



Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Economic Capital for the Trading Book

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfilment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Adrien Chenailler
Delft, the Netherlands
August 2013

Copyright © 2013 by Adrien Chenailler. All rights reserved.



MSc THESIS APPLIED MATHEMATICS

Economic Capital for the Trading Book

Adrien Chenailier

Delft University of Technology

Daily supervisors

Prof. dr. ir. C.W. Oosterlee

M. van Buren, MSc

Responsible professor

Prof. dr. ir. C.W. Oosterlee

Other thesis committee members

Dr. F. Fang

Dr. F. van der Meulen

Dr. P. Cirillo

August 2013

Delft, the Netherlands

Abstract

Economic Capital consists of an internally defined amount of capital that is necessary to overcome adverse market conditions. It plays an important role in risk management and business decisions. This thesis focuses on the Economic Capital of the trading book of an international bank. Several types of risks need to be modelled and, in this thesis, two risks are investigated, namely market risk and credit risk. For market risk, a model is explained and two risk measures are analysed: Value at Risk and Expected Shortfall. The properties of the estimators of these risk measures are explained and a method to compute the one-year market risk component based on scaling a risk measure of a 10-day Profit and Loss distribution is derived. This method shows that Expected Shortfall is more appropriate than Value at Risk for modelling tail risk. The second part of this thesis focuses on the migration matrix employed in the model in order to capture the credit risk present in the trading book. Several methods are employed and compared, and a specific method is analysed to assess the probabilities of default in order to be consistent with other probabilities employed by the bank. Furthermore, several characteristics of a rating process are analysed, such as the Markovity and time-(in)homogeneity.

Acknowledgement

This thesis concludes a year of work in the Quantitative Risk Analytics team (QRA) at Rabobank International. This year was divided between my Masters thesis and an internship project. There are many people I want to thank, particularly my supervisors Kees Oosterlee and Martin van Buren. Kees offered many constructive inputs and feedback about this thesis during our numerous meetings and encouraged me to take this thesis to a more advanced mathematical level. Martin guided me on the world of financial risk management, while his comments and fruitful discussions were key ingredients behind the success of this thesis. I also express my gratitude to my colleagues from QRA for the more or less serious discussions and enjoyable time spent at Rabobank.

I also want to thank Fang Fang for her guidance and for being part of my committee, and the other members of my university committee, notably Pasquale Cirillo and Frank van der Meulen. Finally, many thanks go to my parents and family for their support during my studies. I also owe a lot to my friends in the Netherlands and around the world.

Contents

1	Introduction	1
2	Fundamentals of Financial Risk Management	5
2.1	Mathematical background	5
2.1.1	Definitions	5
2.1.2	Value at Risk	6
2.1.3	Expected Shortfall	10
2.2	Economic Capital and Regulatory Capital	14
2.2.1	Economic Capital for the trading book	14
2.2.2	Regulatory Capital for the trading book	14
2.2.3	Model requirements for Economic Capital	15
2.3	Financial Definitions	16
2.3.1	Liquidity horizon	17
2.3.2	Capital horizon	17
2.3.3	Constant risk assumption	17
3	Data Analysis	19
3.1	Profit and Loss	19
3.2	Historical Simulated PLs	21
3.2.1	Overview of the PLs	21
3.2.2	Properties of the historical PLs	22
3.2.3	Autocorrelation of the historical PLs	24
3.3	Stressed PLs	25
3.3.1	Overview	25
3.3.2	Relevance of stress tests for Economic Capital	26
4	Modelling the Economic Capital	27
4.1	Market Fluctuation Model	27
4.1.1	Definitions and issues	28
4.1.2	Approximations for Value at Risk and Expected Shortfall	28
4.1.3	Sampling algorithm	31
4.2	Incremental Risk Charge	32
4.2.1	Definition and motivations	32
4.2.2	Overview and inputs of the IRC model	33

5	Estimation of Value at Risk and Expected Shortfall	35
5.1	Definition of Errors	35
5.2	Non-Parametric Estimation of Value at Risk	36
5.2.1	Discrete estimators	36
5.2.2	Properties of the Value at Risk estimators	37
5.3	Non-Parametric Estimation of Expected Shortfall	40
5.3.1	Discrete estimators	40
5.3.2	Properties of the estimators	41
5.4	Application to Independent Distributions	43
5.4.1	Definition and methodology	43
5.4.2	Numerical results	43
5.5	Application to Autocorrelated Time Series	49
5.5.1	Processes and methodology	49
5.5.2	Numerical results	52
5.6	Conclusion and Further Research	57
6	Market Fluctuation Risk	58
6.1	Introduction and Model Requirements	58
6.2	Expected Shortfall versus Value at Risk	59
6.3	Theoretical Analysis of the Sampled VaR	61
6.3.1	VaR of a sum of continuous random variables	62
6.3.2	VaR of a sum of discrete random variables	63
6.4	Multi Criteria Analysis of the Scaling Factor	64
6.4.1	Stability of the scaling factor and the Economic Capital	64
6.4.2	Correlation with the sampled VaR	65
6.4.3	Error with the sampled VaR	67
6.5	Rounding and Analysis of the Scaling Factor	68
6.5.1	Analysis of the scaling factor	68
6.5.2	Requirements	70
6.5.3	Quantitative analysis	71
6.6	Impact Analysis	72
6.7	Conclusion and Further Research	74
7	Incremental Risk Charge and Migration Matrix	75
7.1	Problem Description and Definitions	75
7.1.1	Definitions and rating systems	76
7.1.2	Mapping to external ratings	78
7.1.3	Requirements and algorithm of the IRC	79
7.2	Estimation of Migration Matrices	81
7.2.1	Markov chains and definitions	81
7.2.2	The cohort method	83
7.2.3	Estimation of the generator	84
7.2.4	The Aalen-Johansen method	84
7.2.5	Examples and Comparison	85
7.2.6	The rating dynamics	87

7.3	The Computation of Probabilities of Default	89
7.3.1	Requirements	89
7.3.2	Methodology	90
7.3.3	Resulting probabilities of default	97
7.3.4	Historical Probabilities of Default	98
7.4	Computation of the Migration Matrix	99
7.4.1	Introduction and requirements	99
7.4.2	Estimation of the matrix	100
7.4.3	Regularization	101
7.5	Impact Analysis on the IRC	106
7.5.1	Matrices used for the numerical experiments	106
7.5.2	Numerical results for the IRC	108
7.6	Confidence Intervals	110
7.7	Advanced Models to Reflect non-Markovian Migrations	112
7.8	Conclusion and Further Research	117
8	Conclusion and Outlook	118
	References	120
A	Regulatory Capital Model	124
B	Background of Statistics	125
B.1	Definitions and Classical Theorems	125
B.2	Distribution of the Order Statistics	127
B.3	Expected Shortfall and Value at Risk of Classical Distributions	127
B.3.1	Normal distribution	127
B.3.2	Student's t-distribution	128
C	Tables from Chapter 4	129
C.1	Tables for Uncorrelated Time Series	129
C.2	Tables for Correlated Time Series	137
D	Migration Matrices	141

List of Figures

1.1	Evolution of capital held by banks and leverage.	1
2.1	Example of VaR.	7
2.2	Different tails with the same VaR.	10
2.3	Difference TCE and ES.	11
2.4	ES for a Normal distribution (no mean correction).	12
3.1	Overlapping PLs.	21
3.2	4 years of 10-day PLs.	21
3.3	1-day and 2-day PLs for a binary option.	22
3.4	Tails for different PL distributions.	23
3.5	QQplot 18/01/2013 (left) and 25/01/2013 (right).	23
3.6	QQplots of a 10-day PL distribution comparing different distributions.	24
3.7	Autocorrelation of the 10-day PLs.	25
4.1	Validity of the square root rule.	29
4.2	θ for historical PLs and for classic distributions.	30
4.3	Square Root rule for expected shortfall.	31
4.4	Sampling methodology.	32
4.5	Tail IRC distribution.	34
5.1	Bias of the VaR estimators.	44
5.2	Bias for ES, Student's t-distribution.	45
5.3	MSE for VaR, Student's t-distribution with 5 degrees of freedom.	46
5.4	MSE for ES.	46
5.5	Asymptotic normalized standard deviation.	47
5.6	Normalized theoretical standard deviation of the estimators.	48
5.7	Autocorrelation of the PLs.	50
5.8	Autocorrelation of AR and T5 processes.	51
5.9	Autocorrelation in the tail of AR and T5 processes.	52
5.10	Bias for VaR estimators, n=250.	53
5.11	Bias for ES estimators, n=250.	53
5.12	MSE VaR for T5 (left) and AR (right).	54
5.13	MSE ES for T5 (left) and AR (right).	54
5.14	Standard deviations for VaR based on the analytical formula.	55
5.15	Standard deviations for ES based on the analytical formula.	56

5.16	Theoretical standard deviation of the historical PLs (left) and autocorrelation in the tail (right).	56
6.1	Surfaces of VaR (left) and ES (right).	59
6.2	Ratio ES/VaR.	60
6.3	Correlation kurtosis/risk measures.	61
6.4	Density of X_1 knowing the value of the sum for densities , $C=VaR_\alpha(X, T)$	63
6.5	Probability of $X_1 = 0$ knowing the value of the sum for Bernoulli random variables. .	63
6.6	Probability of occurrence of a given PLs on the $\widehat{VaR}_{99.99\%}(X, T)$	64
6.7	NSTD of the scaling factor(top) and of the Market fluctuation risk (bottom) over time.	65
6.8	Correlation between ρ and VaR 99.99% 1-year for historical PLs (left) and standard distributions (right).	66
6.9	Error of the simulation with the sampled VaR.	67
6.10	Scaling factors for historical PLs(bottom) and for some distributions(top). . . .	70
6.11	Market fluctuation risk for several risk measures with nearest half rounding of the scaling factor for P1 (top) and P2 (bottom).	73
6.12	Market fluctuation risk for several risk measures based on the worst PL for P1(left) and P2 (right).	73
7.1	Algorithm for the IRC.	80
7.2	Evolution of the PDs.	88
7.3	Ratio of the $PD_{Ri}(\Delta t)$ to the 1-year PDs.	91
7.4	R2 term structure.	91
7.5	Comparison of GM and AJ methods to extract the PDs.	94
7.6	Scatter plot γ_i , $i = 1, \dots, 17$, (left) and scatter plot $\log(PD_{3/12})$ (right) computed with Equations (7.23) and (7.24).	94
7.7	Linear regression on the γ_i , $i = 1, \dots, 17$, in log space.	95
7.8	Linear interpolated γ_i , $i = 1, \dots, 20$	97
7.9	Term structures of the probabilities of default.	98
7.10	Ratio between extracted PDs from internal rating based PDs and historical 3-month PDs.	99
7.11	Migration matrices computed with AJ and GM methods.	100
7.12	Smoothing of the diagonal elements.	102
7.13	Smooth 3-month migration matrix and for R3 rating.	104
7.14	On the left: sum of M^+ and M^- , the probability to migrate for the AJ matrix and M . On the right: M^+ and M^- , the probability to be upgraded or downgraded for the AJ matrix and M	105
7.15	Final migration matrix.	106
7.16	Tail distributions of the IRC.	108
7.17	Sensitivity to the parameters.	109
7.18	Distance between the migration matrices.	114

Chapter 1

Introduction

The 2007-2008 financial crisis shocked the banks' practice and their risk management frameworks. The regulatory supervision has been made stricter after 30 years of financial deregulation and sometimes not fully controlled innovations (see [47]). These financial innovations had for consequences that banks transferred many banking activities to the trading book. Between 2000 and 2009, the part of the trading book in the total assets increased from 20% to 40% (see [37]). A large part of this increase was due to the creation of new financial derivatives and especially credit derivatives. The consequences were a higher liquidity of the banks' assets but also more volatility in their portfolios and a higher leverage. The risk management of these products failed and banks had to cover large losses during the crisis. Banks had underestimated the risk of these new products. During the aftermath of the crisis, one of the main criticisms focused on the capital that banks were holding for their trading and banking books. This buffer, which is supposed to cover large losses, appeared to be underestimated during the crisis, which is the consequence of two factors, the capital that banks hold, had decreased over the last century and the leverage had increased over the decade preceding the crisis (see Figure 1.1).

