the $\mathcal{H}$-*restriction* of $\boldsymbol{A}^T\boldsymbol{c}^j$. Determine by means of Newton's procedure the minimal solution of the expression $\|(\widehat{\boldsymbol{B}}-\alpha\boldsymbol{I})^{-1}\widehat{\boldsymbol{c}}\|^2$, similar to the one described in Golub & van Loan (1996) (Chapter 12.1). If the restricted solution is not feasible, as it gives rise to negative entries, remove the incorrect indices from $\mathcal{H}$, and repeat the procedure. If the solution is not minimal if observed in the complete space, accept this sub-optimal solution. Continue by returning to step 1 with the next row.

As the initial guess for the factorization, we recommend to start with a nonnegative random matrix A, with its elements between 0 and 1 and with $|\boldsymbol{a}_i| = 1$, as this is easily generated. A more sophisticated initial guess, i.e., a Choleski factorisation, where the negative entries were replaced by zeros, did not lead to improved convergence in general.

The iteration process is stopped by an Aitken error estimate: Let $r_n = \|\boldsymbol{E}_n\|_{\text{frob}}$, and assume $r_n \to r$, then the Aitken estimate of $r_n - r$ reads

$$r_n - r \approx \frac{(r_n - r_{n-1})^2}{r_n - 2r_{n-1} + r_{n-2}}, \tag{6.13}$$

which is based on the hypothesis of linear convergence of the sequence $\{r_n\}$:

$$r_n \approx r + \text{ const } \rho^n. \tag{6.14}$$

This criterion appears to be useful, also in cases where the hypothesis (6.14) is not valid.

It is not easily possible to give definitive statements about the computational effort of the method presented, as its performance varies in practice. The solution to the nonnegative least squares problem requires $4Mm^2$ flops per iteration step, and the number, k_i, of iteration steps varies from approximately 10 to 50, in our tests. The number of problems to be solved for a complete matrix update is M. The equality constraint is not relevant for the computational cost. If we assume that a number, k_o, of matrix updates (outer iterations) is required, then total work is $W \approx 4k_o k_i M^2 m^2$, where k_i and k_o are average values. Practically, k_o may differ from $O(10)$ to even $O(100)$, for 8 digits accuracy, at different problems of similar size. A lower accuracy requirement will lead to a faster convergence. We present some experiments with 8 digits accuracy, resulting in highly satisfactory convergence, in the next section.

6.3 Numerical results

In this section we present some factorization results obtained with the algorithm presented in the previous section. All experiments are performed in Matlab, version 7.4, on an Intel(R) Core (TM) 2 6700 2.66 GHz processor.

The first example is taken from Zhang & Wu (2003). The correlation matrix $\boldsymbol{C}$ is presented in Appendix 6.A and illustrated graphically on a two-dimensional grid in Figure 6.1(a), where the colored surface is formed by the c_{ij} in matrix $\boldsymbol{C}$. Figure 6.1(b)-(d) correspond to the nonnegative

low-rank approximations AA^T, with rank m =2, 3 and 6, respectively. The results obtained are comparable to those in Zhang & Wu (2003), where the non-negativity constraint was not imposed.

The convergence in the Frobenius norm with increasing rank is presented in Figure 6.2(a). With regard to the computation time, an approximation can be found within one second for all $m \leq 11$. Figure 6.2(b) illustrates the 6 largest eigenvalues of $\boldsymbol{C}$ and its nonnegative low-rank approximation with rank $m = 6$.

In our second example $\boldsymbol{C}$ is a 50×50 correlation matrix with its entries defined to be

$$c_{ij} = LongCorr + (1 - LongCorr)e^{\kappa|i-j|},$$

$$\kappa = d_1 - d_2 \max(i, j),$$

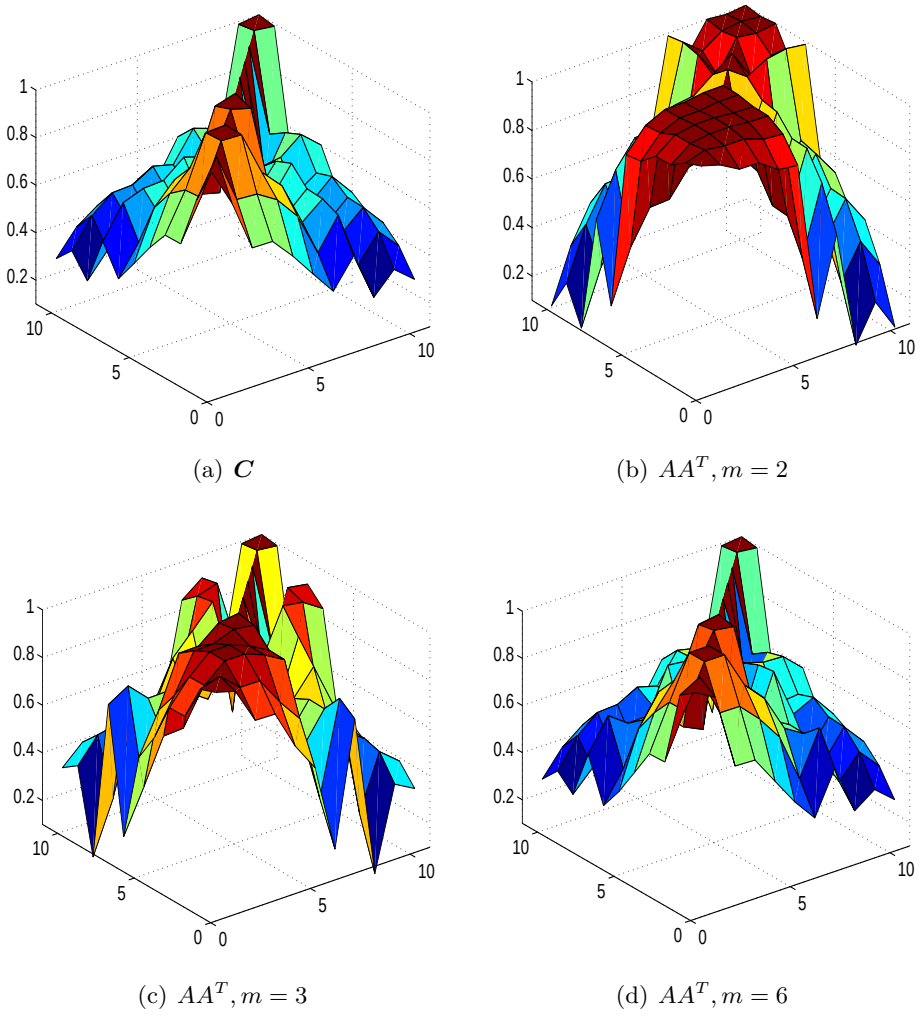
where $LongCorr = 0.3$, $d_1 = -0.12$, $d_2 = 0.005$.

Figure 6.3(a) shows the convergence in the Frobenius norm with increasing rank. Figure 6.3(b) compares the 5 largest eigenvalues of $\boldsymbol{C}$ to their nonnegative low-rank approximations with rank $m = 5, 10$. As the size of the matrix is significantly bigger than the one in the previous example, this computation is more costly. The CPU time for finding the nonnegative low-rank approximations for $m = 1, \cdots, 30$ is on average around 10 seconds. Moreover, Figure 6.3 indicates that as the rank of $\boldsymbol{A}$ increases, the CPU time for finding the nonnegative low-rank approximations also tends to increase.

Finally we consider the use of the NMF in the credit portfolio loss context. We work with a portfolio of 5000 obligors each with default probability $p = 0.01$ and exposure $w = 1$. The full model in the form of (1.12) is a 50 factor model with correlation matrix of $\boldsymbol{\Psi}$ given in our second example. The coefficients γ_{ij} are set randomly to be either 0 or 0.5 with $\sum_{j=1}^{M} \gamma_{ij} = 1.5$ for all i. In Figure 6.5 we show the loss distribution of the portfolio from the full model and from the simplified models based on NMF with rank $m = 6, 10$, with excellent agreement.

6.4 Conclusions

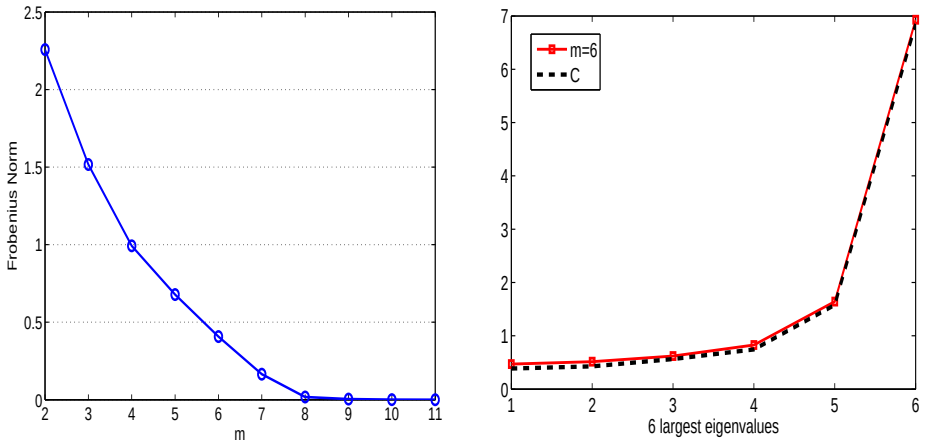
We have presented a dedicated algorithm for the nonnegative factorization of a correlation matrix. The algorithm is based on a two-step procedure. First the non-negativity constraint is dealt with, by means of basic nonnegative least-squares routines, available in Matlab. Secondly, the unit norm



(a) $\boldsymbol{C}$ (b) $AA^T, m = 2$

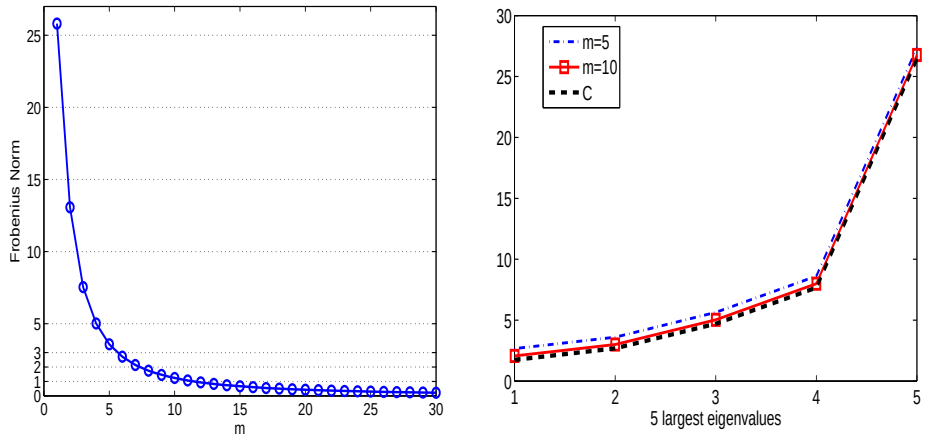
(c) $AA^T, m = 3$ (d) $AA^T, m = 6$

FIGURE 6.1: (a) The correlation structure of $\boldsymbol{C}$ displayed on a 2D grid, (b)-(d) The correlation structure of the nonnegative low-rank approximations to $\boldsymbol{C}$. The ranks of matrices shown in (b)-(d) are m =2, 3 and 6 respectively.



(a) (b)

FIGURE 6.2: (a) Convergence in the Frobenius norm with increasing rank m. (b) The 6 largest eigenvalues of matrix $\boldsymbol{C}$ and its nonnegative low-rank approximations ($m = 6$).



(a) (b)

FIGURE 6.3: (a) Convergence in the Frobenius norm with increasing rank m. (b) The 5 largest eigenvalues of matrix $\boldsymbol{C}$ and its nonnegative low-rank approximations ($m = 5, 10$).

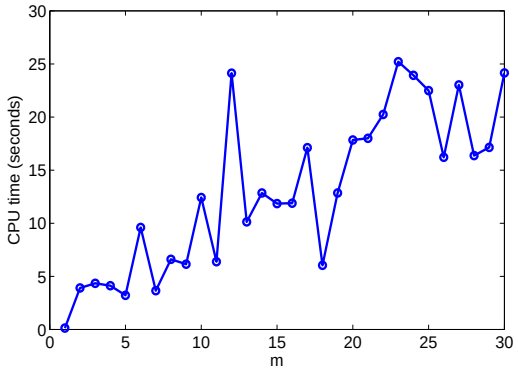


FIGURE 6.4: CPU time with increasing rank m.

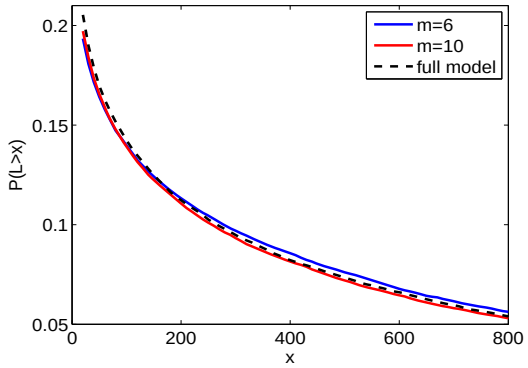


FIGURE 6.5: Portfolio loss distributions from the full model and from the simplified models based on NMF with rank $m = 6, 10$.

condition is taken into account. The algorithm comes with a detailed explanation of all its steps. The methods works well, as is confirmed by some numerical experiments.