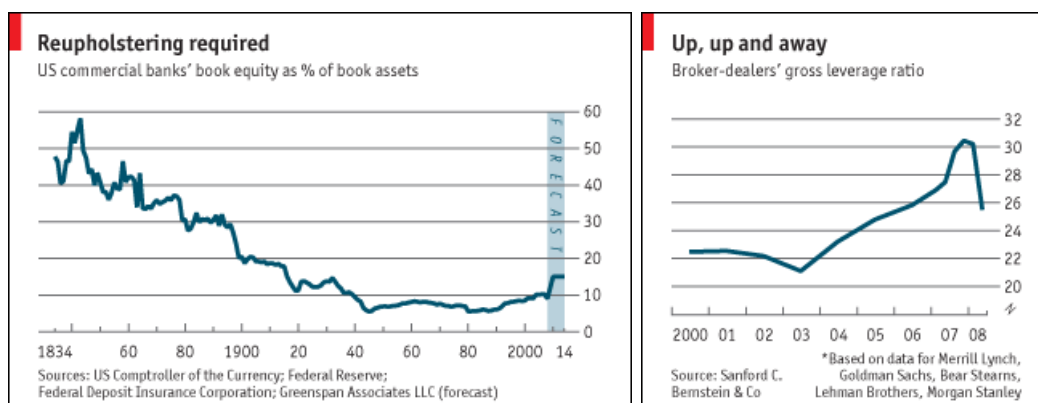


Figure 1.1: Evolution of capital held by banks and leverage.

A bank has to face and manage many risks. These risks are very different but may all have huge consequences. Credit risk is the risk a bank takes when lending money. Typically, the risk is that a borrower does not reimburse a credit. Market risk is a second risk type, which is the risk that a bank takes at the financial market (mainly in its trading book). Two other important risks are the liquidity risk (risk that an asset cannot be traded quickly without a loss) and the

operational risk (the risk of the activities of a bank)¹ This thesis focuses on the financial risks that are in the trading book. There are many risks in the trading book that a bank should cover. These are not covered by one formula. Some measures of risk are computed on a daily basis and are used for the business of a bank. Other measures of market risk only need to be computed on a weekly basis and are reported to the regulators such as the central banks. Regulators are entities at several levels: world, European and national. The most global regulator is the Basel Committee, which is responsible for the development of international rules that then, are approved by governments and central banks. The documents that have been issued by this committee are the Basel I (1998), II (2004), 2.5 (2008) and III (2010) agreements. The influence of the Basel Committee is worldwide. The next level of financial supervision (for the Netherlands) is the European commission, the European parliament and the European Central Bank (ECB). They interpret and translate the Basel agreements into European laws and regulations. The European regulations are known as CRD (Capital Requirement Directive). Currently CRD IV is being implemented. Banks do not report to the European regulators but to their national central banks, De Nederlandsche Bank (DNB) for the Netherlands.

The development of the Economic Capital is closely related to risk measurement development and was first employed by Banker's trust in the 1970's (see [20]). It is now used by every banks and financial institutions. For now, the Economic Capital may be seen as a buffer that a bank thinks it should hold to cover possible losses. This definition will be made more precise in the thesis but there is no universal definition of the Economic Capital. Furthermore, it has to reflect all types of risk (market, credit...). The methodologies usually employed are very diverse for the different risks. Modelling operational risk does not have a lot in common with modelling credit risk, for instance. Therefore, it is common to model them individually and take the sum (possibly with some diversification effects) to have the total Economic Capital of a bank.

The purpose of the thesis is to model the Economic Capital of the trading book [43]. Two major financial risks are modelled, these are the market risk and the credit risk. The difference between market risk and trading book is that the trading book is a group of products, portfolios that are considered to mainly have market risk or that are actively traded. However, there may also be credit risk involved, for instance. "Actively traded" is quite an abstract notion but it mainly states that even a bond quoted on the financial market may be placed outside the trading book if the strategy is to keep it until maturity without any trading involved. The market risk is a type of risk and may be present outside the trading book as well. In this thesis, the assets that are not in the trading book are supposed to be in the banking book. Further, the model presented capitalizes for all the risks present in the trading book.

This thesis deals with several aspects of the modelling of the Economic Capital and may be divided into two parts, one for the market risk (Chapters 3, 4, 5 and 6) and the second part is for the credit risk (Chapters 4 and 7) of the trading book. The sum of the two gives the main part of the Economic Capital for the trading book. For the credit risk, a general model is only explained and the contribution of this thesis is about the inputs, the migration matrix and the probabilities of default.

¹Others types of risk exist such as: reputational risk, settlement risk, profit risk and systemic risk. These are difficult to model and are not as important as the four main risks.

Concerning the market risk, a first aspect is the problem of the data. What data should be used? Does it have any special properties? The amount of data appears to be rather small and is highly autocorrelated, which diminishes its quality and these have an impact on the estimation of the risk measures. Value at Risk has been used for a long time. In this thesis, we also investigate a change to Expected Shortfall and try to determine if such a change is relevant and possible. Given that the dataset is small and autocorrelated, legitimate questions are: What are the statistical properties of the risk measure estimators? Are they accurate and appropriate for Economic Capital modelling? A last objective of the thesis is to analyse the migration matrix and probabilities of default, employed for the credit risk of the trading book, and to compute them accurately while taking in consideration the properties of a rating process. A method to derive the migration matrix is provided and then, a detailed analysis of the Markovity of the rating process is performed. We answer questions such as: Is a classical migration matrix appropriate for the credit risk of the Economic Capital? Is a rating process a Markov process? Is it possible to model the migrations in another way?

This thesis is organized in six main chapters.

The first step is to give precise mathematical definitions to what a risk is and what the concepts that are available to model risk are. In Section 2.1, the mathematical concept of a risk measure is introduced for a general random variable. Some properties are also given for specific risk measures and they are interpreted in financial terms. Then, in Section 2.2 the Economic Capital is defined and linked to another capital: the Regulatory Capital. Requirements for the Economic Capital are also provided.

Chapter 3 provides an overview and an analysis of the data that is available to model the Economic Capital. Subsequently, the concept of Profit and Loss is introduced as a major ingredient in risk measurement.

Based on the theory of financial risk management and the data that is available, a general model for the Economic Capital is presented in Chapter 4. Two of its main components are investigated in this thesis, one captures the market risk of the trading book (market fluctuation risk, modelled in Section 4.1), and the other captures the credit risk (Incremental Risk Charge, explained in Section 4.2). This chapter also investigates the relevance of some classical practice in risk management such as the time horizon scaling of the a risk measure.

The following chapter is more theoretical. One of the main topics of the thesis is to investigate a change from the well known Value at Risk, risk measure to Expected Shortfall. Therefore, in Chapter 5 the statistical properties of the estimators of the two risk measures are analysed. The purpose is to understand the properties of the estimator of the risk measures with autocorrelated time series mimicking the data used in practice for the Economic Capital.

Using the results about the estimators of Value at Risk and Expected Shortfall of Chapter 5 and the model defined in Section 4.1, we find a manageable and accurate approximation of the market fluctuation risk based on a simple scaling of the a risk measure (Chapter 6).

Finally, the last topic is the Incremental Risk Charge (Chapter 7), which captures the second main financial risk, the credit risk of the trading book. After giving the general algorithm employed to compute the IRC, we investigate the main inputs: the migration matrix and the probabilities of default. Several methods are compared and confidence intervals are found. The output should be a stable and reliable migration matrix. A detail analysis of the non-Markovity of the rating process is also performed to understand the rating process better and we try to obtain a more accurate model.

Chapter 2

Fundamentals of Financial Risk Management

In this chapter, we introduce mathematical concepts needed to understand the notion of risk. The definitions and objectives of the Economic Capital and the Regulatory Capital are given. Finally, we give precise definitions of some useful financial terms for Economic Capital modelling.

2.1 Mathematical background

This section provides mathematical background for risk measurement. The concept of risk measure is introduced and then two classical risk measures are defined.

Some basics of market risk measurement as well as vocabulary and notations are introduced. In the thesis, we use a general definition of inverse cumulative/quantile function. In the literature, this is sometimes written as $F_X^{\leftarrow}(x)$. But this notation just adds complexity and the difference only happens in cases not present in this thesis, so we use

$$F_X^{-1}(1-p) = q_{1-p}(X) = \inf\{x | F_X(x) \geq 1-p\}, \quad (2.1)$$

where X is a random variable, F_X its cumulative distribution function and $p \in (0,1)$ is the threshold level.

2.1.1 Definitions

The common definition of a risk measure may be found in the dictionary [6]: A quantitative measure of risk attempts to assess the degree of variation or uncertainty about earnings or returns.

This definition is very general and needs to be specified from a mathematical point of view.

Definition 2.1.1 (Risk measure, [49])

Let (Ω, \mathcal{F}, P) be a probability space and V a non-empty set of \mathcal{F} -measurable real-valued random variables. Then, any mapping $\rho : V \rightarrow \mathbb{R} \cup \{+\infty\}$ is called a risk measure.

With this definition any mapping can be a risk measure. For example, the expectation and the variance are risk measures.

Some properties are very natural and desirable. For instance, if a portfolio contains a risky asset with distribution X and b of cash then the measures of risk $\rho(X + b)$ and $\rho(X) - b$ should be the same. The interpretation is that if there is cash in the portfolio, then the capital requirement of the portfolio is diminished by the amount of cash.

Another desirable property is the monotonicity also named first order stochastic dominance. It implies that if X and Y are two risky portfolios and $P(X < t) > P(Y < t) \forall t \in \mathbb{R}$, then, $\rho(X) > \rho(Y)$ should hold.

Definition 2.1.2 (Monetary risk measure, [49])

A risk measure $\rho : V \rightarrow \mathbb{R} \cup \{+\infty\}$ is called a monetary risk measure if $\rho(0)$ is finite and if ρ satisfies the following conditions for all $X, Y \in V$:

Monotonicity: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

Cash invariance: If $b \in \mathbb{R}$, then $\rho(X + b) = \rho(X) - b$.

A basic example of a monetary risk measure is Value at Risk (VaR) discussed in Section 2.1.2. However, one may remark that some properties that may be desirable are not in Definition 2.1.2. One is the sub-additivity, if X and Y are two risky portfolios then, due to diversification effects, it should hold that the risk of the two portfolios together should not be greater than the sum of the risks of the two portfolios taken separately. Finally, one may request that hX , where h is a positive scalar, is h -times riskier than X .

Adding these two properties to Definition 2.1.2 defines a coherent risk measure (Definition 2.1.3). A coherent risk measure does not discourage investment.

Definition 2.1.3 (Coherent risk measure, [49])

Let X and Y be two random variables.

A monetary measure of risk ρ is a coherent measure of risk if it satisfies the following properties:

Sub-additivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Positive homogeneity: If $h \geq 0$, then $\rho(h.X) = h.\rho(X)$.

These two conditions are rather restrictive and some very common risk measures (for instance Value at Risk) are not coherent.

2.1.2 Value at Risk

Definitions and first examples

The Value at Risk (VaR) has been first used for financial activities by Kenneth Garbade, a banker working for Bankers Trust Cross Market, during the 80's to measure the risk across all the different portfolios. However his work had little impact on the banking industry. The measure has been popularised by JPMorgan with their system, RiskMetrics (see [20] for the complete history of VaR). The aim of VaR is to measure the maximum likely loss given a certain distribution.

Nowdays, VaR is clearly the risk measure which is most commonly used in the financial industry, it can be defined mathematically as follows:

Definition 2.1.4 (Value at Risk (VaR))

The Value at Risk of a random variable X at the confidence level (also called threshold level) α is

given by the largest number x such that the probability that the loss x exceeds the loss $\text{VaR}_\alpha(X)$ is at most $(1 - \alpha)$. Mathematically, if X is a random variable, then VaR_α is defined as:

$$\begin{aligned}\text{VaR}_\alpha(X) &= -\inf\{x \in \mathbb{R} : P(X < x) \geq 1 - \alpha\}, \\ &= -\inf\{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\},\end{aligned}\tag{2.2}$$

where F_X is the Cumulative Distribution Function (CDF) of random variable X .

Remark 2.1.1 In case of a non zero mean, the VaR plus the mean are often used for risk management. This is done because it makes the risk measure more stable. Furthermore, it is mathematically sound to avoid negative VaR in case the mean of random variable is very high.

A simple way to characterize the VaR is by the inverse cumulative distribution function or quantile function as defined in Section 2.1:

$$\text{VaR}_\alpha(X) = -F_X^{-1}(1 - \alpha).\tag{2.3}$$

Example 2.1.1 We consider the basic case where the random variable X , with cumulative distribution function F_X , follows a normal distribution with mean $\mu = 0.5$ (a profit of 0.5 is expected) and variance $\sigma^2 = 1$. VaR is computed for two thresholds: $\alpha = 97.5\%$ and $\alpha = 99.99\%$:

$$\text{VaR}_{97.5\%}(X) = -F^{-1}(1 - 0.975) = -\mu - \sigma\phi^{-1}(1 - 0.975) = 1.46.$$

$$\text{VaR}_{99.99\%}(X) = -F^{-1}(1 - 0.9999) = -\mu - \sigma\phi^{-1}(1 - 0.9999) = 3.2190.$$

Figure 2.1 provides a graphical representation of $\text{VaR}_{97.5\%}(X)$.

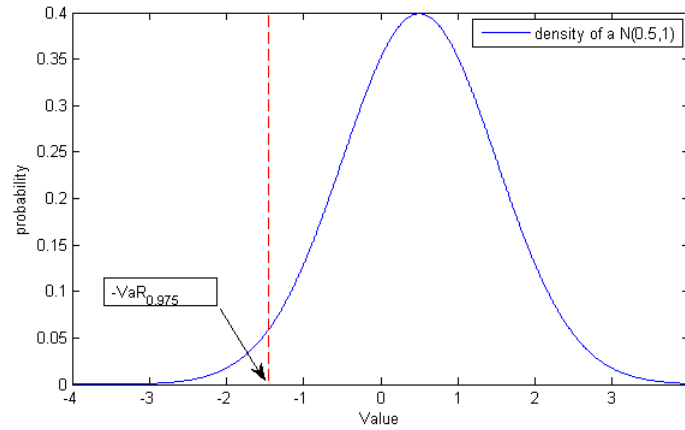


Figure 2.1: Example of VaR.

Along the years VaR has become the main risk measure for the Economic Capital or the Regulatory Capital. International regulators recommended its use in the Basel II agreement (see [38]). Indeed, it is easily understandable and has some nice properties that are investigated in the next section.

Properties

In this section, properties of VaR are explained and analysed. They are summarized in the following:

Property 2.1.1

Let $\alpha \in (0, 1]$ and X, Y be two random variables on a probability space (Ω, \mathcal{F}, P) . Let VaR be the risk measure as defined in Definition 2.1.4, the following properties hold:

- (1) **Monotonicity:** if $X \leq Y$ then $\text{VaR}_\alpha(X) \geq \text{VaR}_\alpha(Y)$.
- (2) **Positive homogeneity :** If $h > 0$ then $\text{VaR}_\alpha(h \cdot X) = h \cdot \text{VaR}_\alpha(X)$.
- (3) **Translation(cash) invariance :** If $b \in \mathbb{R}$ then $\text{VaR}_\alpha(X + b) = \text{VaR}_\alpha(X) - b$.
- (4) **Law invariance:** If $P(X \leq t) = P(Y \leq t)$, $\forall t \in \mathbb{R}$, then, $\text{VaR}_\alpha(X) = \text{VaR}_\alpha(Y)$.

Therefore, VaR is a monetary measure of risk.

Proof:

The four properties are proved independently.

(1) **Monotonicity:** Suppose $X \leq Y$ then $P(X < t) \geq P(Y < t) \forall t \in \mathbb{R}$. Therefore, it holds that: if $x \in \{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\}$ then $x \in \{x \in \mathbb{R} : F_Y(x) \geq 1 - \alpha\}$, Therefore, the following inclusion holds:

$$\{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\} \supset \{x \in \mathbb{R} : F_Y(x) \geq 1 - \alpha\},$$

and we have:

$$\begin{aligned} \inf\{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\} &\leq \inf\{x \in \mathbb{R} : F_Y(x) \geq 1 - \alpha\}, \\ -\inf\{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\} &\geq -\inf\{x \in \mathbb{R} : F_Y(x) \geq 1 - \alpha\}. \end{aligned}$$

(2) **Positive homogeneity:** This is obvious from the fact that $F_{h \cdot X}(x) = F_X(\frac{x}{h})$.

(3) **Translation invariance:** This is immediate from the fact that $F_{X+b}(x) = F_X(x - b)$.

(4) **Law invariance:** If the two cumulative distribution functions are the same then $F_X(t) = F_Y(t)$, $\forall t \in \mathbb{R}$, and the result is straightforward.

□

Despite these properties, one of the main critics to VaR, by the academics, is the lack of sub-additivity (see [36]).

Property 2.1.2 (non sub-additivity)

VaR is not sub-additive and, therefore, is not a coherent risk measure.

We provide a counter-example:

Suppose X and Y are two independent identically distributed random variables such that for both variables with probability 99.1% the outcome is 0 and with probability 0.09% the outcome is -10.

Then we have:

$$\text{VaR}_{0.99}(X) = \text{VaR}_{0.99}(Y) = 0.$$

However, the sum $X+Y$ as the following distribution : With probability 98.21 % the outcome is 0, with probability 0.0081% the outcome is -20 and with probability 1.7819% the result is -10. Therefore, we have:

$$\text{VaR}_{0.99}(X + Y) = 10,$$

and,

$$\text{VaR}_{0.99}(X + Y) > \text{VaR}_{0.99}(X) + \text{VaR}_{0.99}(Y).$$

This contradicts the sub-additivity property. \square

In practice, VaR is very often sub-additive. For most distributions, it is sub-additive in the tail which is the part that is relevant for the Economic Capital. It is the case if the cumulative distribution function decays exponentially, see [13] for more details about the sub-additivity of classical distributions. Sub-additivity in the tail for VaR is defined as follows:

$$\textbf{Sub-additivity in the tail : } \text{VaR}_\alpha(X + Y) \leq \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y), \quad \forall \alpha \geq A,$$

where $A \in (0, 1)$.

Remark 2.1.2 *Being a non-coherent risk measure is not the main drawback. Non-coherency happens mainly in pathological cases that do not happen for Economic Capital because the whole portfolio of a bank is summed up. It is likely to contain a large range of products creating a smooth distribution with a more or less a fat tail. It usually resembles a normal distribution or Student's t-distribution.*

The main problem resides in the fact that, VaR does not fully capture the tail but only one value: the quantile. This may lead to undesirable situations because the behaviour in the end of the tail may be very volatile when only few data are available. VaR does not tell anything about the potential losses beyond its threshold.

In Figure 2.2, the two distributions have the same $\text{VaR}_{90\%}$ but not the same distribution at all. Common sense tells that the second portfolio is riskier. In fact, it is stochastic dominance of order 1. Stochastic dominance is defined as a dominance of the cumulative distribution function (see Equation (2.4)). Let X and Y be random variables with cumulative distribution functions F and G then, Y dominates X (of order 1) if:

$$F(x) \leq G(x) \quad \forall x \in \mathbb{R} \text{ and } \exists x_0 \text{ such as } F(x_0) < G(x_0). \quad (2.4)$$

In Figure 2.2, the cumulative distribution function of random variable in the bottom graph is always greater or equal that one at the top and at the point -250 this inequality is strict.

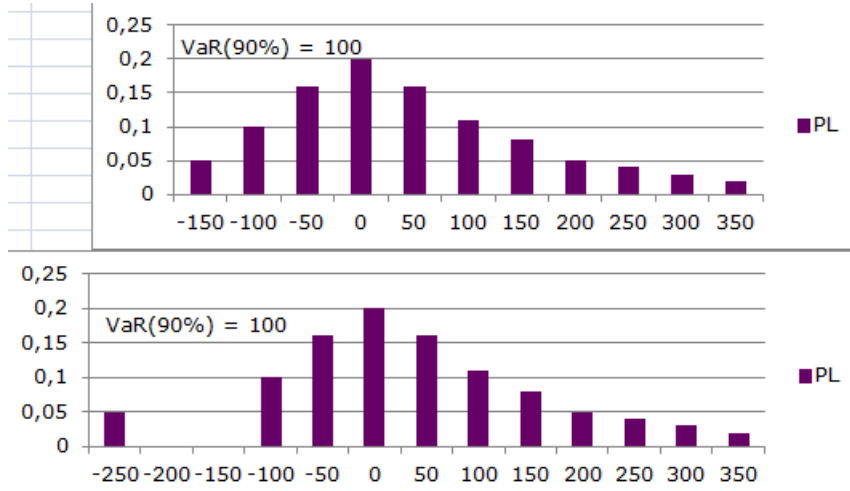


Figure 2.2: Different tails with the same VaR.

Other drawbacks of VaR are explained in [42]. A major problem is that VaR is very sensitive to modelling choices, which is shown the Basel committee report [44]. Several methods, used by banks, to compute VaR are compared and the results are very different for one bank to another (from 1 to 10 for certain asset classes). Another example is described in [50] where it is stated JP Morgan decreases its Economic Capital by changing how VaR is computed.

A last issue, addressed in [42], is the effective use of VaR. It may be the case that when managing the risks with VaR, dealers try to do trades that are not affecting this VaR with a certain threshold.

2.1.3 Expected Shortfall

Expected Shortfall (ES) is emerging as the future main risk measure for market risk. The Basel Committee and the national regulators recommend to start using it in [43] for the Regulatory Capital. The literature has been recommending its use for a long time (see [48],[54], [42] and [49]). Indeed, ES has many desirable properties and is smoother than VaR because it takes in consideration more values in the tail.