6.A Correlation matrix: example one

$$\begin{pmatrix}
1.0000 & 0.8415 & 0.6246 & 0.6231 & 0.5330 & 0.4287 & 0.3274 & 0.4463 & 0.2439 & 0.3326 & 0.2625 \\
0.8415 & 1.0000 & 0.7903 & 0.7844 & 0.7320 & 0.6346 & 0.4521 & 0.5812 & 0.3439 & 0.4533 & 0.3661 \\
0.6246 & 0.7903 & 1.0000 & 0.9967 & 0.8108 & 0.7239 & 0.5429 & 0.6121 & 0.4426 & 0.5189 & 0.4251 \\
0.6231 & 0.7844 & 0.9967 & 1.0000 & 0.8149 & 0.7286 & 0.5384 & 0.6169 & 0.4464 & 0.5233 & 0.4299 \\
0.5330 & 0.7320 & 0.8108 & 0.8149 & 1.0000 & 0.9756 & 0.5676 & 0.6860 & 0.4969 & 0.5734 & 0.4771 \\
0.4287 & 0.6346 & 0.7239 & 0.7286 & 0.9756 & 1.0000 & 0.5457 & 0.6583 & 0.4921 & 0.5510 & 0.4581 \\
0.3274 & 0.4521 & 0.5429 & 0.5384 & 0.5676 & 0.5457 & 1.0000 & 0.5942 & 0.6078 & 0.6751 & 0.6017 \\
0.4463 & 0.5812 & 0.6121 & 0.6169 & 0.6860 & 0.6583 & 0.5942 & 1.0000 & 0.4845 & 0.6452 & 0.5673 \\
0.2439 & 0.3439 & 0.4426 & 0.4464 & 0.4969 & 0.4921 & 0.6078 & 0.4845 & 1.0000 & 0.6015 & 0.5200 \\
0.3326 & 0.4533 & 0.5189 & 0.5233 & 0.5734 & 0.5510 & 0.6751 & 0.6452 & 0.6015 & 1.0000 & 0.9889 \\
0.2625 & 0.3661 & 0.4251 & 0.4299 & 0.4771 & 0.4581 & 0.6017 & 0.5673 & 0.5200 & 0.9889 & 1.0000
\end{pmatrix}$$

Chapter 7

Saddlepoint Approximations for Expectations

7.1 Introduction

We consider the saddlepoint approximations of $\mathbb{E}[(X-K)^+]$ and $\mathbb{E}[X|X \geq K]$, where X is the sum of n independent random variables X_i, $i = 1, \ldots, n$, and K is a known constant. These two expectations can be frequently encountered in finance and insurance. In option pricing, $\mathbb{E}[(X-K)^+]$ is the payoff of a call option (Rogers & Zane 1999). It also plays an integral role in the pricing of the Collateralized Debt Obligations (CDO) (Yang et al. 2006, Antonov et al. 2005). In insurance, $\mathbb{E}[(X-K)^+]$ is known as the stop-loss premium. The term $\mathbb{E}[X|X \geq K]$ corresponds to the expected shortfall, also known as the tail conditional expectation, of a credit or insurance portfolio, which plays an increasingly important role in risk management in financial and insurance institutions.

We derive two types of saddlepoint expansions for the two quantities. The first type of approximation formulas for $\mathbb{E}[(X-K)^+]$ is based on Esscher tilting and the Edgeworth expansion. The second type of approximations is obtained by two distinct approaches. The resulting formulas distinguish themselves from all existing approximation formulas by their remarkable simplicity. We also establish error convergence rates for both types of approximations in the *i.i.d.* case. The approximations are further extended to cover the case of lattice variables. The lattice case is largely ignored, even in applications where lattice variables are more relevant, for example, the pricing of CDOs.

The two quantities are related as follows,

$$\mathbb{E}[X|X \geq K] = \frac{\mathbb{E}[(X-K)^+]}{\mathbb{P}(X \geq K)} + K, \tag{7.1}$$

$$\mathbb{E}[(X-K)^+] = \mathbb{E}\left[X\mathbf{1}_{\{X \geq K\}}\right] - K\mathbb{P}(X \geq K), \tag{7.2}$$

$$\mathbb{E}[X|X \geq K] = \frac{\mathbb{E}\left[X\mathbf{1}_{\{X \geq K\}}\right]}{\mathbb{P}(X \geq K)}. \tag{7.3}$$

It is also straightforward to extend our results to the functions $\mathbb{E}[(K-X)^+]$ and $\mathbb{E}[X|X < K]$. The connections are well known and we put them here only for completeness.

$$\begin{aligned}
&\mathbb{E}[(K-X)^+] = \mathbb{E}[(X-K)^+] - \mathbb{E}[X] + K,\\
&\mathbb{E}[X\mathbf{1}_{\{X<K\}}] = \mathbb{E}[X] - \mathbb{E}[X\mathbf{1}_{\{X \geq K\}}],\\
&\mathbb{E}[X|X < K] = \left(\mathbb{E}[X] - \mathbb{E}[X\mathbf{1}_{\{X \geq K\}}]\right)/\mathbb{P}(X < K).
\end{aligned}$$

For simplicity of notation, we define

$$\begin{cases} C &:= \mathbb{E}[(X-K)^+],\\ S &:= \mathbb{E}[X|X \geq K],\\ J &:= \mathbb{E}\left[X\mathbf{1}_{\{X \geq K\}}\right]. \end{cases} \tag{7.4}$$

7.2 Densities and tail probabilities

Dating back to Esscher (1932), the saddlepoint approximation has been recognized as a valuable tool in asymptotic analysis and statistical computing. It has found a wide range of applications in finance and insurance, reliability theory, physics and biology. The saddlepoint approximation literature so far mainly focuses on the approximation of densities (Daniels 1954) and tail probabilities (Lugannani & Rice 1980, Daniels 1987). For a comprehensive exposition of saddlepoint approximations, see Jensen (1995).

We start with some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X_i, i = 1 \ldots n$ be n independently and identically distributed continuous random variables all defined on the given probability space and $X = \sum_{i=1}^{n} X_i$. Suppose the moment generating function (MGF) of X_1 is analytic and given by $M_1(t)$ for t in some open neighborhood of zero, the MGF of the sum X is then simply the product of the MGF of X_i, i.e.,

$$M(t) = (M_1(t))^n.$$

Let $\kappa(t) = \log M(t)$ be the Cumulant Generating Function(CGF) of X. The density and tail probability of X can be represented by the following inversion formulas

$$f_X(K) = \frac{1}{2\pi \mathtt{i}} \int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} \exp(\mathcal{K}(t) - tK)\mathrm{d}t, \tag{7.5}$$

$$\mathbb{P}(X \geq K) = \frac{1}{2\pi \mathtt{i}} \int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} \frac{\exp(\mathcal{K}(t) - tK)}{t}\mathrm{d}t \quad (\tau > 0). \tag{7.6}$$

Throughout this chapter we adopt the following notation:

- $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the *p.d.f.* and *c.d.f.* of a standard normal random variable,
- $\mathcal{K}_1(t) = \log M_1(t)$ be the CGF of X_1.
- $\mu := \mathbb{E}[X]$ and $\mu_1 = \mathbb{E}[X_1]$ are the expectation of X and X_1 under $\mathbb{P}$,
- T represents the saddlepoint that gives $\mathcal{K}'(T) = K$,
- $\lambda_r := \mathcal{K}^{(r)}(T)/\mathcal{K}''(T)^{r/2}$ is the standardized cumulant of order r evaluated at T, and $\lambda_{1,r} := \mathcal{K}_1^{(r)}(T)/\mathcal{K}_1''(T)^{r/2}$,
- $Z := T\sqrt{\mathcal{K}''(T)}$ and $Z_1 := T\sqrt{\mathcal{K}_1''(T)}$,
- $W := sgn(T)\sqrt{2[KT - \mathcal{K}(T)]}$ and $W_1 := sgn(T)\sqrt{2[KT/n - \mathcal{K}_1(T)]}$ with $sgn(T)$ being the sign of T.

It is obvious that $\mu = n\mu_1$, $Z = \sqrt{n}Z_1$, $W = \sqrt{n}W_1$, $\lambda_3 = \lambda_{1,3}/\sqrt{n}$ and $\lambda_4 = \lambda_{1,4}/n$.

In the sequel we should write formulas in terms of X_1 (i.e., formulas with subscript 1 such as Z_1, W_1, etc) when we wish to study the order of the approximation errors. Otherwise, we write the formulas in terms of X (i.e., Z, W, etc) for notational simplicity. The latter is more general in the sense that it is also applicable when the random variables X_i are not identically distributed.

The saddlepoint approximation for densities is given by the Daniels (1954) formula

$$f_X(K) = \phi(\sqrt{n}W_1)\frac{T}{\sqrt{n}Z_1}\left[1 + \frac{1}{n}\left(\frac{\lambda_{1,4}}{8} - \frac{5\lambda_{1,3}^2}{24}\right) + O\left(n^{-2}\right)\right] \tag{7.7}$$

$$\approx \phi(W)\frac{T}{Z}\left(1 + \frac{\lambda_4}{8} - \frac{5\lambda_3^2}{24}\right) =: f_D. \tag{7.8}$$

For tail probabilities, two types of distinct saddlepoint expansions exist. The first type of expansion is given by

$$\mathbb{P}(X \geq K) = e^{\frac{n}{2}(Z_1^2 - W_1^2)}[1 - \Phi(\sqrt{n}Z_1)]\left[1 + O\left(n^{-\frac{1}{2}}\right)\right] \tag{7.9}$$

$$\approx e^{-\frac{W^2}{2}+\frac{Z^2}{2}}[1 - \Phi(Z)] =: P_1, \tag{7.10}$$

$$\mathbb{P}(X \geq K) = \left[P_1\left(1 - \frac{n\lambda_{1,3}}{6}Z_1^3\right) + \phi(\sqrt{n}W_1)\frac{\lambda_{1,3}}{6\sqrt{n}}\left(nZ_1^2 - 1\right)\right]\left[1 + O\left(n^{-1}\right)\right] \tag{7.11}$$

$$\approx P_1\left(1 - \frac{\lambda_3}{6}Z^3\right) + \phi(W)\frac{\lambda_3}{6}\left(Z^2 - 1\right) =: P_2, \tag{7.12}$$

in the case $T \geq 0$. For $T < 0$ similar formulas are available, see Daniels (1987). The second type of expansion is obtained by Lugannani & Rice (1980), with

$$\mathbb{P}(X \geq K) = [1 - \Phi(\sqrt{n}W_1)] + \phi(\sqrt{n}W_1)\left[\frac{1}{\sqrt{n}}\left(\frac{1}{Z_1} - \frac{1}{W_1}\right) + O\left(n^{-\frac{3}{2}}\right)\right] \tag{7.13}$$

$$\approx 1 - \Phi(W) + \phi(W)\left[\frac{1}{Z} - \frac{1}{W}\right] =: P_3, \tag{7.14}$$

$$\mathbb{P}(X \geq K) = P_3 + \phi(\sqrt{n}W_1)\left\{n^{-\frac{3}{2}}\left[\frac{1}{Z_1}\left(\frac{\lambda_{1,4}}{8} - \frac{5\lambda_{1,3}^2}{24}\right) - \frac{\lambda_{1,3}}{2Z_1^2} - \frac{1}{Z_1^3} + \frac{1}{W_1^3}\right] + O\left(n^{-\frac{5}{2}}\right)\right\} \tag{7.15}$$

$$\approx P_3 + \phi(W)\left[\frac{1}{Z}\left(\frac{\lambda_4}{8} - \frac{5\lambda_3^2}{24}\right) - \frac{\lambda_3}{2Z^2} - \frac{1}{Z^3} + \frac{1}{W^3}\right] =: P_4. \tag{7.16}$$

Widely known as the Lugannani-Rice formula, P_3 is most popular among the four tail probability approximations for both simplicity and accuracy. A good review of saddlepoint approximations for the tail probability is given in Daniels (1987).

7.3 Measure change approaches

Before we derive the formulas for $\mathbb{E}[(X-K)^+]$ and $\mathbb{E}[X|X \geq K]$, we would like to briefly review a different approach to approximating the two quantities. This usually involves a change of measure and borrows the saddlepoint expansions for densities or tail probabilities.