Definition and first examples

First, the concept of ES is introduced and some basic examples are provided.

Definition 2.1.5 (Expected Shortfall)

The Expected Shortfall (ES) of a random variable X on a probability space (Ω, \mathcal{F}, P) and such that $X \in L^p(\Omega)$ for a threshold level α , is defined as:

$$\begin{aligned} \text{ES}_\alpha(X) &= \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_t(X) dt = \frac{-1}{1-\alpha} \int_\alpha^1 F_X^{-1}(1-t) dt, \\ &= -\frac{1}{1-\alpha} \mathbb{E}[X \cdot \mathbb{I}_{X \leq F_X^{-1}(1-\alpha)} + F_X^{-1}(1-\alpha)(1-\alpha - P[X \leq F_X^{-1}(1-\alpha)])], \end{aligned} \quad (2.5)$$

where F_X is the cumulative distribution function of X .

Remark 2.1.3 The mean of the distribution is often added to VaR for risk management. Therefore, it is also often added to ES.

The formal definition is rather technical because it involves integrals and/or inverse cumulative distribution function. Therefore, a more manageable concept is introduced: The tail conditional expectation.

Definition 2.1.6 *The Tail Value at Risk (TVaR) or Tail Conditional Expectation (TCE) of a random variable X on a probability space (Ω, \mathcal{F}, P) , and such that $X \in L^p(\Omega)$ for a level α , is defined as:*

$$\text{TCE}_\alpha(X) = -E[X|X \leq -\text{VaR}_\alpha(X)]. \quad (2.6)$$

At first sight, the definitions of TCE and ES seem the same. It is true for continuous random variables. The TCE takes the average of all values below the VaR, but for discrete random variables there may be more values than $(1 - \alpha) \cdot n$ in the TCE the outcomes. If $(1 - \alpha) \cdot n$ is not an integer than the TCE is the ES with threshold level $\lceil (1 - \alpha) \cdot n \rceil$. Figure 2.3 shows a counter example where TCE is not equal to ES:

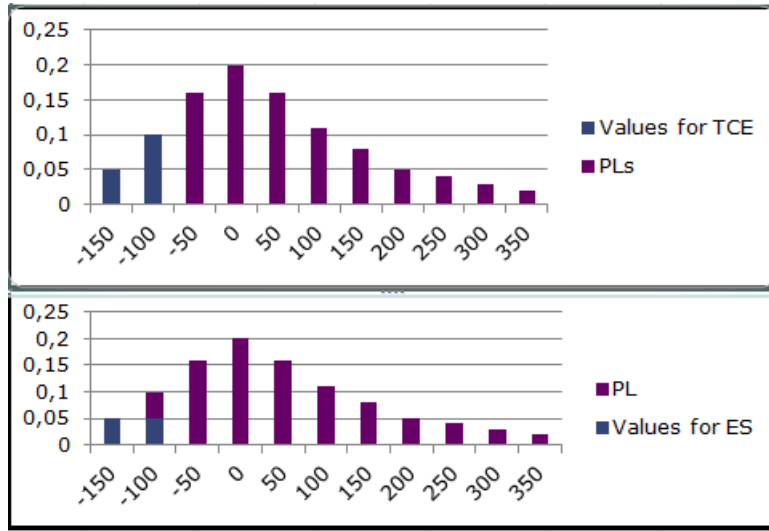


Figure 2.3: Difference TCE and ES.

In this case, the TCE is 116,7 and the ES is 125.

Consider a normally distributed random variable X with mean μ , standard deviation σ , cumulative distribution function F_X and density f_X .

For normal distribution, an analytical formula may be derived for a given α (it is equal to the TCE) and for $\mu = 0$ (this is just to make the computation easier, the mean may just be added to the final result):

$$\begin{aligned} \text{ES}_\alpha(X) &= -\frac{1}{1-\alpha} \mathbb{E}[X|X \leq -\text{VaR}_\alpha(X)] = -\frac{1}{1-\alpha} \int_{-\infty}^{-\text{VaR}_\alpha(X)} x f_X(x) dx, \\ \text{ES}_\alpha(X) &= -\frac{1}{1-\alpha} \int_{-\infty}^{-\text{VaR}_\alpha(X)} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx. \end{aligned}$$

This can be integrated directly:

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\text{VaR}_\alpha(X)^2}{2\sigma^2}},$$

Example 2.1.2 Suppose X is a normal random variable with parameters $\mu = 0.5$ and $\sigma = 1$. ES is the following:

$$ES_{0.975}(X) = \frac{1}{0.025} \frac{1}{\sqrt{2\pi}} e^{\frac{(-\text{VaR}_\alpha(X)+0.5)^2}{2}} - 0.5 = 1.8376.$$

Figure 2.4 shows the concept of ES graphically.

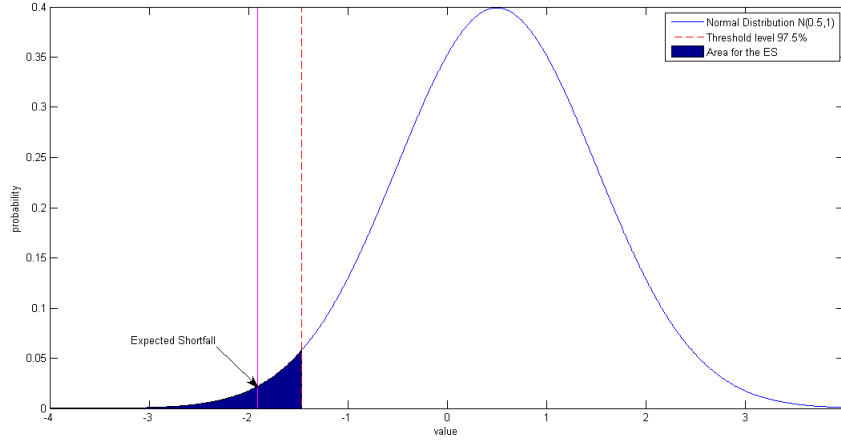


Figure 2.4: ES for a Normal distribution (no mean correction).

Properties

The first property represents one of the main theoretical differences between VaR and ES . Unlike VaR , ES is sub-additive. Therefore, it does not discourage investment. This is one of the reason why the regulators and the Basel committee support a change from VaR to ES in [43]. The following statement summarizes the main properties of ES :

Property 2.1.3

ES is monotonic, cash invariant, sub-additive and homogeneous. Therefore, it is a coherent risk measure.

Proof:

The four properties are proved independently:

Let $\alpha \in (0, 1]$ and X, Y be two random variables on a probability space (Ω, \mathcal{F}, P) .

- *Monotonicity*

Suppose $X \leq Y$ then, $VaR_\alpha(X) \geq VaR_\alpha(Y)$. We integrate on both sides and scale by $\frac{1}{1-\alpha}$. Then, we use the properties of the integral:

$$\frac{1}{1-\alpha} \int_\alpha^1 VaR_t(X) dt \geq \frac{1}{1-\alpha} \int_\alpha^1 VaR_t(Y) dt.$$

- *Cash invariance*

Suppose $b \in \mathbb{R}$, then by cash invariance of VaR

$$\frac{1}{1-\alpha} \int_{\alpha}^1 \text{VaR}_t(X+b) dt = \frac{1}{1-\alpha} \left(\int_{\alpha}^1 \text{VaR}_t(X) dt - b \right).$$

- *Sub-additivity*

First we define $q_{1-\alpha}(X) := F_X^{-1}(1-\alpha)$, and:

$$\mathbb{I}_{X \leq q_{1-\alpha}(X)}^q = \mathbb{I}_{X \leq q_{1-\alpha}(X)} + \frac{1-\alpha - F_X(q_{1-\alpha}(X))}{\text{Prob}(X = q_{1-\alpha}(X))} \mathbb{I}_{X=q_{1-\alpha}(X)},$$

where F_X is the cumulative distribution function, and $\text{Prob}(X = q_{1-\alpha}(X))$ is the regular \mathbb{P} measure if X is a discrete random variable and equals 1 if has a density at the point $q_{1-\alpha}(X)$. The following holds:

$$\mathbb{E}[\mathbb{I}_{X \leq q_{1-\alpha}(X)}^q] = 1 - \alpha,$$

and

$$0 \leq \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q \leq 1.$$

We want to show that

$$A := (1-\alpha) \cdot (-\text{ES}_{\alpha}(X) - \text{ES}_{\alpha}(Y) + \text{ES}_{\alpha}(X+Y)) \leq 0.$$

We use the second definition of ES (2.1.5).

$$A = -\mathbb{E}[(X+Y) \cdot \mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q] + \mathbb{E}[X \cdot \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q] + \mathbb{E}[Y \cdot \mathbb{I}_{Y \leq q_{1-\alpha}(Y)}^q].$$

We group the terms depending on X and on Y :

$$A = \mathbb{E}[X(-\mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q + \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q) + Y \cdot (\mathbb{I}_{Y \leq q_{1-\alpha}(Y)}^q - \mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q)].$$

Using that $X \cdot \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q \leq q_{1-\alpha}(X)$ and $Y \cdot \mathbb{I}_{Y \leq q_{1-\alpha}(Y)}^q \leq q_{1-\alpha}(Y)$ and the following:

$$\begin{cases} -\mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q + \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q \leq 0 & \text{if } X > q_{1-\alpha}(X), \\ -\mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q + \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q \geq 0 & \text{if } X < q_{1-\alpha}(X). \end{cases}$$

Then we can write:

$$A \leq q_{1-\alpha}(X) \cdot \mathbb{E}[\mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q - \mathbb{I}_{X \leq q_{1-\alpha}(X)}^q] + q_{1-\alpha}(Y) \cdot \mathbb{E}[-\mathbb{I}_{Y \leq q_{1-\alpha}(Y)}^q + \mathbb{I}_{X+Y \leq q_{1-\alpha}(X+Y)}^q],$$

$$A \leq q_{1-\alpha}(X)(\alpha - \alpha) + q_{1-\alpha}(Y)(\alpha - \alpha) = 0.$$

This proves the sub-additivity property.

- *Positive homogeneity:*

Let $h \geq 0$, then

$$\text{ES}_{\alpha}(h.X) = \frac{1}{1-\alpha} \int_{\alpha}^1 \text{VaR}_t(h.X) dt.$$

By positive homogeneity of VaR we have:

$$\text{ES}_{\alpha}(h.X) = \frac{h}{1-\alpha} \int_{\alpha}^1 \text{VaR}_t(X) dt = h \cdot \text{ES}_{\alpha}(X),$$

This proves the positive homogeneity of ES.

□

From a mathematical point of view, ES is better than VaR because it does not discourage investment (sub-additivity) and it captures the tail risk instead of unique value (see [54]). These few properties of ES provide some qualitative arguments in favour ES.

2.2 Economic Capital and Regulatory Capital

This section gives high level definitions of the concepts and the objectives of both Economic Capital and Regulatory Capital. Both are supposed to represent amounts of capital. Their values are called Capital because they represent (theoretical) amounts of cash/money necessary to survive or cover large losses over a given period of time. They are two very close concepts and the modelling choices for one have consequences on the other.

2.2.1 Economic Capital for the trading book

Economic Capital is usually defined as the capital level that a financial institution would hold to cover losses with a certain probability or confidence level, which is often related to a desired rating [25] in absence of regulation. A rating expresses an opinion about the ability and the willingness of an issuer to meet its financial obligations [52]. A probability of default (over a certain period of time) may be assigned to these ratings. Therefore, the capital should cover all the possible scenarios up to this probability.

An alternative definition is given in [16]: The Economic Capital is defined as the amount of money a financial institution thinks it should hold in order to face adverse market conditions. The second definition is more general as it does not refer to credit ratings.

The Economic Capital has many properties and objectives.

It is an internal model that is mainly used within the bank. For instance, it is used in the computation of the Risk-Adjusted Return On Capital (RAROC) which represents the profitability divided by the amount of Economic Capital the bank should hold for a certain business line. It is one of the main decision tools for investment decisions. Therefore, the Economic Capital has a direct impact on the commercial activities of a bank.

Unlike the Regulatory Capital, banks are not required to hold the Economic Capital. It is, however, mandatory for international banks to have their own Economic Capital model and be able to justify the modelling choices [38]. Unlike for the Regulatory Capital too, banks have a lot of freedom when modelling and implementing it. However, there are some usual requirements that should be satisfied and these are explained in Section 2.2.3.

2.2.2 Regulatory Capital for the trading book

The Regulatory Capital is defined as the amount of capital the bank is required to hold as buffer for adverse market conditions [16].

The first characteristic is that the amount of capital required is given by a formula defined by the regulator (for international banks). However, some freedom is given for the implementation to take in consideration the risk characteristics of each bank. This implementation has to be

validated by the national regulator of the country.

Although the Basel committee does not explicitly state that the objective is to cover 99.9% of the possible scenarios that may happen over one year, other terms, such as the IRC or the credit risk, of the Regulatory Capital are to be computed with this threshold (see [39] and [38]). It implies that the amount of money a bank has to hold is (theoretically) exceeded one year in a thousand years.

The current model for the trading book is not a model that is meant to last for a long time. It has been proposed by the Basel Committee after the credit crisis and the Lehman Brothers bankruptcy (see [43], [39]) and was supposed to be completely implemented in 2012. The primary goal of the current model is to increase quickly the Regulatory Capital requirements. It is sometimes called Basel 2.5, which reflects the fact that it is a temporary model.

The complete model is provided in Appendix A.

2.2.3 Model requirements for Economic Capital

In this thesis, we are interested in modelling the Economic Capital. Although, the Regulatory Capital is not in scope of this thesis, we may refer to it to assess a possible convergence of the two models or a divergence, if necessary. The issues below give a range of requirements given by the management of banks and by the Basel Committee in [41]. A first requirement is simply that it should cover all the risks present in the trading book. Of course, it should cover the market risk, but it should also take in consideration eventual credit risks.

Confidence levels and ratings

As mentioned in the definition of Economic Capital (Section 2.2.1), the level of risk determines the bank's objective in terms of credit rating. The credit rating is an indicator of the riskiness of a borrower. High ratings are often related to low interest rates (low risk) when borrowing money. In [25], it is shown that an S&P rating of AA (the third best rating out of 23 ratings) implies that the Economic Capital of the bank is sufficient to cover 99.96% of the possible scenarios over one year. It may be inferred that a bank looking for the best rating possible (AAA) should cover 99.99% of the 1-year scenarios. This decision is usually taken by the board of a bank. We are interested in modelling an Economic Capital of a bank aiming for the best rating. As stated in Section 2.2.1, the Economic Capital is mainly used for internal purposes and investment choices. Therefore, a safe bank either has a high Economic Capital or selects safe clients and investments (or a mix of these two strategies).

Stability of the Economic Capital through the cycle

The credit crisis has demonstrated some modelling practice that should be avoided: one is a capital model too much dependent on the current market conditions. For instance, an Economic Capital based on the market fluctuation of a one-year rolling period (last year) did not appear to be appropriate. When the 2008-crisis hit, banks were punished twice: they were making large losses and their capital requirements were increasing giving rise to a higher vulnerability because banks were not able to reach this level of capital.

Allocation of the Economic Capital

An important requirement of Economic Capital is that it has to be usable by the business side. Therefore, it should be possible to allocate a part of the total Economic Capital to a certain business unit depending on the risk this unit takes. Capital allocation is not discussed in this thesis.

Consistency with Regulatory Capital

Consistency may be a useful requirement for the Economic Capital. It implies that a bank allocates the risk to business lines and takes investment decisions in concordance with the capital it holds as regulatory obligation. On the other hand, keeping two measures of risk allows to detect different types of risks: One type of risk may be captured by one measure and not by the other. The latter option limits the model risk, which is the risk that the model is not accurate. The drawbacks of modelling the risk with only one risk measure are investigated in [54]. The author shows that several risk measures should be used to measure the risk accurately.

Transparency of the framework

The model adopted should be transparent and simple enough in order to be understood by the users of the Economic Capital within the bank (see [41]). As it is used for business decisions, having a model which is understandable helps the users to understand its significance and the weight they should give to it in a decision making process. This process is known as the ICAAP: Internal Capital Adequacy Assessment Process.

Validation of the model

As mentioned in Section 2.2.1, the model has to be explained to the regulator. Therefore, the model should be based on realistic and understandable assumptions. Furthermore, a bank has to show that the model is suitable to be used by the business. Basic verifications have to be performed such as sensitivity analysis, qualitative review and data quality checks [41]. This is also part of the ICAAP.

2.3 Financial Definitions

This section provides some technical definitions that are needed to understand choices made when modelling the Economic Capital. Typical questions are:

- How much time should a bank suffer a loss on a particular position without the possibility to exit the position?
- How much time should the Economic Capital cover?

The following definitions discuss these issues.

2.3.1 Liquidity horizon

In practice, the liquidity horizon is the amount of time a bank suffers a loss without the possibility to exit the position. Formally, it is defined as follows:

Definition 2.3.1 (Liquidity horizon, [43])

Liquidity horizon is the time required to exit or hedge a risky position in a stressed market environment without materially affecting the market price.

A typical liquidity horizon is 10 days. This is the one used for the Regulatory Capital. An example to understand this concept is the following: 10000 stocks of Apple are by far easier to sell or hedge than 10000 Bermudean swaptions (complex financial product) Danish bond to Swiss bond. The reason is that the products are very different. On one hand, Apple stock is liquid (20,000,000 stocks exchanged per day on average) and is quoted on the stock exchange. On the other hand, Bermudean swaptions are Over The Counter (OTC) products, they have no quotation and are more confidential derivatives.

Therefore, 10 days to exit an Apple position may be a long of time, but it is not long for certain OTC products. So a liquidity horizon longer than a day is justified for some products.

2.3.2 Capital horizon

The second relevant question was for how long a bank should capitalize for the Economic Capital. How long may a stress period last? This issue is covered in the so-called capital horizon.

Definition 2.3.2 (Capital horizon, [38])

The capital horizon is the amount of time for which a bank or a financial institution is required to capitalize.

If a bank is required to capitalize for one year and the liquidity horizon is 10 days, then, it means that it is supposed to resist a one year crisis and it is capitalizing for 25 times the 10-day liquidity horizon (250 trading days per year are assumed and the liquidity horizon is 10 days).

2.3.3 Constant risk assumption

This section defines an important assumption of most of the current risk models.

Assumption 2.3.1 (Constant risk assumption, [39])

The level of risk remains the same at the start of each liquidity horizon.

At the beginning of each 10-day period a bank rebalances its positions in order to have the same risk as at the beginning of the simulation (the previous period). This assumption is conservative because it implies that if there is large loss from a certain trading position, then one may try to hedge it at the end the 10-day period and, at the beginning of the next period this loss can occur again. A concrete example is in case of default. Suppose that a bank has a large bond position on France and France goes bankrupt on a given 10-day period. Then, France may go bankrupt again during the next period and the bankruptcy of France is counted twice within the same scenario.

This concept was introduced by the Basel committee in [39] for the Incremental Risk Charge (see Section 4.2).

This concept reflects the fact that even if a bank may exit the position on one period of time, it will continue to make business in the next period and, therefore, will keep taking risks.

The opposite assumption would be a constant position assumption. It would mean that if the market is going down and the bank loses money then, the risk of the bank should decrease. If at time 0 a bank hold 4000€ and after 6 months the value of the bank's positions is only 2000€ then, the bank should have less risk for the remaining 6 months of the year it has to capitalize.

Chapter 3

Data Analysis

Based on the definitions from Chapter 2, the financial data needed and available to model the Economic Capital is introduced and explained.

Data is an essential ingredient when modelling Economic Capital. At most banks, the availability of data is rather restrictive and this may influence modelling choices. This Chapter also contains some important modelling decisions because the data is a part of the model.

3.1 Profit and Loss

First, the concepts of Profit & Loss (PL), PL distribution and stress test are introduced in a financial context. Then, two types of PLs are distinguished: simulated PLs and realized PLs. If not specified otherwise, in the thesis the abbreviation PLs is used for simulated PLs. They are the basis of the model.

Definition 3.1.1 (Simulated PL)

A simulated PL is the potential profit or loss given a certain scenario for a given portfolio.

In practice, it means that predetermined variations are applied to financial products of a portfolio and the total variation of the portfolio value gives a simulated PL. This scenario may last more than one day. In this thesis, 10-day scenarios are used.

Typically, a scenario is an event that happened in the past, is based on an expert opinion or is the result of Monte Carlo simulation.

Definition 3.1.2 (Realized PL)

A realized PL is an observed variation of a portfolio at a given day.

This is a profit or loss of the bank at a certain day with the portfolio of that day. Realized PLs are used to backtest a model. They represent the historical variations of a portfolio.

A PL distribution is a random variable X on a probability space (Ω, \mathcal{F}, P) where Ω is a set of PLs (discrete or continuous). If the set is discrete, the probability measure P taken is a uniform discrete distribution. A PL distribution may be all simulated PLs of today's portfolio based on daily scenarios of the last two years. In that case, the cardinal of Ω is 500 and each of the PLs has probability $\frac{1}{500}$.

Definition 3.1.3 (Stress test)

A stress test is an extreme, expert based scenario affecting a certain risk factor of an asset class (equity, interest rate, ...) or determined by a complete scenario (Euro crisis, economic meltdown...).

Stress tests are very common at banks. They are simulated PLs in which the scenarios are expert based. They are detailed in Section 3.3.

In Definition 3.1.1, it is stated that the PLs are the variations of a portfolio for given scenarios. These scenarios need to be generated for all different products present in a bank trading book. The general set up is that all assets are linked to one or more risk factors. A risk factor represents a given class/type of assets. A risk factor may be, for instance, "Equity-NorthAmerica-Financial-Institution". Then, the equity prices of North American banks are linked to this risk factor. These risk factors need to be simulated. Three main methods are used in the banking industry (see [41] and [44]):

- The first method is called historical PLs. This method employs historical variations to generate scenarios of the risk factors. A day (or some consecutive days for longer liquidity horizon) from the past gives the variation of the risk factors and, therefore, of the corresponding PL.
- The second method relies on a Monte Carlo simulation. The risk factors are simulated using many stochastic processes (geometric Brownian motions, for instance). Then, the entire portfolio is evaluated and this gives a PL. This operation is repeated to obtain a PL distribution.
- Finally, the last method is probably the least used. It makes use of stress tests as defined in the previous section. Many stress tests are performed and they provide the PL distribution.

These different options all have their advantages and drawbacks: Historical PLs are the easiest to use because the data is almost directly available and the correlation between the risk factors is the historical correlation, so no calibration is needed. However, the number of scenarios is limited (often less than a thousand) and they only reflect what happened in the past. The Monte Carlo method is the most advanced method. A lot of combinations are possible for the PLs and the number of scenarios that may be generated is potentially unlimited. However, calibration of the parameters and the correlations may be difficult. The cost of this method is also the biggest.