An inversion formula similar to those for densities and tail probabilities also exists for $\mathbb{E}[(X-K)^+]$, which is given by

$$\mathbb{E}\left[(X-K)^+\right] = \frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\frac{\exp(\mathcal{K}(t)-tK)}{t^2}\mathrm{d}t \quad (\tau>0). \tag{7.17}$$

Yang et al. (2006) rewrite the inversion formula to be

$$\mathbb{E}\left[(X-K)^+\right] = \frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\exp(\mathcal{K}(t)-\log t^2-tK)\mathrm{d}t. \tag{7.18}$$

Take $\mathcal{K}_{\mathbb{Q}}(t)=\mathcal{K}(t)-\log t^2$, where subscript $\mathbb{Q}$ denotes a probability measure different from the original measure $\mathbb{P}$, the right-hand side of (7.18) is then in the form of (7.5) and the Daniels formula (7.8) can be used for approximation. It should be pointed out, however, that in this case always two saddlepoints exist. Moreover, the MGF of X under the new measure $\mathbb{Q}$ is problematic as $M_{\mathbb{Q}}(0)\to\infty$,which suggests that $\mathbb{Q}$ is *not a probability measure*.

Bounded random variables

Studer (2001) considers the approximation of the expected shortfall, in two models of the associated random variable.

The first case deals with bounded random variables. Without loss of generality, we only consider the case that X has a nonnegative lower bound. Define the probability measure $\mathbb{Q}$ on $(\Omega,\mathcal{F})$ by $\mathbb{Q}(A)=\int_A X/\mu\mathrm{d}\mathbb{P}$ for $A\in\mathcal{F}$, then

$$\begin{aligned}\mathbb{E}[X|X\geq K] &= \frac{1}{\mathbb{P}(X\geq K)}\int_{\{X\geq K\}}X\mathrm{d}\mathbb{P} = \frac{\mu}{\mathbb{P}(X\geq K)}\int_{\{X\geq K\}}\frac{X}{\mu}\mathrm{d}\mathbb{P}\\ &= \frac{\mu}{\mathbb{P}(X\geq K)}\mathbb{Q}(X\geq K).\end{aligned} \tag{7.19}$$

Hence the expected shortfall is transformed to be a multiple of the ratio of two tail probabilities. The MGF of X under probability $\mathbb{Q}$ reads

$$M_{\mathbb{Q}}(t)=\int e^{tX}\frac{X}{\mu}\mathrm{d}\mathbb{P}=\frac{M'(t)}{\mu}=\frac{M(t)\mathcal{K}'(t)}{\mu}$$

as $\mathcal{K}'(t)=[\log M(t)]'=M'(t)/M(t)$. It follows that

$$\mathcal{K}_{\mathbb{Q}}(t)=\log M_{\mathbb{Q}}(t)=\mathcal{K}(t)+\log\left(\mathcal{K}'(t)\right)-\log(\mu). \tag{7.20}$$

For bounded variables in general it is only necessary to apply a linear transform on the random variable X beforehand so that the new variable has a nonnegative lower bound and thus $\mathbb{Q}(\cdot)$ is a valid probability measure.

The saddlepoint approximation for tail probability can be applied for both probabilities $\mathbb{P}$ and $\mathbb{Q}$ in (7.19). A disadvantage of this approach is that two saddlepoints need to be found as the saddlepoints under the two probability measures are generally different.

Log-return model

The second case in Studer (2001) deals with $\mathbb{E}[e^X|X \geq K]$ rather than $\mathbb{E}[X|X \geq K]$. The expected shortfall $\mathbb{E}[e^X|X \geq K]$ can also be written to be a multiple of the ratio of two tail probabilities. Define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by $\mathbb{Q}(A) = \int_A e^X/M(1)\mathrm{d}\mathbb{P}$ for $A \in \mathcal{F}$, then

$$\begin{aligned}\mathbb{E}[e^X|X \geq K] &= \frac{1}{\mathbb{P}(X \geq K)} \int_{\{X \geq K\}} e^X \mathrm{d}\mathbb{P} = \frac{M(1)}{\mathbb{P}(X \geq K)} \int_{\{X \geq K\}} \frac{e^X}{M(1)} \mathrm{d}\mathbb{P} \\ &= \frac{M(1)}{\mathbb{P}(X \geq K)} \mathbb{Q}(X \geq K). \end{aligned} \tag{7.21}$$

The MGF and CGF of X under probability $\mathbb{Q}$ are given by

$$\begin{aligned} M_{\mathbb{Q}}(t) &= \int e^{tX} \frac{e^X}{M(1)} \mathrm{d}\mathbb{P} = \frac{M(t+1)}{M(1)}, \\ \mathcal{K}_{\mathbb{Q}}(t) &= \mathcal{K}(t+1) - \mathcal{K}(1). \end{aligned}$$

This also forms the basis for the approach used in Rogers & Zane (1999) for option pricing where the log-price process follows a Lévy process.

7.4 Classical saddlepoint approximations

In this and in the sections to follow we give, in the spirit of Daniels (1987), two types of explicit saddlepoint approximations for $\mathbb{E}[(X-K)^+]$. For each type of approximation, we give a lower order version and a higher order version. The approximations to $\mathbb{E}[X|X \geq K]$ then simply follow from (7.1). No measure change is required and only one saddlepoint needs to be computed.

Following Jensen (1995), we call this first type of approximations the classical saddlepoint approximations. Approximation formulas for $\mathbb{E}[(X-K)^+]$ of this type already appeared in Antonov et al. (2005), however with no discussion on the error terms. They are obtained by means of routine

application of the saddlepoint approximation to (7.17), i.e., on the basis of the Taylor expansion of $\mathcal{K}(t) - tK$ around $t = T$. Here we provide a simpler and more statistically-oriented derivation that employs Esscher tilting and the Edgeworth expansion. Rates of convergence for the approximations are readily available with our approach in the i.i.d. case. Another advantage of our approach is that it leads to explicit saddlepoint approximations in the log-return model in Studer (2001), which is not possible with the approach in Antonov et al. (2005).

For now we assume that the saddlepoint $t = T$ that solves $\mathcal{K}'(t) = K$ is positive. The expectation $\mathbb{E}[(X-K)^+]$ is reformulated under an exponentially tilted probability measure,

$$\begin{aligned}\mathbb{E}\left[(X-K)^+\right] &= \int_K^\infty (x-K) f(x) \mathrm{d}x \\ &= e^{-\frac{nW_1^2}{2}} \int_K^\infty (x-K) e^{-T(x-K)} \tilde{f}(x) \mathrm{d}x, \end{aligned} \tag{7.22}$$

where $\mathcal{K}'(T) = K$ and $\tilde{f}(x) = f(x)\exp(Tx - \mathcal{K}(T))$. The MGF associated with $\tilde{f}(x)$ is given by $\tilde{M}(t) = M(T+t)/M(T)$. It immediately follows that the mean and variance of a random variable $\tilde{X}$ with density $\tilde{f}(\cdot)$ are given by $\mathbb{E}\tilde{X} = K$ and $Var(\tilde{X}) = \mathcal{K}''(T) = n\mathcal{K}_1''(T)$. Writing $\xi = (x-K)/\sqrt{n\mathcal{K}_1''(T)}$ and $\tilde{f}(x)dx = g(\xi)d\xi$, (7.22) reads

$$\mathbb{E}\left[(X-K)^+\right] = e^{-\frac{nW_1^2}{2}} \sqrt{n\mathcal{K}_1''(T)} \int_0^\infty \xi e^{-\sqrt{n}Z_1\xi} g(\xi) \mathrm{d}\xi. \tag{7.23}$$

Suppose that $g(\xi)$ is approximated by a normal distribution such that $g(\xi) = \phi(\xi)[1 + O(n^{-\frac{1}{2}})]$. The integral in (7.23) then becomes

$$\begin{aligned}\int_0^\infty \xi e^{-\sqrt{n}Z_1\xi} g(\xi) \mathrm{d}\xi &= \int_0^\infty \xi e^{-\sqrt{n}Z_1\xi} \phi(\xi) \left[1 + O\left(n^{-\frac{1}{2}}\right)\right] \mathrm{d}\xi \\ &= \frac{\exp(\frac{nZ_1^2}{2})}{\sqrt{2\pi}} \int_0^\infty \xi e^{-\frac{(\xi+\sqrt{n}Z_1)^2}{2}} \mathrm{d}\xi \left[1 + O\left(n^{-\frac{1}{2}}\right)\right] \\ &= \left\{\frac{1}{\sqrt{2\pi}} - \sqrt{n}Z_1 e^{\frac{nZ_1^2}{2}} \left[1 - \Phi(\sqrt{n}Z_1)\right]\right\} \left[1 + O\left(n^{-\frac{1}{2}}\right)\right]. \end{aligned} \tag{7.24}$$

Inserting (7.24) in (7.23) leads to the following approximation

$$\begin{aligned}&\mathbb{E}\left[(X-K)^+\right] \\ &= e^{-\frac{nW_1^2}{2}} \left\{\sqrt{\frac{n\mathcal{K}_1''(T)}{2\pi}} - Tn\mathcal{K}_1''(T) e^{\frac{nZ_1^2}{2}} \left[1 - \Phi(\sqrt{n}Z_1)\right]\right\} \left[1 + O\left(n^{-\frac{1}{2}}\right)\right]. \end{aligned} \tag{7.25}$$

By deleting the error term in (7.25) and representing the remaining terms in quantities related to X, we obtain,

$$\mathbb{E}\left[(X-K)^+\right] \approx e^{-\frac{W^2}{2}}\left\{\sqrt{\frac{\mathcal{K}''(T)}{2\pi}} - T\mathcal{K}''(T)e^{\frac{Z^2}{2}}\left[1-\Phi(Z)\right]\right\} =: C_1. \tag{7.26}$$

Higher order terms enter if $g(\xi)$ is approximated by its Edgeworth expansion, e.g.,

$$g(\xi) = \phi(\xi)[1 + \frac{\lambda_{1,3}}{6\sqrt{n}}(\xi^3 - 3\xi) + O(n^{-1})].$$

This gives

$$\begin{aligned}
&\mathbb{E}\left[(X-K)^+\right] \\
&= C_1\left[1+O(n^{-1})\right] + e^{-\frac{nW_1^2}{2}}\sqrt{\mathcal{K}_1''(T)}\frac{\lambda_{1,3}}{6}\int_0^\infty \xi e^{-Z\xi}\phi(\xi)(\xi^3-3\xi)\mathrm{d}\xi \\
&= C_1\left[1+O(n^{-1})\right] + e^{-\frac{nW_1^2}{2}}\sqrt{\mathcal{K}_1''(T)}\frac{\lambda_{1,3}}{6}\frac{e^{\frac{Z^2}{2}}}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{(\xi+Z)^2}{2}}\left(-\xi^4+3\xi^2\right)\mathrm{d}\xi \\
&= C_1\left[1+O(n^{-1})\right] + e^{\frac{n}{2}(Z_1^2-W_1^2)}\sqrt{\mathcal{K}_1''(T)}\frac{\lambda_{1,3}}{6} \\
&\quad\times\left\{[1-\Phi(\sqrt{n}Z_1)](n^2Z^4+3nZ^2) - \phi(\sqrt{n}Z_1)(n^{\frac{3}{2}}Z_1^3+2\sqrt{n}Z_1)\right\} \qquad (7.27) \\
&\approx C_1 + e^{\frac{Z^2}{2}-\frac{W^2}{2}}\sqrt{\mathcal{K}''(T)}\frac{\lambda_3}{6}\left\{[1-\Phi(Z)](Z^4+3Z^2) - \phi(Z)(Z^3+2Z)\right\} \\
&=: C_2. \qquad (7.28)
\end{aligned}$$

The approximations C_1 and C_2 are in agreement with the formulas given by Antonov et al. (2005). But our derivation has the advantage helping us better understand the order of the approximations.