The use of stress tests is not very common. They employ a large number of scenarios based on the opinions of experts. The advantages are that any combination may be created and many scenarios may be generated. However, they often lead to more fundamental problems. The correlations between two stress scenarios are often very hard to calibrate. The maintenance costs of the model (to verify if the stressed test are still appropriated) are also quite high.

Due to the low cost and the robustness of the historical PLs method, this is used in the rest of the thesis. Stress tests may also be used because their results are directly available.

3.2 Historical Simulated PLS

We first investigate the historical PLS that are available and, then, investigate some properties they may have.

3.2.1 Overview of the PLS

Historical PLS are the first type of data analysed. These are simulated PLS computed using scenarios from the period 2008-2011 (4 years).

After having calibrated the risk factors and mapped the products from the trading book to them, the PLS are computed in different ways. Linear products are directly evaluated. For more complex products (ie. financial derivatives), two methods are used: full revaluation or the Delta-Gamma approximation. The first one employs the pricing functions whereas the latter uses an approximation of the product prices. The Delta-Gamma approximations use a quadratic approximation of the price of a derivative with the price of the underlying asset. The full revaluation is more accurate and captures the "real" behaviour of the financial products.

The PLS reflect 10-day variations of a certain portfolio. Therefore, if VaR or ES is computed on these PLS, then the result is a risk measure with a 10-day liquidity horizon and a 10-day capital horizon. However, these are 10-day overlapping PLS, it means that two successive PLS share 9 days. The PL of September 6th 2011 is overlapping with the PLS of September 7th 2011, this is represented in Figure 3.1:

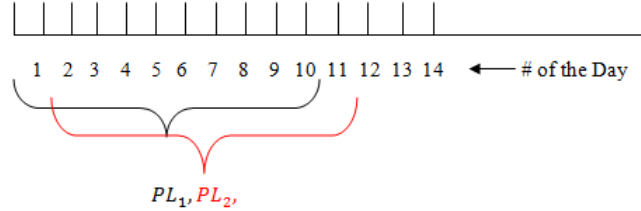


Figure 3.1: Overlapping PLS.

The PLS come in the form of a vector per book. The sum (over all the books so the trading book) of them can be plotted as in Figure 3.2 for example. In fact, it gives the 10-day PLS of the entire bank's trading book. In this figure, the stress period with a larger amplitude after September 2008 (credit crisis and Lehman bankruptcy) is clearly visible.

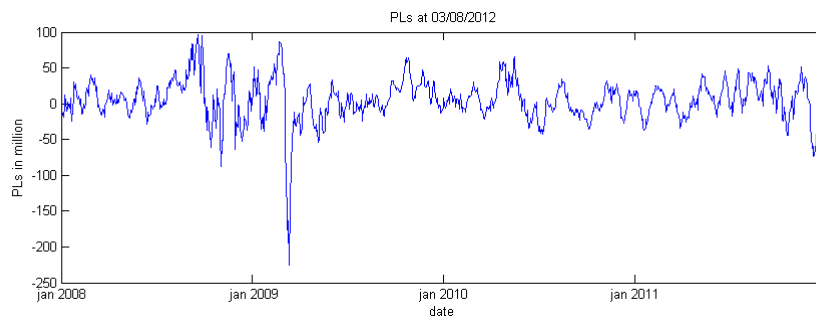


Figure 3.2: 4 years of 10-day PLS.

Remark 3.2.1

A last comment is about the use of 10-day PLs whereas it is easier to compute 1-day PLs. Unfortunately, information about the 1-day PLs does not imply that the corresponding 10-day PLs are known. The 10-day liquidity horizon is taken in order to capture more non-linearity (the pay-off of the product does not vary linearly with the underlying asset) of certain products. The two following quantities are different: The sum of the losses of two days, the 2-day PL.

Mathematically, it may be written as it follows: Let x be the price of an asset, Δx_1 and Δx_2 two daily variations of this price and R be the pricing function of a derivative. For non-linear products, the following holds:

$$(R(x + \Delta x_1) - R(x)) + (R(x + \Delta x_2) - R(x)) \neq R(x + \Delta x_1 + \Delta x_2) - R(x). \quad (3.1)$$

We analyse the example of a cash-or-nothing binary option. The current price of the underlying asset is 11. The option expires in two days with strike 10, so that the current price function is almost a step function.

The pay-off of this binary option is given by:

$$C(x) = \begin{cases} 1 & \text{if } x \geq 10, \\ 0 & \text{otherwise,} \end{cases} \quad (3.2)$$

where x is the price of the underlying asset.

Suppose that none of the historical 1-day PL of the underlying asset lead to a loss larger than 1. So, the 1-day PLs are all very small for this option. The 2-day PL may be very different. Assume that two consecutive 1-day PLs of the asset are -0.7. Then, the 2-day PL equals -1.4 and the 2-day PL of the option represents a large loss different from the sum of the two 1-day PLs. This is illustrated in Figure 3.3: none of the 1-day PLs changes (significantly) the price of the binary option but a 2-day PL does.

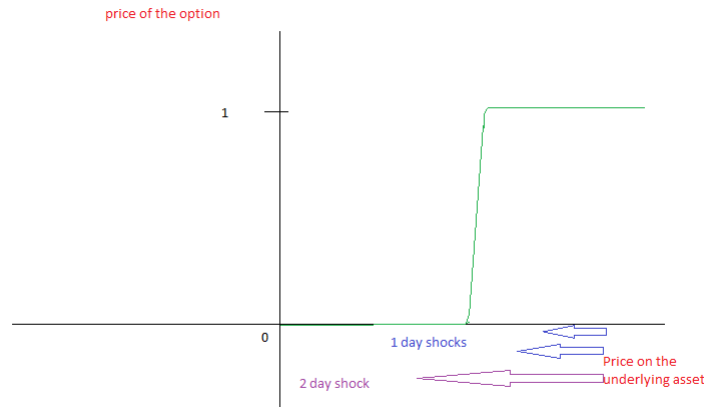


Figure 3.3: 1-day and 2-day PLs for a binary option.

3.2.2 Properties of the historical PLS

In finance, it usually holds that the PL distributions have fatter tails than normal distributions (see [35]). However, this is not certain because many different products and asset classes are

summed up. Figure 3.4 gives examples of tails of PLs distributions with a 10-day liquidity horizon computed with the variations of the window 2008-2011. This is done for three different portfolio snapshots. If a distribution (normal or other) can be assumed then, it would make the modelling quite easy because we could approximate the PL distribution and have the nice properties of the normal distribution, for instance.

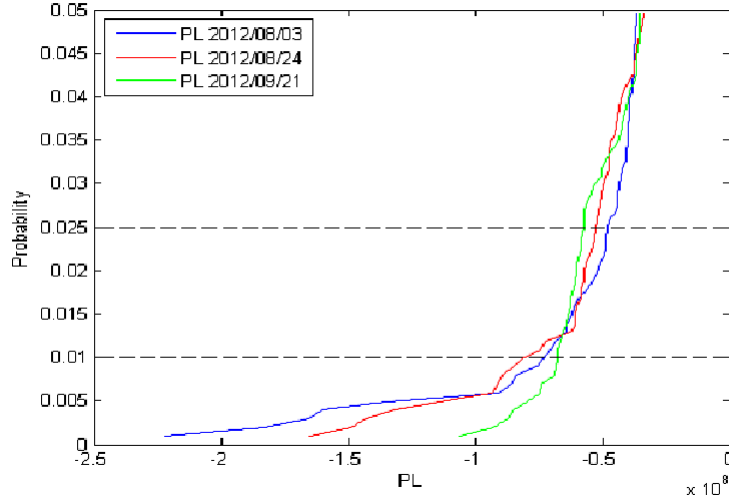


Figure 3.4: Tails for different PL distributions.

The normality of the 10-day PL distribution is investigated by means of graphical tests only. Two quantile-quantile plots of the summed PLs from all the trading books are plotted (dates are randomly selected) in Figure 3.5.

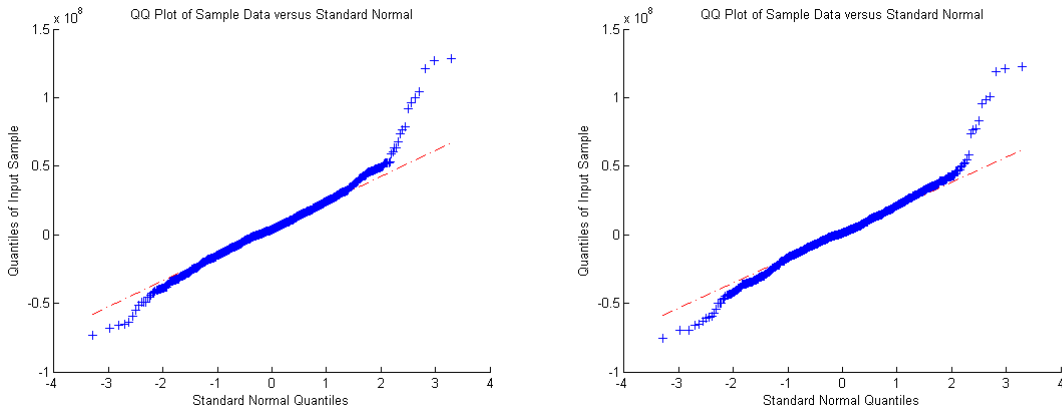


Figure 3.5: QQplot 18/01/2013 (left) and 25/01/2013 (right).

As expected the returns exhibit a fatter tail than a normal distribution. In our case, the hypothesis that the returns are normally distributed is rejected. These graphical tests are conclusive enough. We do not perform probabilistic tests.

As normality could not be proven, other graphical tests are performed.

Quantile-Quantile plots of the PLs with t-distributions are analysed (figure 3.6) for one date

only: 15/03/2013. The general conclusion is that a t -distribution with 5 or 6 degrees of freedom fits better the lower tail than the other distributions. Another remark is that the two tails are different. The profit tail is fatter than the loss tail. An asymmetric probability density is more suitable to model the whole distribution. However, the objective is not to model the PL distribution perfectly but to have realistic and manageable approximation.

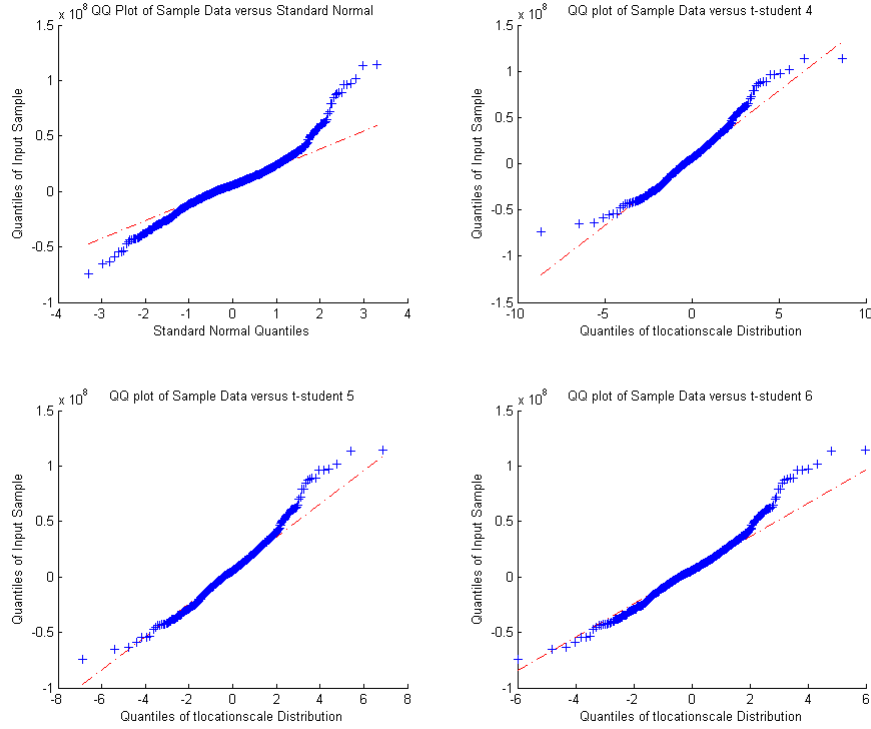


Figure 3.6: QQplots of a 10-day PL distribution comparing different distributions.