Negative saddlepoint

We have assumed that the saddlepoint is positive, when deriving C_1 and C_2 in (7.26) and (7.28), or, in other words, $\mu < K$. If the saddlepoint T equals 0, or equivalently, $\mu = K$, it is straightforward to see that C_1 and C_2 both reduce to the following formula,

$$\mathbb{E}[(X-\mu)^+] = \sqrt{\frac{\mathcal{K}''(0)}{2\pi}} =: C_0. \tag{7.29}$$

In case that $\mu > K$, we should work with $Y = -X$ and $\mathbb{E}[Y\mathbf{1}_{\{Y\geq -K\}}]$ instead since

$$\mathbb{E}[X\mathbf{1}_{\{X\geq K\}}] = \mu + \mathbb{E}[-X\mathbf{1}_{\{-X\geq -K\}}] = \mu + \mathbb{E}[Y\mathbf{1}_{\{Y\geq -K\}}].$$

The CGF of Y is given by $\mathcal{K}_Y(t) = \mathcal{K}_X(-t)$. The saddlepoint that solves $\mathcal{K}_Y(t) = -K$ is $-T > 0$ so that C_1 and C_2 can again be used. Note that

$$\mathcal{K}_Y^{(r)}(t) = (-1)^r \mathcal{K}_X^{(r)}(-t),$$

where the superscript (r) denotes the r-th derivative. Therefore, in the case of a negative saddlepoint, $\mathbb{E}[(X-K)^+]$ can be approximated by

$$C_1^- = \mu - K + e^{-\frac{W^2}{2}} \left\{ \sqrt{\mathcal{K}''(T)/(2\pi)} + T\mathcal{K}''(T) e^{\frac{Z^2}{2}} \Phi(Z) \right\}, \tag{7.30}$$

$$C_2^- = C_1^- - e^{\frac{Z^2}{2} - \frac{W^2}{2}} \sqrt{\mathcal{K}''(T)} \frac{\lambda_3}{6} \left\{ \Phi(Z)(Z^4 + 3Z^2) + \phi(Z)(Z^3 + 2Z) \right\}. \tag{7.31}$$

Log-return model revisited

We now show how to deal with the log-return model in Studer (2001) without working with two probability measures simultaneously. We work with $\mathbb{E}\left[e^X \mathbf{1}_{\{X\geq K\}}\right]$ which equals $\mathbb{E}\left[e^X | X \geq K\right] \mathbb{P}(X \geq K)$. Replace x in (7.22) by e^x and make the same change of variables,

$$\mathbb{E}\left[e^X \mathbf{1}_{\{X\geq K\}}\right] = e^{-\frac{W^2}{2}} \int_0^\infty e^{K + \xi\sqrt{n\mathcal{K}''(T)}} e^{-Z\xi} g(\xi) \mathrm{d}\xi.$$

Approximating $g(\xi)$ by the standard normal density, we obtain

$$\begin{aligned} \mathbb{E}\left[e^X \mathbf{1}_{\{X\geq K\}}\right] &\approx e^{-\frac{W^2}{2} + K + \frac{\dot{Z}^2}{2}} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{(\xi + \dot{Z})^2}{2}} \mathrm{d}\xi \\ &= e^{-\frac{W^2}{2} + K + \frac{\dot{Z}^2}{2}} [1 - \Phi(\dot{Z})], \end{aligned} \tag{7.32}$$

where $\dot{Z} = (T-1)\sqrt{\mathcal{K}''(T)}$. Eq (7.32) is basically $e^K P_1$, where P_1 is given by (7.10), with Z replaced by $\dot{Z}$. It is easy to verify that this approximation is exact when X is normally distributed. A higher order approximation would be

$$\begin{aligned} &\mathbb{E}\left[e^X \mathbf{1}_{\{X\geq K\}}\right] \\ \approx &e^{-\frac{W^2}{2} + K + \frac{\dot{Z}^2}{2}} \left\{ [1 - \Phi(\dot{Z})] \left(1 - \frac{\lambda_3}{6\sqrt{n}} \dot{Z}^3\right) + \frac{\lambda_3}{6\sqrt{n}} \phi(\dot{Z})(\dot{Z}^2 - 1) \right\}. \end{aligned}$$

7.5 The Lugannani-Rice type formulas

The second type of saddlepoint approximations to $\mathbb{E}[(X-K)^+]$ can be derived in a very similar way as was done in section 4 of Daniels (1987) where the Lugannani-Rice formula to tail probability was derived. As a result we shall call the obtained formulas "Lugannani-Rice type formulas".

To start, we derive the following inversion formula for $\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right]$.

Theorem 7.1. *Let $\mathcal{K}(t)=\log M(t)$ be the cumulant generating function of a continuous random variable X. Then*

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right]=\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\mathcal{K}'(t)\frac{\exp(\mathcal{K}(t)-tK)}{t}\mathrm{d}t\quad(\tau>0).\tag{7.33}$$

Proof. We start with the case that X has a positive lower bound. Employing the same change of measure as in (7.19), we have $\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right]=\mu\mathbb{Q}(X\geq K)$, where

$$\mathbb{Q}(X\geq K)=\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\frac{\exp(\mathcal{K}_{\mathbb{Q}}(t)-tK)}{t}\mathrm{d}t\quad(\tau>0).$$

Plug in $\mathcal{K}_{\mathbb{Q}}(t)$, which is given by (7.20), we find

$$\begin{aligned}\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right]&=\mu\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\frac{\exp\left[\mathcal{K}(t)+\log\mathcal{K}'(t)-\log\mu-tK\right]}{t}\mathrm{d}t\\&=\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\mathcal{K}'(t)\frac{\exp(\mathcal{K}(t)-tK)}{t}\mathrm{d}t.\end{aligned}$$

In the case that X has a negative lower bound, $-a$, with $a>0$, we define $Y=X+a$ so that Y has a positive lower bound. Then, the CGF of Y and its first derivative are given by $\mathcal{K}_Y(t)=\mathcal{K}(t)+ta$ and $\mathcal{K}'_Y(t)=\mathcal{K}'(t)+a$, respectively. Since

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right]=\mathbb{E}\left[(Y-a)\mathbf{1}_{\{Y-a\geq K\}}\right]=\mathbb{E}\left[Y\mathbf{1}_{\{Y-a\geq K\}}\right]-a\mathbb{P}(Y-a\geq K),$$

and

$$\mathbb{E}\left[Y\mathbf{1}_{\{Y-a\geq K\}}\right]=\frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty}\mathcal{K}'(t)\frac{\exp(\mathcal{K}(t)-tK)}{t}\mathrm{d}t+a\mathbb{P}(Y-a\geq K),$$

we are again led to (7.33). Extension to variables bounded from above is straightforward.

For unbounded X, we take $X_L = \max(X, L)$, where $L < -1/\tau$ is a constant. Since X_L is bounded from below, we have

$$\begin{aligned}\mathbb{E}\left[X_L \mathbf{1}_{\{X_L \geq K\}}\right] &= \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} \mathcal{K}'_{X_L}(t) \frac{\exp(\mathcal{K}_{X_L}(t) - tK)}{t} \mathrm{d}t, \\ &= \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} M'_{X_L}(t) \frac{\exp(-tK)}{t} \mathrm{d}t, \end{aligned} \tag{7.34}$$

where $M'_{X_L}(\tau) = M'(\tau) + \int_{-\infty}^{L} (Le^{\tau L} - xe^{\tau x}) d\mathbb{P}(x)$. For $L < -1/\tau$, $M'_{X_L}(\tau)$ increases monotonically as L decreases and approaches $M'(\tau)$ as $L \to -\infty$. Note also that $\mathbb{E}\left[X\mathbf{1}_{\{X \geq K\}}\right] = \mathbb{E}\left[X_L \mathbf{1}_{\{X_L \geq K\}}\right]$ for all $L < K$. Now take the limit of both sides of (7.34) as $L \to -\infty$. Due to the monotone convergence theorem, we again obtain

$$\begin{aligned}\mathbb{E}\left[X\mathbf{1}_{\{X \geq K\}}\right] &= \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} M'(t) \frac{\exp(-tK)}{t} \mathrm{d}t \\ &= \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\infty}^{\tau + \mathtt{i}\infty} \mathcal{K}'(t) \frac{\exp(\mathcal{K}(t) - tK)}{t} \mathrm{d}t. \end{aligned}$$

□

We look at $K = nx$ for fixed x and let $\mathcal{K}'_1(T) = x$ so that $\mathcal{K}'(T) = n\mathcal{K}'_1(T) = nx = K$.

We follow Daniels (1987) to approximate $\mathcal{K}_1(t) - tx$ over an interval containing both $t = 0$ and $t = T$ by a quadratic function. Here, T need not to be positive any more. Since $nx = K$ we have $-\frac{1}{2}W_1^2 = \mathcal{K}_1(T) - Tx$ with W_1 taking the same sign as T. Let w be defined between 0 and W_1 such that

$$\frac{1}{2}(w - W_1)^2 = \mathcal{K}_1(t) - tx - \mathcal{K}_1(T) + Tx. \tag{7.35}$$

Then we have

$$\frac{1}{2}w^2 - W_1 w = \mathcal{K}_1(t) - t\mathcal{K}'_1(T), \tag{7.36}$$

and $t = 0 \Leftrightarrow w = 0$, $t = T \Leftrightarrow w = W_1$. Differentiate both sides of (7.36) once and twice to obtain

$$w\frac{\mathrm{d}w}{\mathrm{d}t} - W_1 \frac{\mathrm{d}w}{\mathrm{d}t} = \mathcal{K}'_1(t) - \mathcal{K}'_1(T), \quad \left(\frac{\mathrm{d}w}{\mathrm{d}t}\right)^2 + (w - W_1)\frac{d^2 w}{\mathrm{d}t^2} = \mathcal{K}''_1(t).$$

So, in the neighborhood of $t = T$ (or, equivalently, $w = W_1$) we have $\frac{\mathrm{d}w}{\mathrm{d}t} = \sqrt{\mathcal{K}''_1(T)}$. Note that $\mu_1 = \mathbb{E}[X_1] = \mathcal{K}'_1(0)$. In the neighborhood of

$t = 0$ (or, equivalently, $w = 0$), we have

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \sqrt{\mathcal{K}_1''(0)} \quad \text{if } T = 0, \tag{7.37}$$

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{\mathcal{K}_1'(T) - \mathcal{K}_1'(0)}{W_1} = \frac{x - \mu_1}{W_1} \quad \text{if } T \neq 0.$$

Hence, in the neighborhood of $t = 0$ we have $w \propto t$. Moreover,

$$\frac{1}{t}\frac{\mathrm{d}t}{\mathrm{d}w} \sim \frac{1}{w}, \quad \frac{\mathcal{K}_1'(t)}{t}\frac{\mathrm{d}t}{\mathrm{d}w} \sim \frac{\mu_1}{w}. \tag{7.38}$$

Based on Theorem 7.1, the inversion formula for $\mathbb{E}\left[X\mathbf{1}_{\{X \geq nx\}}\right]$ can be formulated to be

$$\begin{aligned}
\mathbb{E}\left[X\mathbf{1}_{\{X \geq nx\}}\right] &= \frac{1}{2\pi\mathtt{i}} \int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} n\mathcal{K}_1'(t) e^{n(\frac{1}{2}w^2 - W_1 w)} \frac{1}{t}\frac{\mathrm{d}t}{\mathrm{d}w}\mathrm{d}w \\
&= \frac{n}{2\pi\mathtt{i}} \int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} e^{n(\frac{1}{2}w^2 - W_1 w)} \left[\frac{\mu_1}{w} + \frac{\mathcal{K}_1'(t)}{t}\frac{\mathrm{d}t}{\mathrm{d}w} - \frac{\mu_1}{w}\right]\mathrm{d}w \\
&= n\mu_1 \int_{\tau-\mathtt{i}\infty}^{\tau+\mathtt{i}\infty} \frac{1}{2\pi\mathtt{i}} e^{n(\frac{1}{2}w^2 - W_1 w)} \frac{\mathrm{d}w}{w} \\
&\quad + \frac{n e^{-\frac{nW_1^2}{2}}}{2\pi\mathtt{i}} \int_{W_1-\mathtt{i}\infty}^{W_1+\mathtt{i}\infty} e^{\frac{1}{2}n(w-W_1)^2} \left[\frac{\mathcal{K}_1'(t)}{t}\frac{\mathrm{d}t}{\mathrm{d}w} - \frac{\mu_1}{w}\right]\mathrm{d}w.
\end{aligned} \tag{7.39}$$

The first integral takes the value $1 - \Phi(\sqrt{n}W_1) = 1 - \Phi(W)$. The second integral has no singularity because of (7.38). Hence there is no problem to change the integration contour from the imaginary axis along $\tau > 0$ to that along W_1, as done in (7.39), not even if W_1 and T are negative. The major contribution to the second integral comes from the saddlepoint. The terms in the brackets are expanded around T and integrated to give an expansion of the form

$$n\phi(\sqrt{n}W_1)(b_1 n^{-\frac{1}{2}} + b_3 n^{-\frac{3}{2}} + b_5 n^{-\frac{5}{2}} + \ldots). \tag{7.40}$$

By Watson's lemma this is an asymptotic expansion in a neighborhood of W_1. For more details see Lemma 4.5.2 in Kolassa (2006). The coefficient b_1 in (7.40) can be obtained by only taking into account the leading terms of the Taylor expansion of

$$\frac{\mathcal{K}_1'(t)}{t}\frac{\mathrm{d}t}{\mathrm{d}w} - \frac{\mu_1}{w} = \frac{\mathcal{K}_1'(t)}{t}\frac{\mathrm{d}t}{\mathrm{d}w}\Big|_T - \frac{\mu_1}{w}\Big|_{W_1} + \ldots = \frac{x}{Z_1} - \frac{\mu_1}{W_1} + \ldots. \tag{7.41}$$

Therefore we are led to

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq nx\}}\right] = n\mu_1\left[1-\Phi(\sqrt{n}W_1)\right] + n\phi(\sqrt{n}W_1)\left[\frac{1}{\sqrt{n}}\left(\frac{x}{Z_1}-\frac{\mu_1}{W_1}\right)+O\left(n^{-\frac{3}{2}}\right)\right] \tag{7.42}$$

Subtracting $K\mathbb{P}(X\geq K)$ from (7.42) with the tail probability approximated by the Lugannani-Rice formula P_3, we see immediately that

$$\mathbb{E}\left[(X-nx)^+\right] = n(\mu_1-x)\left[1-\Phi(\sqrt{n}W_1)-\frac{\phi(\sqrt{n}W_1)}{\sqrt{n}W_1}+O\left(n^{-\frac{3}{2}}\right)\right]. \tag{7.43}$$

Rewrite (7.42) and (7.43) in quantities related to X and deleting the error terms we obtain the following approximation,

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx \mu\left[1-\Phi(W)\right]+\phi(W)\left[\frac{K}{Z}-\frac{\mu}{W}\right] =: J_3. \tag{7.44}$$

$$\mathbb{E}\left[(X-K)^+\right] \approx (\mu-K)\left[1-\Phi(W)-\frac{\phi(W)}{W}\right] =: C_3. \tag{7.45}$$

C_3 is a surprisingly neat formula requiring only knowledge of W. A more statistical approach to derive the approximation J_3 in (7.44) can be found in Appendix A.

Next we consider the coefficient b_3 in (7.40). Write $U := \mathcal{K}_1''(T)T - \mathcal{K}_1'(T)$. The Taylor expansion of $\mathcal{K}_1'(t)/t$ around T gives

$$\frac{\mathcal{K}_1'(t)}{t} = \frac{\mathcal{K}_1'(T)}{T}+(t-T)\frac{U}{T^2}+\frac{(t-T)^2}{2}\left[\frac{\mathcal{K}_1'''(T)}{T}-\frac{2U}{T^3}\right]+\dots. \tag{7.46}$$

Furthermore, expand $\exp(n[\mathcal{K}_1(t)-tx])$ in the same way as Daniels (1954),

$$\begin{aligned}
&\exp(n[\mathcal{K}_1(t)-tx])\\
&=\exp\left(n[\mathcal{K}_1(T)-Tx]+\frac{1}{2}n\mathcal{K}_1''(T)(t-T)^2\right.\\
&\qquad\left.+\frac{n}{6}\mathcal{K}'''(T)(t-T)^3+\frac{n}{24}\mathcal{K}^{(4)}(t-T)^4+\dots\right)\\
&=\exp\left(n[\mathcal{K}_1(T)-Tx]+\frac{1}{2}n\mathcal{K}_1''(T)(t-T)^2\right)\\
&\times\left[1+\frac{n}{6}\mathcal{K}_1'''(T)(t-T)^3+\frac{n}{24}\mathcal{K}_1^{(4)}(T)(t-T)^4+\frac{n^2}{72}\mathcal{K}_1'''(T)^2(t-T)^6+\dots\right].
\end{aligned} \tag{7.47}$$

Put (7.46) and (7.47) together, we have, on the line $t = T + \mathtt{i}y$,

$$\begin{aligned}
&\frac{n}{2\pi\mathtt{i}}\int_{T-\mathtt{i}\infty}^{T+\mathtt{i}\infty} e^{n[\mathcal{K}_1(t)-tx]}\frac{\mathcal{K}_1'(t)}{t}\mathrm{d}t \\
=&\frac{ne^{-\frac{nW_1^2}{2}}}{2\pi\mathtt{i}}\int_{T-\mathtt{i}\infty}^{T+\mathtt{i}\infty} e^{\frac{1}{2}n\mathcal{K}_1''(T)(t-T)^2} \\
&\times\left[1+\frac{n}{6}\mathcal{K}_1'''(T)(t-T)^3+\frac{n}{24}\mathcal{K}_1^{(4)}(T)(t-T)^4+\frac{n^2}{72}\mathcal{K}_1'''(T)^2(t-T)^6+\ldots\right] \\
&\times\left\{\frac{\mathcal{K}_1'(T)}{T}+(t-T)\frac{U}{T^2}+\frac{(t-T)^2}{2}\left[\frac{\mathcal{K}_1'''(T)}{T}-\frac{2U}{T^3}\right]+\ldots\right\}\mathrm{d}t \\
=&\frac{ne^{-\frac{nW_1^2}{2}}}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}n\mathcal{K}_1''(T)y^2}\left[1-\frac{n}{6}\mathcal{K}_1'''(T)\mathtt{i}y^3+\frac{n}{24}\mathcal{K}_1^{(4)}(T)y^4+\right. \\
&\left.-\frac{n^2}{72}\mathcal{K}_1'''(T)^2y^6+\ldots\right]\left\{\frac{\mathcal{K}_1'(T)}{T}+\mathtt{i}y\frac{U}{T^2}-\frac{y^2}{2}\left[\frac{\mathcal{K}_1'''(T)}{T}-\frac{2U}{T^3}\right]+\ldots\right\}\mathrm{d}y \\
=&n\phi(\sqrt{n}W_1)\left\{\frac{\mathcal{K}_1'(T)}{\sqrt{n}Z_1}+n^{-\frac{3}{2}}\left[\frac{\mathcal{K}_1'(T)}{Z_1}\left(\frac{\lambda_{1,4}}{8}-\frac{5}{24}\lambda_{1,3}^2\right)\right.\right. \\
&\left.\left.+\frac{U\lambda_{1,3}}{2Z_1^2}-\frac{\lambda_{1,3}}{2T}+\frac{U}{Z_1^3}\right]+O\left(n^{-\frac{5}{2}}\right)\right\} \\
=&n\phi(W)\left[\frac{x}{\sqrt{n}Z_1}+n^{-\frac{3}{2}}\left(\frac{x\lambda_{1,4}}{8Z_1}-\frac{5x\lambda_{1,3}^2}{24Z_1}+\frac{1}{TZ_1}-\frac{x\lambda_{1,3}}{2Z_1^2}-\frac{x}{Z_1^3}\right)+O\left(n^{-\frac{5}{2}}\right)\right].
\end{aligned} \tag{7.48}$$

Notice that (7.48) is itself a saddlepoint approximation to $\mathbb{E}[X\mathbf{1}_{\{X\geq K\}}]$ for $K > \mu$. However, it becomes inaccurate when T approaches zero due to the presence of a pole at zero in the integrand. Meanwhile expanding $1/w$ in the second integral in (7.39) around W_1 gives

$$\begin{aligned}
&\frac{ne^{-\frac{nW_1^2}{2}}}{2\pi\mathtt{i}}\int_{W_1-\mathtt{i}\infty}^{W_1+\mathtt{i}\infty} e^{\frac{1}{2}n(w-W_1)^2}\frac{\mu_1}{w}\mathrm{d}w \\
=&\frac{n\mu_1e^{-\frac{nW_1^2}{2}}}{2\pi\mathtt{i}}\int_{W_1-\mathtt{i}\infty}^{W_1+\mathtt{i}\infty} e^{\frac{1}{2}n(w-W_1)^2}\left[\frac{1}{W_1}-\frac{(w-W_1)}{W_1^2}+\frac{(w-W_1)^2}{W_1^3}+\ldots\right]\mathrm{d}w \\
=&n\mu_1\phi(W)\left[\frac{1}{\sqrt{n}W_1}-\frac{1}{(\sqrt{n}W_1)^3}+O\left(n^{-\frac{5}{2}}\right)\right].
\end{aligned} \tag{7.49}$$

Adding (7.48) and (7.49) to $1-\Phi(\sqrt{n}W_1)$ and then subtracting nx times

(7.15) we obtain

$$\mathbb{E}[(X-nx)^+] = n(\mu_1 - x)\left\{[1-\Phi(\sqrt{n}W_1)] + \frac{\phi(\sqrt{n}W_1)}{\sqrt{n}W_1}\right\} + n\phi(\sqrt{n}W_1)\left\{n^{-\frac{3}{2}}\left[\frac{1}{TZ_1} + \frac{\mu_1 - x}{W_1^3}\right] + O\left(n^{-\frac{5}{2}}\right)\right\}, \quad (7.50)$$

which can be rewritten to be

$$\mathbb{E}[(X-K)^+] \approx C_3 + \phi(W)\left[\frac{1}{TZ} + (\mu - K)\frac{1}{W^3}\right] =: C_4. \quad (7.51)$$

Remark 7.2. Interestingly, Martin (2006) gives an approximation formula for
$\mathbb{E}[(X-K)^+]$, decomposing the expectation to one term involving the tail probability and another term involving the probability density,

$$\mathbb{E}\left[(X-K)^+\right] \approx (\mu - K)\mathbb{P}(X \geq K) + \frac{K-\mu}{T} f_X(K).$$

Martin (2006) suggests to approximate $\mathbb{P}(X \geq K)$ by the Lugannani-Rice formula P_3 in (7.13) and $f_X(K)$ by the Daniels formula f_D in (7.8). In the i.i.d. case, this leads to an approximation $C_M := n(\mu_1 - x)P_3 + n(x - \mu_1)f_D/T$ with a rate of convergence $n^{-1/2}$ as the first term has an error of order $n^{-1/2}$ and the second term has an error of order $n^{-3/2}$. We propose to replace P_3 by its higher order version, P_4 in (7.15). This gives the following formula,

$$\mathbb{E}\left[(X-K)^+\right] \approx C_3 + (\mu - K)\phi(W)\left(\frac{1}{W^3} - \frac{\lambda_3}{2Z^2} - \frac{1}{Z^3}\right). \quad (7.52)$$

Not only eq. (7.52) is simpler than C_M as λ_4 is not involved, but also it has a higher rate of convergence of order $n^{-3/2}$. However compared to C_4 eq. (7.52) contains a term of λ_3 and is certainly more complicated to evaluate. Note further that if we neglect in C_M the terms of the higher order standard cumulants λ_3 and λ_4 in f_D we get precisely C_3 as given in (7.45). For these reasons, C_4 is to be preferred.

Zero saddlepoint

It is mentioned in Daniels (1987) that in case that the saddlepoint $T = 0$, or in other words, $\mu = K$, the approximations to tail probability P_1 to P_4 all reduce to

$$\mathbb{P}(X \geq K) = \frac{1}{2} - \frac{\lambda_3(0)}{6\sqrt{2\pi}}.$$

We would like to show that, under the same circumstances, C_3 and C_4 also reduce to the formula C_0 in (7.29). To show that $C_3 = C_0$ when $T = 0$, we point out that

$$\lim_{T\to 0} C_3 = \lim_{T\to 0} \frac{\mathcal{K}'(0) - \mathcal{K}'(T)}{T} \left[T(1 - \Phi(W)) - \phi(W)\frac{T}{W} \right].$$

Note that when $T \to 0$, $\frac{\mathcal{K}'(0)-\mathcal{K}'(T)}{T} \to -\mathcal{K}''(0)$, $T(1 - \Phi(W)) \to 0$ and $\frac{T}{W} \to [\mathcal{K}''(0)]^{-\frac{1}{2}}$ (see (7.37)). This implies that $\lim_{T\to 0} C_3 = C_0$. Similarly we also have $\lim_{T\to 0} C_4 = C_0$.

7.6 Lattice variables

So far we have only considered approximations to continuous variables. Let us now turn to the lattice case. This is largely ignored in the literature, even in applications where lattice variables are much more relevant. For example, in the pricing of CDOs, the random variable concerned is essentially the number of defaults in the pool of companies and is thus discrete.

Suppose that $\hat{X}$ only takes integer values k with nonzero probabilities $p(k)$. The inversion formula of $\mathbb{E}[(\hat{X} - K)^+]$ can then be formulated as

$$\begin{aligned}
\mathbb{E}[(\hat{X} - K)^+] &= \sum_{k=K+1}^{\infty} (k - K)p(k) \\
&= \sum_{k=K+1}^{\infty} (k - K)\frac{1}{2\pi\mathtt{i}} \int_{\tau-\mathtt{i}\pi}^{\tau+\mathtt{i}\pi} \exp(\mathcal{K}(t) - tk)\mathrm{d}t \\
&= \frac{1}{2\pi\mathtt{i}} \int_{\tau-\mathtt{i}\pi}^{\tau+\mathtt{i}\pi} \exp(\mathcal{K}(t) - tK) \sum_{m=1}^{\infty} me^{-tm}\mathrm{d}t \\
&= \frac{1}{2\pi\mathtt{i}} \int_{\tau-\mathtt{i}\pi}^{\tau+\mathtt{i}\pi} \frac{\exp(\mathcal{K}(t) - tK)}{t^2} \frac{t^2e^{-t}}{(1 - e^{-t})^2}\mathrm{d}t \quad (\tau > 0).
\end{aligned}$$

For $K > \mu$, we proceed by expanding the two terms in the integrand separately. According to a truncated version of the Watson's Lemma (see Lemma 4.5.1 and 4.5.2 in Kolassa 2006), for an integrand in the form of $\exp(\frac{n\alpha}{2}(t-T)^2)\sum_{j=0}^{\infty}(t-T)^j$, the change in the contour of integration for t from $\tau \pm \mathtt{i}\infty$ to $\tau \pm \mathtt{i}\pi$ leads to a negligible difference which is exponentially small in n. Blackwell & Hodges (1959) declare further that the integral over the range $\tau + \mathtt{i}y$ where $|y| > \log n/\sqrt{n}$ is negligible. This means we are able to incorporate the formulas for continuous variables C_1 and C_2 in

the approximations for lattice variables. We find, for lattice variables, the following approximations corresponding to C_1 and C_2 in (7.26) and (7.28), respectively,

$$\hat{C}_1 = C_1 \frac{T^2 e^{-T}}{(1-e^{-T})^2}, \tag{7.53}$$

$$\begin{aligned}\hat{C}_2 = {} & C_2 \frac{T^2 e^{-T}}{(1-e^{-T})^2} \\ & + e^{-\frac{W^2}{2}+\frac{Z^2}{2}} \{\phi(Z) - Z[(1-\Phi(Z)]\} \frac{T e^{-T}\left(2 - T - 2e^{-T} - T e^{-T}\right)}{\sqrt{\mathcal{K}''(T)}(1-e^{-T})^3}.\end{aligned} \tag{7.54}$$