3.2.3 Autocorrelation of the historical PLs

Autocorrelation among the data is an important factor in financial time series. It makes cumulative losses become more severe because a loss is more likely to be followed by another loss. Therefore, the Economic Capital is likely to increase if the autocorrelation increases. Over several portfolio snapshots, it can be seen that the autocorrelation is positive. Figure 3.7 shows the 10-day autocorrelation of a granular (in fact, the entire) portfolio containing a wide range of products and asset classes. As stated in Section 3.2 the 10-day PLs are overlapping, but this is the autocorrelation without overlap. To compute it, ten time series have been extracted so that there are no more overlapping dates within a time series. The autocorrelation is computed individually for these time series and is averaged.

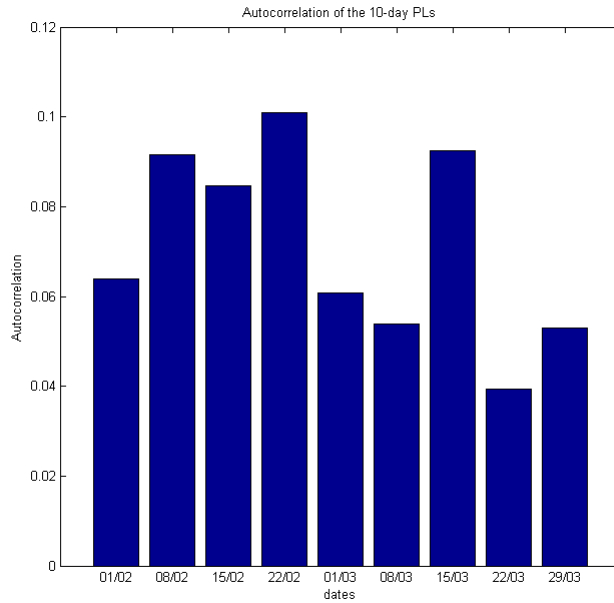


Figure 3.7: Autocorrelation of the 10-day PLs.

However, this autocorrelation should not be underestimated. During stressed market conditions, the 10-day autocorrelation tends to increase significantly (see [8]).

3.3 Stressed PLS

3.3.1 Overview

At most banks, the data also includes stress tests results. Indeed, it is a regulatory obligation to perform stress tests and to report the results to the regulators (see [40]). Therefore, banks may also use them to compute their Economic Capital or for other internal purposes. Stress tests may be based on scenarios calibrated on historical data and verified by experts or purely hypothetically (Euro crisis, for instance). They typically impact one asset class (equity, commodity, interest rate,...). An example of a stress test is to assume that equity products go up by 10% over 10 days. For our purposes, 59 tests are available and they are summarized in Table 3.1. They are stress tests calibrated on historical extreme events and validated by experts. This table gives the asset classes that are impacted by the stress tests. The second column gives the number of scenarios that are available. Several parameters are changed in stress tests. For instance, equity stress tests are scenarios with low and high returns, but the volatility is also changed. One of the scenarios is equity prices do not change but the volatility increases by 40%. This gives many possible combinations.

Table 3.1: 10-day stress tests.

Risk Factor	# of scenarios	Min (€)	Max (€)
Commodities	8 scenarios	-2.210.623	2.128.732
Credits	2 scenarios	-207.603.911	245.211.165
Equities	19 scenarios	-23.441.714	10.103.217
Bonds(6 months)	2 scenarios	-418.936.851	332.719.486
Inflations	2 scenarios	14.542	20.821
Interest rates	12 scenarios	-30.539.552	44.927.168
FX	12 scenarios	-865.615	1.448.775
Treasuries	2 scenarios	-16.598.341	16.941.515

In Section 3.3.2, the properties and the relevance of the stress tests for capital models are analysed.

3.3.2 Relevance of stress tests for Economic Capital

Qualitatively, there are good reasons to argue that stress tests should not be used to model the Economic Capital:

Banks calibrate some of their stress tests on historical data to check whether they are extreme variations or not. Therefore, most of the information contained in stress tests is also available in the historical PLs. A second issue is that a given stress test does not affect the whole portfolio. Therefore, if stress tests are used, the results need to be aggregated together. This is rather complex because correlations between stress tests are difficult to estimate or expert based so artificial because dependent on somebody's expertise rather than mathematics. A method to have a consistent aggregation may be found in [32]. A model purely based on stress tests is possible but requires more stress test data than available for our purposes. A hybrid model (historical PLs + stress tests) poses some challenges when trying make to a consistent framework. A last argument is that, allocation of the risks can be hard to justify and expert based.

To conclude, stress tests are not used in the rest of the thesis.

Chapter 4

Modelling the Economic Capital

In Chapter 2, the Economic Capital has been defined from a high level perception. In this chapter, modelling choices are explained. The general model for the Economic Capital is given and some analysis is performed to assess some common practice in risk management such as the scaling of risk measures to a different time horizon.

From Chapter 3, we learnt that we will use historical PLs. However, there are only a small number of them available.

A bank aiming for the best rating possible, the amount of capital is supposed to cover the worst 99.99% the possible 1-year scenarios (0.01 % is the 1-year probability of default of a AAA rated company according to S&P). Based on definitions from Section 2.1, the Economic Capital should be equal to the 1-year VaR with a threshold level of 99.99%. However, there is an issue: in Section 3.2, data were discussed and only 10-day PLs were available, not the 1-year PL distribution. Therefore, taking the VaR directly on the PLs is not a viable option. Furthermore, it is a requirement to use 10-day PLs to reflect a 10-day liquidity horizon. In order to model this, the market fluctuation risk is introduced. In a second part, other types of risk are presented: The migration and default risks. They are not fully captured by the market fluctuation risk so another method is needed to model them.

Many banks add other terms to their Economic Capital. However, they are often specific to the bank and do not count for a large proportion of the Economic Capital.

The Economic Capital for the trading book (EC) is given by the following formula:

$$EC = L \cdot \text{VaR}_{99.99\%}(X) + IRC + SR, \quad (4.1)$$

where X is the 1-year PLs distribution, IRC is the Incremental Risk Charge and SR is the Specific Risk of the bank. $\text{VaR}_{99.99\%}(X)$ is the market fluctuation risk (also called VaR component in the literature, however this name is non-sense, when considering a change to expected shortfall). L is a given parameter depending on the liquidity of the trading book ($0 \leq L \leq 1$, currently around 0.75). $\text{VaR}_{99.99\%}(X)$ is discussed in Section 4.1 and the IRC is discussed in Section 4.2. L and SR are not in scope.

4.1 Market Fluctuation Model

The market fluctuation risk of the Economic Capital is the main, and historically the first, component of the Economic Capital. In this section, the methodology used to compute the

$\text{VaR}_{99.99\%}(X)$ is explained. However, due to business requirements, for simplicity and for the sake of risk management, the Economic Capital is computed by an approximation.

4.1.1 Definitions and issues

First, new useful concepts and notations to model the Economic Capital are defined. We stated in the introduction of this chapter that there was an issue when the liquidity horizon of the PLs is not the same as the desired capital horizon. Therefore, there is a need to define a risk measure computed on a capital horizon longer than the liquidity horizon.

Definition 4.1.1 (n-period risk measure)

Let $\rho : V \rightarrow \mathbb{R} \cup \{+\infty\}$ be a risk measure on (Ω, \mathcal{F}, P) as defined in Definition 2.1.1. Let $X = (X_1, X_2, \dots, X_n) \in V^n$ be n identically distributed real random variables. Let Δt be the liquidity horizon of X_i for all $i \in \{1, \dots, n\}$ and $n \cdot \Delta t = T$ be the capital horizon. Then,

$$\rho'(X, T) = \rho'(X, n \cdot \Delta t) = \rho \left(\sum_{i=1}^n X_i \right), \quad (4.2)$$

is an n -period risk measure.

Remark 4.1.1 Dependency between the random variables is assumed. Very often in the thesis the two notations are merged: with one parameter $\rho(X)$ it is a Δt capital horizon and with two parameters $\rho(X, T)$ is a T capital horizon.

For Economic Capital modelling, 10-day (10d) PLs are available and the objective is to model a 1-year (1y) Economic Capital. Assuming that there are 250 trading days in a year, we have:

$$\rho(X, 1y) = \rho \left(\sum_{i=1}^{25} X_i \right) = \rho \left(\sum_{i=1}^{25} X_i, 10d \right), \quad (4.3)$$

where $(X_i)_{i=1,2,\dots,25}$ are identically distributed as the 10-day PLs $\forall i \in \{1, 2, \dots, 25\}$.

4.1.2 Approximations for Value at Risk and Expected Shortfall

The distribution of $\sum_{i=1}^{25} X_i$ is unknown in general or is expensive to compute. Therefore, an approach to model the risk of the sum of identical random variables based on the risk of one of these random variables has to be developed. This particular strategy has been used for a long time in risk management (see [12]). In this section, a basic approximation of the 10-day VaR, given the 1-day VaR is derived. Scaling from 10 days to one year cannot be done as there is not enough data to have a 1-year PL distribution. In case the time series is not autocorrelated and with some regularity conditions (discussed after Property 4.1.1) the following approximation holds for VaR:

Property 4.1.1

To compute VaR for longer or shorter liquidity horizon we use the following approximation:

$$\text{VaR}_\alpha(X, n\Delta t) \approx \sqrt{n} \cdot \text{VaR}_\alpha(X, \Delta t), \quad (4.4)$$

where $X = \{X_1, \dots, X_n\}$ are identically distributed, Δt is the liquidity horizon of X_i , for all $i \in \{1, \dots, n\}$, and $n \cdot \Delta t = T$ is the capital horizon. This approximation is exact in case i.i.d. normally distributed returns.

Proof:

Suppose X_i are normally distributed for all $i \in \{1, \dots, n\}$. Then, the quantile function is given by (see [3]):

$$F_{X_i}^{-1}(p) = \mu + \sigma\sqrt{2}\text{erf}^{-1}(2p - 1),$$

where μ is the mean, σ is the standard deviation and erf is the error function. Assume the mean is 0, then the VaR depends only on the standard deviation. The standard deviation of a sum of n i.i.d. normal random variables is again normally distributed with standard deviation \sqrt{n} . Therefore, the approximation is valid. □

This approximation was recommended in Basel II [38]. However, the literature [12] has shown that it should be used with caution because it may underestimate or overestimate (depending on division or multiplication) the risk. It should be used for short periods of time. A scaling from one day to one year is not appropriate as shown in [48]. Counter examples (where this approximation fails) may be found in [12] and in Figure 4.1.