For the approximations to $\mathbb{E}[\hat{X}|\hat{X} \geq K]$, we also need the lattice version for the tail probability

$$\mathbb{P}(\hat{X} \geq K) \approx e^{-\frac{W^2}{2}+\frac{Z^2}{2}}[1-\Phi(Z)]\frac{T}{1-e^{-T}} =: \hat{P}_1 \tag{7.55}$$

or its higher order version

$$\begin{aligned}\mathbb{P}(\hat{X} \geq K) \approx {} & e^{-\frac{W^2}{2}+\frac{Z^2}{2}} \frac{T}{1-e^{-T}} \times \left\{ [1-\Phi(Z)] \left(2 - \frac{\lambda_3}{6} Z^3 - \frac{T}{e^T - 1}\right) \right. \\ & \left. + \phi(Z) \left[\frac{\lambda_3}{6}(Z^2-1) + \frac{1}{Z} - \frac{T}{Z(e^T-1)} \right] \right\} =: \hat{P}_2.\end{aligned} \tag{7.56}$$

Recall that the Lugannani-Rice formula for lattice variables reads

$$\mathbb{P}(\hat{X} \geq K) \approx 1 - \Phi(W) + \phi(W)\left[\frac{1}{\hat{Z}} - \frac{1}{W}\right] =: \hat{P}_3, \tag{7.57}$$

where $\hat{Z} = (1-e^{-T})\sqrt{\mathcal{K}''(T)}$. A similar lattice formula can also be obtained for J_3, which we will denote by $\hat{J}_3$. We first write down the inversion formula of the tail probability of a lattice variable,

$$\mathbb{Q}(\hat{X} \geq K) = \sum_{k=K}^{\infty} \mathbb{Q}(\hat{X} = k) = \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\pi}^{\tau + \mathtt{i}\pi} \frac{\exp(\mathcal{K}_{\mathbb{Q}}(t) - tK)}{1 - e^{-t}} \mathrm{d}t. \tag{7.58}$$

Combining (7.58) with Theorem 7.1, we obtain

$$\mathbb{E}\left[\hat{X}\mathbf{1}_{\{\hat{X} \geq K\}}\right] = \frac{1}{2\pi \mathtt{i}} \int_{\tau - \mathtt{i}\pi}^{\tau + \mathtt{i}\pi} \mathcal{K}'(t) \frac{\exp(\mathcal{K}(t) - tK)}{1 - e^{-t}} \mathrm{d}t.$$

By the same change of variables as in section 7.5, we have

$$\begin{aligned}\mathbb{E}\left[\hat{X}\mathbf{1}_{\{\hat{X}\geq K\}}\right] &= \frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\pi}^{\tau+\mathtt{i}\pi}\mathcal{K}'(t)e^{\frac{1}{2}w^2-Ww}\frac{1}{1-e^{-t}}\frac{\mathrm{d}t}{\mathrm{d}w}\mathrm{d}w \\ &= \frac{1}{2\pi\mathtt{i}}\int_{\tau-\mathtt{i}\pi}^{\tau+\mathtt{i}\pi}e^{\frac{1}{2}w^2-Ww}\left[\frac{\mu}{w}+\frac{\mathcal{K}'(t)}{1-e^{-t}}\frac{\mathrm{d}t}{\mathrm{d}w}-\frac{\mu}{w}\right]\mathrm{d}w.\end{aligned}$$

Now we can proceed exactly as in section 7.5 as $\lim_{t\to 0} 1-e^{-t} = t$. This leads to

$$\hat{J}_3 = \mu\left[1-\Phi(W)\right]+\phi(W)\left[\frac{K}{\hat{Z}}-\frac{\mu}{W}\right], \tag{7.59}$$

$$\hat{C}_3 = (\mu-K)\left[1-\Phi(W)-\frac{\phi(W)}{W}\right] \equiv C_3. \tag{7.60}$$

Including higher order terms we obtain

$$\hat{C}_4 = \hat{C}_3+\phi(W)\left[\frac{e^{-T}}{\hat{Z}(1-e^{-T})}+(\mu-K)\frac{1}{W^3}\right]. \tag{7.61}$$

A higher order version of $\hat{P}_3$ can be derived similarly,

$$\begin{aligned}\mathbb{P}(\hat{X}\geq K) \approx 1-\Phi(W)+\phi(W)\Bigg[&\frac{1}{\hat{Z}}\left(1+\frac{\lambda_4}{8}-\frac{5\lambda_3^2}{24}\right) \\ &-\frac{e^{-T}\lambda_3}{2\hat{Z}^2}-\frac{e^{-T}(1+e^{-T})}{2\hat{Z}^3}-\frac{1}{W}+\frac{1}{W^3}\Bigg] =: \hat{P}_4.\end{aligned} \tag{7.62}$$

This can be used to estimate $\mathbb{E}[\hat{X}|\hat{X}\geq K]$.

The rates of convergence of $\hat{C}_1$ to $\hat{C}_4$ in the i.i.d. case are identical to their non-lattice counterparts and we shall not elaborate further.

7.7 Numerical results

By two numerical experiments we evaluate the quality of the various approximations that are derived in the earlier sections.

In our first example $X = \sum X_i$ where X_i are i.i.d. *exponentially* distributed with density $p(x) = e^{-x}$. The CGF of X reads $\mathcal{K}(t) = -n\log(1-t)$. The saddlepoint to $\mathcal{K}'(t) = K$ is given by $T = 1-n/K$. Moreover, we have

$$\mathcal{K}''(T) = \frac{K^2}{n}, \lambda_3 = \frac{2}{\sqrt{n}}, \lambda_4 = \frac{6}{n}.$$

Their exact values are available as $X \sim Gamma(n,1)$. The tail probability is then given by

$$\mathbb{P}(X \geq K) = 1 - \frac{\gamma(n,K)}{\Gamma(n)},$$

and

$$\mathbb{E}[X\mathbf{1}_{\{X\geq K\}}] = n\left[1 - \frac{\gamma(n+1,K)}{\Gamma(n+1)}\right],$$

where Γ and γ are the gamma function and the incomplete gamma function, respectively.

In the second example we set $X = \sum X_i$ where X_i are i.i.d. *Bernoulli* variables with $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p = 0.15$. Its CGF is given by $\mathcal{K}(t) = n\log\left(1 - p + pe^t\right)$. Here the saddlepoint to $\mathcal{K}'(t) = K$ equals $T = \log\left[\frac{K(1-p)}{(n-K)p}\right]$ and

$$\mathcal{K}''(T) = \frac{K(n-K)}{n}, \lambda_3 = \frac{n-2K}{\sqrt{nK(n-K)}}, \lambda_4 = \frac{n^2 - 6nK + 6K^2}{nK(n-K)}.$$

In this specific case, X is binomially distributed with

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

which means that C and S as defined in (7.4) can also be calculated exactly.

We report in Tables 7.1 and 7.2 on the approximations obtained in the exponential case and in Tables 7.3 and 7.4 approximations in the Bernoulli case. For the approximations to S we take $S_r = C_r/P_r + K$ for $r = 1,2,3,4$. The saddlepoint approximations in the Bernoulli case are based on the formulas for lattice variables derived in section 7.6.

In general we see that all approximations work remarkably well in our experiments. The higher order Lugannani-Rice type formula, S_4, C_4 and their lattice versions, produce almost exact approximations. Particularly worth mentioning is the quality of approximations C_4 and $\hat{C}_4$, that use the same information as C_1 and $\hat{C}_1$, but show errors that are significantly smaller than C_2 and $\hat{C}_2$.

7.8 Conclusions

We have derived two types of saddlepoint approximations to $\mathbb{E}[(X-K)^+]$ and $\mathbb{E}[X|X \geq K]$, where X is the sum of n independent random variables and K is a known constant. For each type of approximation, we have

n	K	Exact	C_1	C_2	C_3	C_4
10	10.5	1.0375	1.0791	1.0465	1.0478	1.0374
10	12.5	0.4529	0.5048	0.4528	0.4608	0.4528
10	14.5	0.1761	0.2007	0.1747	0.1808	0.1761
100	105	2.0331	2.0852	2.0341	2.0360	2.0331
100	125	3.7283e-2	3.8873e-2	3.7240e-2	3.7508e-2	3.7283e-2
100	145	9.5270e-5	9.8546e-5	9.5210e-5	9.6553e-5	9.5269e-5

TABLE 7.1: Exact values of $\mathbb{E}[(X-K)^+]$ and their saddlepoint approximations. $X = \sum_{i=1}^{n} X_i$ where X_i is exponentially distributed with density $f(x) = e^{-x}(x \geq 0)$.

n	K	Exact	S_1	S_2	S_3	S_4
10	10.5	13.1124	12.9673	13.1382	13.1383	13.1124
10	12.5	14.7482	14.7512	14.7582	14.7874	14.7481
10	14.5	16.5066	16.5654	16.4998	16.5599	16.5065
100	105	111.7826	111.7313	111.7883	111.7924	111.7826
100	125	128.9751	129.0343	128.9715	128.9990	128.9571
100	145	147.9199	147.9626	147.9175	147.9592	147.9199

TABLE 7.2: Exact values of $\mathbb{E}[X|X \geq K]$ and their saddlepoint approximations. $X = \sum_{i=1}^{n} X_i$ where X_i is exponentially distributed with density $f(x) = e^{-x}(x \geq 0)$.

n	K	Exact	$\hat{C}_1$	$\hat{C}_2$	$\hat{C}_3$	$\hat{C}_4$
15	3	2.6017e-1	2.8790e-1	2.6656e-1	2.7583e-1	2.5980e-1
15	4	8.2821e-2	9.2946e-2	8.4153e-2	9.0764e-2	8.2651e-2
15	5	2.1115e-2	2.3480e-2	2.1364e-2	2.4050e-2	2.1057e-2
100	18	4.2046e-1	4.3660e-1	4.2330e-1	4.2579e-1	4.2045e-1
100	23	2.3355e-2	2.4313e-2	2.3529e-2	2.4075e-2	2.3353e-2
100	28	4.2976e-4	4.4395e-4	4.3281e-4	4.5375e-4	4.2969e-4

TABLE 7.3: Exact values of $\mathbb{E}[(X-K)^+]$ and their saddlepoint approximations. $X = \sum_{i=1}^{n} X_i$ where X_i is Bernoulli distributed with $p(X_i = 1) = 0.15$.

given a lower order version and a higher order version. We have also established the error convergence rates for the approximations in the *i.i.d.* case. The approximations have been further extended to cover the case of lattice variables. Numerical examples show that all these approximations work remarkably well. The Lugannani-Rice type formulas to $\mathbb{E}[(X-K)^+]$ are particularly attractive because of their simplicity.

7.A Alternative derivation of the Lugannani-Rice type formulas

The approximation J_3 in (7.44) can also be derived by a more statistical approach. Let us replace the density of X by its saddlepoint approximation (7.8), we then obtain

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx \frac{1}{\sqrt{2\pi}}\int_K^\infty x\frac{e^{\mathcal{K}(t)-xt}}{\sqrt{\mathcal{K}''(t)}}\left[1+\frac{\lambda_4(t)}{8}-\frac{5\lambda_3(t)^2}{24}\right]\mathrm{d}x \tag{7.63}$$

where $x=\mathcal{K}'(t)$. Let again $K=\mathcal{K}'(T)$. A change of variables from x to t gives

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx \frac{1}{\sqrt{2\pi}}\int_T^\infty \mathcal{K}'(t)\sqrt{\mathcal{K}''(t)}e^{\mathcal{K}(t)-\mathcal{K}'(t)t}\left[1+\frac{\lambda_4(t)}{8}-\frac{5\lambda_3(t)^2}{24}\right]\mathrm{d}t$$

Let $w^2/2=\mathcal{K}'(t)t-\mathcal{K}(t)$ and $W^2/2=\mathcal{K}'(T)T-\mathcal{K}(T)$ so that $w\mathrm{d}w=t\mathcal{K}''(t)\mathrm{d}t$, $t=0\Leftrightarrow w=0$, $t=T\Leftrightarrow w=W$. A second change of variables from t to w gives

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx \frac{1}{\sqrt{2\pi}}\int_W^\infty e^{-\frac{w^2}{2}}\frac{w\mathcal{K}'(t)}{t\sqrt{\mathcal{K}''(t)}}\left[1+\frac{\lambda_4(t)}{8}-\frac{5\lambda_3(t)^2}{24}\right]\mathrm{d}w,$$

which is precisely in the form of eq. (3.2.1) in Jensen (1995). According to Theorem 3.2.1 therein, one finds

$$\begin{aligned}\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx& [1-\Phi(W)]\left\{q(0)\left[1+\frac{\lambda_4(0)}{8}-\frac{5\lambda_3(0)^2}{24}\right]+\frac{q''(0)}{2}\right\}\\ &+\phi(W)\frac{q(W)-q(0)}{W},\end{aligned} \tag{7.64}$$

where $q(w)=\frac{w\mathcal{K}'(t)}{t\sqrt{\mathcal{K}''(t)}}$. Let $\tilde{q}(w)=\frac{w}{t\sqrt{\mathcal{K}''(t)}}$, then $q(w)=\mathcal{K}'(t)\tilde{q}(w)$.

Lemma 7.3.

$$\tilde{q}(w) = 1 - \frac{1}{6}\lambda_3(0) + \left[\frac{5}{24}\lambda_3(0)^2 - \frac{1}{8}\lambda_4(0)\right] w^2 + O(|w|^3),$$
$$\tilde{q}(0) = 1, \qquad \tilde{q}''(0) = -2\left[\frac{\lambda_4(0)}{8} - \frac{5\lambda_3^2(0)}{24}\right],$$
$$t = \frac{1}{\sqrt{\mathcal{K}''(0)}}\left[w - \frac{1}{3}\lambda_3(0)w^2 + O(|w|^3)\right].$$

Proof. See Jensen(1995) Lemma 3.3.1. □

According to Lemma 7.3, we have

$$q(0) = \mu, \quad q(W) = \frac{W\mathcal{K}'(T)}{T\sqrt{\mathcal{K}''(T)}}, \tag{7.65}$$

$$q''(w) = \tilde{q}''(w)\mathcal{K}'(t) + 2\tilde{q}'(w)\mathcal{K}''(t)\frac{\mathrm{d}t}{\mathrm{d}w} + \tilde{q}(w)\left[\mathcal{K}'''(t)\left(\frac{\mathrm{d}t}{\mathrm{d}w}\right)^2 + \mathcal{K}''(t)\frac{d^2t}{\mathrm{d}w^2}\right],$$

where $\frac{\mathrm{d}t}{\mathrm{d}w} = \frac{1}{\sqrt{\mathcal{K}''(0)}}\left[1 - \frac{2}{3}\lambda_3(0)w\right]$, $\frac{d^2t}{\mathrm{d}w^2} = \frac{-2\lambda_3(0)}{3\sqrt{\mathcal{K}''(0)}}$. When $w = 0$ we find

$$\begin{aligned} q''(0) =& -2\left[\frac{\lambda_4(0)}{8} - \frac{5\lambda_3^2(0)}{24}\right]\mu + 2\left[-\frac{\lambda_3(0)}{6}\right]\frac{\mathcal{K}''(0)}{\sqrt{\mathcal{K}''(0)}} \\ &+ \frac{\mathcal{K}'''(0)}{\mathcal{K}''(0)} + \mathcal{K}''(0)\frac{-2\lambda_3(0)}{3\sqrt{\mathcal{K}''(0)}} \\ =& -2\left[\frac{\lambda_4(0)}{8} - \frac{5\lambda_3^2(0)}{24}\right]\mu. \end{aligned} \tag{7.66}$$

Plugging (7.65) and (7.66) in (7.64) we again get

$$\mathbb{E}\left[X\mathbf{1}_{\{X\geq K\}}\right] \approx \mu\left[1 - \Phi(W)\right] + \phi(W)\left[\frac{\mathcal{K}'(T)}{T\sqrt{\mathcal{K}''(T)}} - \frac{\mu}{W}\right] \equiv J_3. \tag{7.67}$$

n	K	Exact	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_4$
15	3	3.6574	3.8166	3.6702	3.6952	3.6563
15	4	4.4670	4.5614	4.4711	4.5099	4.4660
15	5	5.3422	5.3991	5.3425	5.3879	5.3412
100	18	19.7762	19.9213	19.7874	19.7984	19.7761
100	23	24.0548	24.1191	24.0619	24.0870	24.0547
100	28	28.7012	28.7330	28.7053	28.7400	28.7011

TABLE 7.4: Exact values of $\mathbb{E}[X|X \geq K]$ and their saddlepoint approximations. $X = \sum_{i=1}^{n} X_i$ where X_i is Bernoulli distributed with $p(X_i = 1) = 0.15$.

Chapter 8

Conclusions and Outlook

8.1 Conclusions

In this dissertation we have investigated the measurement and allocation of portfolio credit risk with a factor model.

We have described a new procedure to embed the saddlepoint approximation as a useful tool in portfolio credit loss modeling. The saddlepoint approximation is applied to the conditional moment generating function of the portfolio loss given the common factor in the Vasicek one-factor model. The saddlepoint approximations, esp. the higher order approximations, are able to produce accurate results on both the VaR and the VaR contribution. The ES and ES contribution can also be computed satisfactorily. We have also illustrated that the saddlepoint approximation works well for small-sized portfolios and portfolios with exposure concentration, where Vasicek's asymptotic formulas fail. We further point out that the saddlepoint approximation is a flexible method which can be applied in quite general situations, for example, multi-factor models, non-Gaussian factor models and models with random LGD.

Moreover we have examined various numerical methods, including the normal approximation, saddlepoint approximation, importance sampling, for the purpose of calculating the credit portfolio VaR and VaRC under the Vasicek one-factor model. We find that each method provides a viable solution to VaR/VaRC estimation for lower granular portfolios and portfolios with medium exposure concentration. However there is no perfect method that prevails under all circumstances and the choice of preferred method turns out to be a trade-off among speed, accuracy and robustness.

The Generalized Beta Regression framework for modeling systematic risk in loss given default (LGD) is proposed in the context of credit port-

folio losses. The GBR framework provides great flexibility in random LGD modeling and accommodates well skewness and heteroscedastic errors. The quantities in the GBR models have simple economic interpretation. We have shown that parameter estimation and model selection are straightforward in this framework. Moreover, it has been demonstrated that the portfolio loss distribution can be efficiently evaluated by both the normal approximation and the saddlepoint approximation.

For the calculation of the tail probability in multi-factor credit portfolio loss models, we have proposed algorithms based on adaptive integration, with either a deterministic multiple integration rule or a Monte Carlo type random rule. Both algorithms are asymptotically convergent and consistently outperform the plain Monte Carlo method. The adaptive Monte Carlo integration algorithm is able to provide reliable probabilistic error bounds. To be able to take advantage of the adaptive integration algorithm, an assumption of nonnegative coefficients in the multi-factor model is made. A dedicated algorithm for the nonnegative factorization of a correlation matrix has been presented to make sure that this assumption is satisfied.

Finally, two types of saddlepoint approximations to $\mathbb{E}[(X-K)^+]$ and $\mathbb{E}[X|X \geq K]$, where X is the sum of n independent random variables and K is a known constant, are derived. For each type of approximation, we have given a lower order version and a higher order version. We have also established the error convergence rates for the approximations in the *i.i.d.* case. The approximations have been further extended to cover the case of lattice variables. Numerical examples show that all these approximations work remarkably well. The Lugannani-Rice type formulas to $\mathbb{E}[(X-K)^+]$ are particularly attractive because of their simplicity.

8.2 Outlook

We have only considered default-mode models in which loss only occurs when an obligor defaults in a fixed time horizon. However, credit losses can also come from the deterioration in an obligor's creditworthiness, particularly for exposures with longer maturity. It would be a useful extension to quantify portfolio credit risk on a marked-to-market basis that explicitly measures the potential impact of both defaults and credit migrations.

Another straightforward extension would be replacing the Gaussian distribution of the common factor(s) by Lévy models, as outlined in Albrecher et al. (2007). The Gaussian model has long been criticized for producing too light tails and having no tail dependence. Especially since the emer-

gence of the credit crisis, the spreads of the senior tranches of market indices (e.g., CDX) have widened dramatically to a level that the industry standard Gaussian copula model can not produce even with 100% correlation. As the Lévy models are capable of generating heavier tail than the Gaussian copula model, they can be utilized to provide a better fit to the market price.

A different approach to improve the performance of the Gaussian copula model is taking into account the randomness in the LGD. Along this direction, the GBR models for LGD proposed in Chapter 4 can be readily incorporated in the valuation of CDOs.

The Lévy models and the random LGD model can in fact be combined conveniently. The saddlepoint approximations for $\mathbb{E}[(X-K)^+]$ derived in Chapter 7 can be immediately employed for CDO pricing with such a Lévy model with random LGD.

Empirically, the GBR models for LGD should be tested against real world LGD data, when the dataset is available.

It would also be interesting to explore portfolio credit losses models beyond the factor models in which the obligors are independent conditional on a realization of the common factors. An attractive alternative would be the contagion models, in which the default of a firm has a negative impact on the other firms. These models are said to be able to explain better the *clustering of default*.

Bibliography

Abate, J., Choudhury, G. L. & Whitt, W. (2000), An introduction to numerical transform inversion and its application to probability models, *in* W. K. Grassman, ed., 'Computational Probability', Kluwer, Norwell, MA.

Acerbi, C. & Tasche, D. (2002), 'On the coherence of expected shortfall', *Journal of Banking and Finance* **26**(7), 1487–1503.

Albrecher, H., Ladoucette, S. & Schoutens, W. (2007), A generic one-factor Lévy model for pricing synthetic cdos, *in* M. C. Fu, R. A. Jarrow, J.-Y. J. Yen & R. J. Elliott, eds, 'Advances in Mathematical Finance', Birkhäuser, Boston.

Altman, E. I., Brady, B., Resti, A. & Sironi, A. (2005), 'The link between default and recovery rates: Theory, empirical evidences, and implications', *Journal of Business* **78**(6), 2203–2227.

Andersen, L. & Sidenius, J. (2004), 'Extensions to the Gaussian copula: Random recovery and random factor loadings', *Journal of Credit Risk* **1**(1), 29–70.

Andersen, L., Sidenius, J. & Basu, S. (2003), 'All your hedges in one basket', *RISK* (November), 67–72.

Annaert, J., Garcia, J. B. C., Lamoot, J. & Lanine, G. (2006), Don't fall from the saddle: the importance of higher moments of credit loss distributions, Technical Report 06/367, Faculty of Economics and Business Administration, Ghent University, Belgium.

Antonov, A., Mechkov, S. & Misirpashaev, T. (2005), Analytical techniques for synthetic CDOs and credit default risk measures, Technical report, Numerix.

Antonov, I. A. & Saleev, V. M. (1979), 'An economic method of computing LP_τ-sequences', *USSR Computational Mathematics and Mathematical Physics* **19**, 252–256.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Basel Committee on Bank Supervision (2005), 'Basel II: International convergence of capital measurement and capital standards: a revised framework'.

Beran, J. & Ocker, D. (2005), 'Small sample asymptotics for credit risk portfolios', *Journal of Computational & Graphical Statistics* **14**(2), 339–351.

Berntsen, J. (1989), 'Practical error estimation in adaptive multidimensional quadrature routine', *Journal of Computational and Applied Mathematics* **25**(3), 327–340.

Berntsen, J., Espelid, T. O. & Genz, A. (1991), 'An adaptive algorithm for the approximate calculation of multiple integrals', *ACM Transactions on Mathematical Software* **17**(4), 437–451.

Berry, M., Brown, M., Langville, A., Pauca, V. & Plemmons, R. (2007), 'Algorithms and applications for approximate nonnegative matrix factorization', *Computational Statistics & Data Analysis* **52**(1), 155–173.

Blackwell, D. & Hodges, J. L. (1959), 'The probability in the extreme tail of a convolution', *The Annals of Mathematical Statistics* **31**, 1113–1120.

Bluhm, C., Overbeck, L. & Wagner, C. (2002), *An Introduction to Credit Risk Modeling*, Chapman & Hall / CRC.

Boyle, P., Broadie, M. & Glasserman, P. (1997), 'Monte carlo methods for security pricing', *Journal of Economic Dynamics and Control* **21**, 1267–1321.

Breslow, N. & Clayton, D. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* **88**, 9–25.

Bruche, M. & González-Aguado, C. (2008), Recovery rates, default probabilities and the credit cycle, Discussion papers, CEMFI. Available at SSRN: http://ssrn.com/abstract=934348.

Catral, M., Han, L., Neumann, M. & Plemmons, R. (2004), 'On reduced rank nonnegative matrix factorization for symmetric nonnegative matrices', *Linear Algebra and its Applications* **393**, 107–126.

Credit Suisse Financial Products (1997), *CreditRisk$^+$: a CreditRisk management framework*, Credit Suisse Financial products.

Crouhy, M., Galai, D. & Mark, R. (2000), 'A comparative analysis of current credit risk models', *Journal of Banking & Finance* **24**(1), 59–117.

Daniels, H. E. (1954), 'Saddlepoint approximations in statistics', *The Annals of Mathematical Statistics* **25**(4), 631–650.

Daniels, H. E. (1987), 'Tail probability approximations', *International Statistical Review* **55**, 37–48.

Davison, A. & Wang, S. (2002), 'Saddlepoint approximations as smoothers', *Biometrika* **89**(4), 933–938.

Debuysscher, A. & Szegö, M. (2003), The Fourier transform method - technical document, Technical report, Moody's Investors Service.

Denault, M. (2001), 'Coherent allocation of risk capital', *Journal of Risk* **4**(1), 7–21.

Düllmann, K. & Trapp, M. (2004), Systematic risk in recovery rates - an empirical analysis of US corporate credit exposures, Deutsche Bundesbank Discussion Paper Series 2: Banking and Financial Supervision, No 02/2004.

Esscher, F. (1932), 'On the probability function in the collective theory of risk', *Skandinavisk Aktuarietidskrift* **15**, 175–195.

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta regression for modeling rates and proportions', *Journal of Applied Statistics* **31**, 799–815.

Frye, J. (2000), 'Depressing recoveries', *RISK* (November), 108–111.

Genz, A. (1984), Testing multiple integration software, *in* B. Ford, J.-C. Rault & F. Thommaset, eds, 'Tools, Methods and Languages for Scientific and Engineering Computation', North Holland, New York, pp. 208–217.

Genz, A. & Kass, R. (1997), 'Subregion adaptive integration of functions having a dominant peak', *Journal of Computational and Graphical Statistics* **6**(1), 92–111.

Genz, A. & Malik, A. (1980), 'An adaptive algorithm for numerical integration over an n-dimensional rectangular region', *Journal of Computational and Applied Mathematics* **6**(4), 295–302.

Giese, G. (2005), 'The impact of PD/LGD correlations on credit risk capital', *RISK* (April), 79–84.

Giese, G. (2006), 'A saddle for complex credit portfolio models', *RISK* (July), 84–89.

Glasserman, P. (2006), 'Measuring marginal risk contributions in credit portfolios', *Journal of Computational Finance* **9**(2), 1–41.

Glasserman, P. & Li, J. (2005), 'Importance sampling for credit portfolios', *Management Science* **51**(11), 1643–1656.

Glasserman, P. & Ruiz-Mata, J. (2006), 'Computing the credit loss distribution in the Gaussian copula model: a comparison of methods', *Journal of Credit Risk* **2**(4), 33–66.

Glasserman, P., Kang, W. & Shahabuddin, P. (2008), 'Fast Simulation of Multifactor Portfolio Credit Risk', *Operations Research* **56**(5), 1200–1217.

Golub, G. & van Loan, C. (1996), *Matrix computations*, 3rd edn, The Johns Hopkins University Press, London.

Gordy, M. B. (2002), 'Saddlepoint approximation of CreditRisk$^+$', *Journal of Banking and Finance* **26**, 1335–1353.

Gordy, M. B. (2003), 'A risk-factor model foundation for ratings-based bank capital rules', *Journal of Financial Intermediation* **12**(3), 199–232.

Gourieroux, C., Laurent, J.-P. & Scaillet, O. (2000), 'Sensitivity analysis of values at risk', *Journal of Empirical Finance* **7**, 225–245.

Groenen, P. & Pietersz, R. (2007), 'Efficient rank reduction of correlation matrices', *Linear Algebra and its Applications* **422**(2-3), 629–653.

Gupton, G. & Stein, R. (2002), *LossCalc: Model for predicting loss given default (LGD)*, Moody's Investors Service.

Gupton, G., Finger, C. & Bhatia, M. (1997), *CreditMetrics – Technical document*, J.P.Morgan New York.

Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized Additive Models*, New York: Chapman & Hall.

Hillebrand, M. (2006), 'Modelling and estimating dependent loss given default', *RISK* (September), 120–125.

Hirschmann, A. O. (1964), 'The paternity of an index', *American Economic Review* **54**, 761–762.

Hu, Y. T. & Perraudin, W. (2002), The dependence of recovery rates and defaults, Working paper, Birkbeck College.

Jensen, J. (1995), *Saddlepoint Approximations*, Oxford University Press.

Kalkbrener, M. (2005), 'An axiomatic approach to capital allocation', *Mathematical Finance* **15**(3), 425–437.

Kalkbrener, M., Lotter, H. & Overbeck, L. (2004), 'Sensible and efficient capital allocation for credit portfolios', *RISK* (January), S19–S24.

Kealhofer, S. (2001), *Portfolio Management of Default Risk*, KMV Corporation.

Kolassa, J. E. (2006), *Series Approximation Methods in Statistics*, 3rd edn, Springer, New York.

Lawson, C. L. & Hanson, B. J. (1974), *Solving Least Squares Problems*, Prentice-Hall.

Lee, D. D. & Seung, H. S. (1999), 'Learning the parts of objects by non-negative matrix factorization', *Nature* **401**(6755), 788–791.

Lee, Y. & Nelder, J. A. (1998), 'Generalized linear models for the analysis of quality-improvement experiments', *Canadian Journal of Statistics* **26**(1), 95–105.

Lee, Y. & Nelder, J. A. (2001), 'Hierarchical generalized linear models: a synthesis of generalized linear models, random effect models and structured dispersions', *Biometrika* **88**, 987–1006.

Lepage, G. P. (1978), 'A new algorithm for adaptive multidimensional integration', *Journal of Computational Physics* **27**, 192–203.

Lepage, G. P. (1980), Vegas: An adaptive multi-dimensional integration routine., Technical Report CLNS-80/447, Newman Laboratory of Nuclear Studies, Cornell University.

Lugannani, R. & Rice, S. (1980), 'Saddlepoint approximations for the distribution of the sum of independent random variables', *Advances in Applied Probability* **12**, 475–490.

Lyness, J. N. & Kaganove, J. J. (1976), 'Comments on the nature of automatic quadrature routines', *ACM Transactions on Mathematical Software* **2**(1), 65–81.

Martin, R. (2004), *Credit portfolio modeling handbook*, Credit Suisse Financial products.

Martin, R. (2006), 'The saddlepoint method and portfolio optionalities', *RISK* (December), 93–95.

Martin, R., Thompson, K. & Browne, C. (2001*a*), 'Taking to the saddle', *RISK* (June), 91–94.

Martin, R., Thompson, K. & Browne, C. (2001*b*), 'VAR: who contributes and how much?', *RISK* (August), 99–102.

Mausser, H. & Rosen, D. (2004), Allocating credit capital with VaR contributions, Technical report, Algorithmics Inc.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, London: Chapman & Hall.

McNeil, A. J. & Wendin, J. P. (2006), 'Dependent credit migrations', *Journal of Credit Risk* **2**(3), 87–114.

McNeil, A. J. & Wendin, J. P. (2007), 'Bayesian inference for generalized linear mixed models of portfolio credit risk', *Journal of Empirical Finance* **14**, 131–149.

Merino, S. & Nyfeler, M. A. (2005), 'Applying importance sampling for estimating coherent credit risk contributions', *Quantitative Finance* **4**, 199–207.

Merton, R. C. (1974), 'On the pricing of corporate debt: the risk structure and interest rates', *The Journal of Finance* **29**, 449–470.

Nelder, J. A. & Lee, Y. (1991), 'Generalized linear models for the analysis of Taguchi-type experiments', *Applied Stochastic Models and Data Analysis* **7**, 107–120.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society: Series A* **135**(3), 370–384.

Pan, J. & Thompson, R. (2007), 'Quasi-monte Carlo estimation in generalized linear mixed models', *Computational Statistics & Data Analysis* **51**, 5756–5775.

Pauca, V., Paul, P. & Plemmons, R. (2006), 'Nonnegative matrix factorization for spectral data analysis', *Linear Algebra and its Applications* **416**, 29–47.

Pietersz, R. & Groenen, P. (2004), 'Rank reduction of correlation matrices by majorization', *Quantitative Finance* **4**(6), 649–662.

Press, W. H. & Farrar, G. R. (1990), 'Recursive stratified sampling for multidimensional Monte Carlo integration', *Computers in Physics* **4**(2), 190–195.

Pykhtin, M. (2003), 'Unexpected recovery risk', *RISK* (August), 74–78.

Rogers, L. C. G. & Zane, O. (1999), 'Saddlepoint approximations to option prices', *The Annals of Applied Probability* **9**(2), 493–503.

Rösch, D. & Scheule, H. (2005), 'A multi-factor approach for systematic default and recovery risk', *Journal of Fixed Income* **15**(2), 63–75.

Rosen, J. (1960), 'The gradient projection method for nonlinear programming, part 1: Linear constraints', *SIAM Journal of Applied Mathematics* **8**, 181–217.

Schuermann, T. (2004), What do we know about loss given default?, *in* D. Shimko, ed., 'Credit Risk Models and Management, Second Edition', Risk Books, London, pp. 249–274.

Shelton, D. (2004), Back to normal. proxy integration: a fast accurate method for CDO and CDO-squared pricing, Technical report, Citigroup Global Structured Credit Research.

Studer, M. (2001), Stochastic Taylor expansions and saddlepoint approximations for risk management, PhD thesis, ETH Zürich.

Tasche, D. (1999), Risk contributions and performance measurement, Technical report, Technische Universitat Muenchen.

Tasche, D. (2002), 'Expected shortfall and beyond', *Journal of Banking and Finance* **26**(7), 1519–1533.

Tchistiakov, V., de Smet, J. & Hoogbruin, P.-P. (2004), 'A credit loss control variable', *RISK* (July), 81–85.

Thompson, K. & Ordovás, R. (2003), 'Credit ensembles', *RISK* (April), 67–72.

van Dooren, P. & de Ridder, L. (1976), 'An adaptive algorithm for numerical integration over an n-dimensional cube', *Journal of Computational and Applied Mathematics* **2**(3), 207–217.

Vandendorpe, A., Ho, N.-D., Vanduffel, S. & Dooren, P. V. (2008), 'On the parameterization of the creditrisk+ model for estimating credit portfolio risk', *Insurance: Mathematics and Economics* **42**(2), 736–745.

Vasicek, O. (2002), 'Loan portfolio value', *RISK* (December), 160–162.

Wallace, D. L. (1958), 'Asymptotic approximations to distributions', *The Annals of Mathematical Statistics* **29**(3), 635–654.

Wedderburn, R. (1974), 'Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method', *Biometrika* **61**(3), 439–447.

Wilde, T. (2001), 'Probing granularity', *RISK* (August), 103–106.

Wilson, T. (1997*a*), 'Portfolio credit risk I', *RISK* (September), 111–117.

Wilson, T. (1997*b*), 'Portfolio credit risk II', *RISK* (October), 56–61.

Wu, L. (2003), 'Fast at-the-money calibration of the libor market model using lagrange multipliers', *Journal of Computational Finance* **6**(2), 39–77.

Yang, J., Hurd, T. & Zhang, X. (2006), 'Saddlepoint approximation method for pricing CDOs', *Journal of Computational Finance* **10**(1), 1–20.

Zeng, B. & Zhang, J. (2001), *An Empirical Assessment of Asset Correlation Models*, KMV Corporation.

Zhang, Z. & Wu, L. (2003), 'Optimal low-rank approximation to a correlation matrix', *Linear Algebra and its Applications* **364**, 161–187.

Zheng, H. (2006), 'Efficient hybrid methods for portfolio credit derivatives', *Quantitative Finance* **6**, 349–357.

Curriculum vitae

Xinzheng Huang was born in Wuxi, PR China on December 13, 1979. After he finished his high school at Wuxi No.1 high school in July 1997, he went to Fudan University in Shanghai to study statistics, where he obtained his Bachelor of Science degree in 2001. From September 2003 to August 2005 he attended the Master program in Applied Mathematics at the Delft University of Technology. He received his Master of Science diploma with his thesis entitled *Numerical Valuation of American Options under Exponential Lévy Processes*, under the supervision of prof. Cornelis W. Oosterlee. In December 2005, he continued his study at the Delft University of Technology as a PH.D student. In June 2009 he joined Fortis Bank Nederland as a quantitative risk analyst.

List of publications

1. Huang, X. & Oosterlee, C. W (2009), 'Saddlepoint approximations for expectations', submitted.

2. Sonneveld, P., van Kan, J. J. I. M., Huang, X. & Oosterlee, C. W (2009), 'Nonnegative matrix factorization of a correlation matrix', *Linear Algebra and its Applications*, **431**(3-4), 334–349.

3. Huang, X. & Oosterlee, C. W (2009), 'Adaptive integration for multi-factor portfolio credit loss models', *Journal of Computational and Applied Mathematics*, **231**(2), 506–516.

4. Huang, X. & Oosterlee, C. W (2008), 'Generalized beta regression models for random Loss-Given-Default', submitted.

5. Huang, X., Oosterlee, C. W. & Mesters, M. (2007), 'Computation of VaR and VaR contribution in the Vasicek portfolio credit loss model: A comparative study', *Journal of Credit Risk* **3**(3), 75–96.

6. Huang, X., Oosterlee, C. W. & van der Weide, J. A. M. (2007), 'Higher order saddlepoint approximations in the Vasicek portfolio credit loss model', *Journal of Computational Finance* **11**(1), 93–113.