Example 4.1.1

Property 4.1.1 is investigated based on one year of PLs (250 PLs). Various quantiles are determined for the 10-day liquidity horizon based on the realized 10-day PLs and based on the square root rule with 1-day realized PLs. The following formula is used to approximate the 10-day VaR:

$$\text{VaR}_\alpha(X, 10d) \approx \sqrt{10} \times \text{VaR}_\alpha(X, 1d). \quad (4.5)$$

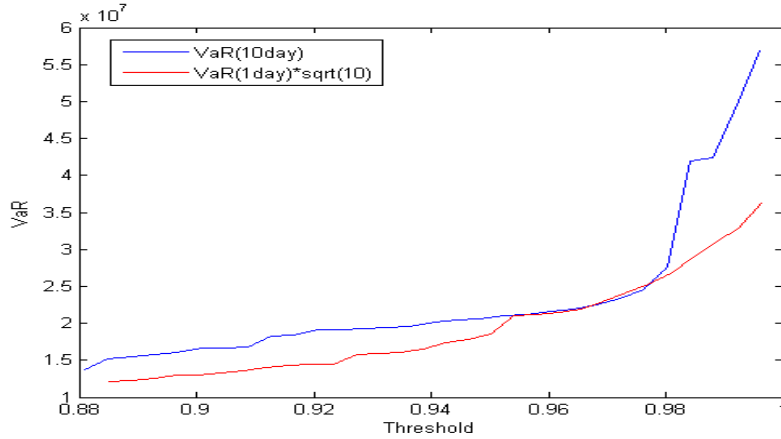


Figure 4.1: Validity of the square root rule.

Extreme quantiles (99% and more) exhibit a large difference between the exact quantile and the approximation. There are two reasons for that. The first one is the non-linearity of the PLs (due to financial derivatives) and the second is that we use overlapping 10-day PLs. Therefore, a 1-day large loss may influence the distribution for up to 10 days.

In the literature [45], some improvements to this method have been proposed. A basic improvement consists in keeping the same items as the previous approximation but use an exponent different than 0.5 (i.e. square root):

$$\text{VaR}_\alpha(X, n\Delta t) \approx n^\theta \cdot \text{VaR}_\alpha(X, \Delta t), \quad (4.6)$$

where $\theta \in \mathbb{R}$. In practice, parameter θ is likely to be close to 0.5, as shown in Figure 4.2 on empirical PLs ($T=10$ -day, $\Delta t = 1$ -day). In this figure, the average of θ over 15 dates is used. The parameter θ increases as the threshold decreases up to certain level (98%). This is a consequence of overlapping data: 10-day PLs may be affected by one very large 1-day loss. Then up to ten 10-day PLs may be large because influence by this 1-day PL. Therefore several 10-day PLs may be very close because of a single 1-day PL. However, the 1-day PLs are not overlapping. So for several extreme thresholds, the 10-day VaR corresponding may be closed and the ratio $\frac{\text{VaR}_\alpha(X, n\Delta t)}{\text{VaR}_\alpha(X, \Delta t)}$ decreases as α increases because $\text{VaR}_\alpha(X, \Delta t)$ increases faster than $\text{VaR}_\alpha(X, n\Delta t)$.

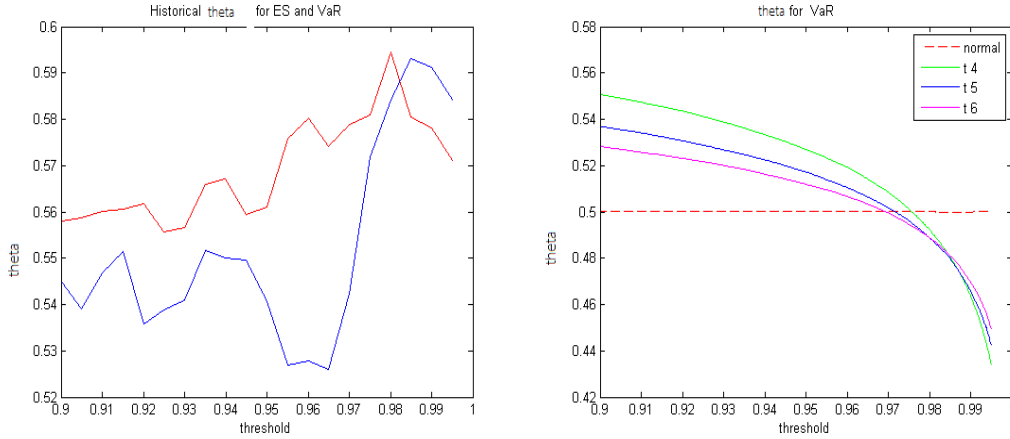


Figure 4.2: θ for historical PLs and for classic distributions.

The square root Approximation 4.1.1 of the VaR should also holds for ES. It comes from the very definition of ES: $\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_t(X) dt$.

Therefore, it should also hold that:

$$\text{ES}_\alpha(X, n\Delta t) \approx \sqrt{n} \cdot \text{ES}_\alpha(X, \Delta t).$$

As for the VaR this rule should be used with caution for ES. The same verification is performed as for VaR: The ES of the 10-day PLs and $\sqrt{10} \times \text{ES}_\alpha(X, 1d)$ are provided in Figure 4.3.

As expected, due to the fat tail of the PLs this approximation is underestimating the risk. The multiple underestimations of the true VaR, made at most of the thresholds, are adding up for ES.

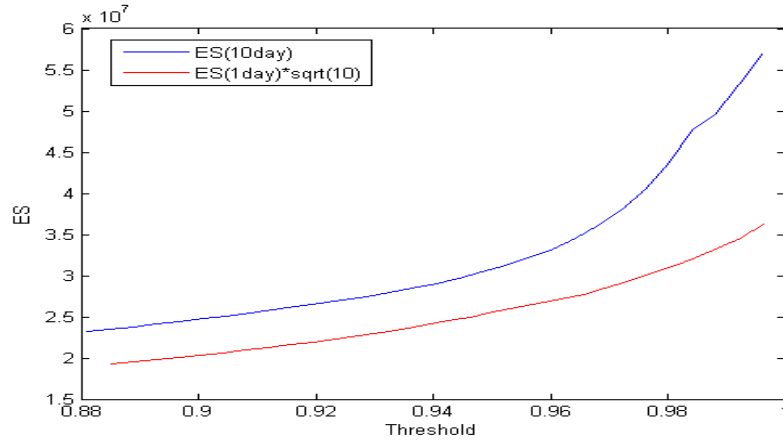


Figure 4.3: Square Root rule for expected shortfall.

The conclusion is that these approximations should be used with caution. Furthermore, they are too restrictive. To model the Economic Capital, an approximation of $\text{VaR}_{99.99\%}(X, T)$ is required. It is not possible to use the square root scaling because taking the 99.99% VaR on 1000 10-day PLs is not possible. Furthermore, there is no independence between the 10-day PLs. However, these approximations suggest that a Scaling Factor for a risk measure ρ (SF_ρ) may be found such that:

$$\text{VaR}_{99.99\%}(X, n\Delta t) \approx SF_\rho \cdot \rho(X, \Delta t), \quad (4.7)$$

If the scaling factor is calibrated so that the following holds:

$$SF_\rho := \frac{\text{VaR}_{99.99\%}(X, T)}{\rho(X, \Delta t)}, \quad (4.8)$$

then this may be a suitable scaling (in practice Δt is 10 days and T is one year). However, some analysis needs to be done to determine whether equation (4.8) gives an accurate scaling. Furthermore, exact computation of $\text{VaR}_{99.99\%}(X, T)$ is still needed. Subsequently, a widely used sampling algorithm may be used.

4.1.3 Sampling algorithm

Here we explain the method to compute an accurate $\text{VaR}_{99.99\%}(X, T)$. The general concept is based on a sampling method where 25 10-day PLs are randomly drawn (with Gaussian correlation copula with correlation conservatively fixed at 20%) out of the 1000 historical PLs. They are summed up and this operation is repeated a large number of times. Once there is a sufficient number of scenarios, the 99.99% quantile is taken.

The main mathematical insight is that the quantile of a sum of 25 Gaussian correlated random variables is taken, is accordance with Equation (4.3).

Let F be the cumulative distribution function of the PL distribution. Then, the sampling Algorithm 1 is employed (also summarized in Figure 4.4):

```

for  $i = 1$  to # iterations do
  rand = Random Number in  $[0,1]$  ;
   $D(1) = F^{-1}(\text{rand})$  ;
  for period = 2 : 25 do
     $Z =$  random number from a  $N(0,1)$  distribution ;
    rand =  $c \cdot \text{rand} + \sqrt{1 - c^2} \cdot Z$  ;
     $D(\text{period}) = F^{-1}(\text{rand})$  ;
  end
end

```

Algorithm 1: Sampling procedure.

c is the non-overlapping autocorrelation of the 10-day PLs.¹

D contains the 1-year PL distribution, so VaR or ES can be computed directly from it.

The algorithm may also be represented as follows:

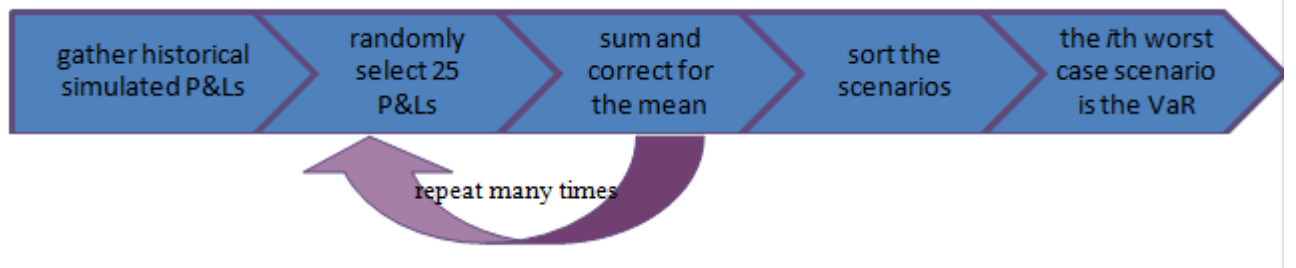


Figure 4.4: Sampling methodology.

4.2 Incremental Risk Charge

4.2.1 Definition and motivations

IRC stands for Incremental Risk Charge. It represents the migration risk and the default risk which is not fully modelled by the market fluctuation risk.

The IRC has been incorporated to the Regulatory Capital [39] in response to the large credit exposure of banks in their trading books to certain issuers. In their ICAAP, banks (major global banks) are also required to show that their Economic Capital model captures the IRC [41]. Without IRC (so only with the market fluctuation risk), owning bonds from, for example, UBS is not very risky because in the past fluctuations of the bond price were not very large. However, bankruptcy of UBS may be possible and it would lead to a large loss. So, the default of an issuer is not fully modelled by the market fluctuation risk of the Economic Capital for trading book. Indeed, if an issuer goes bankrupt, then, most of the time, there is dissolution, so the bonds will not be available anymore. However, it is not because an issuer never went bankrupt before, that it will never happen in the future. Therefore, there is a need of a new

¹This is not Pearson correlation but Gaussian copula correlation. It can be shown that 20% is still very conservative for the 10-day PLs.

model to measure this risk. The same holds for the migration risk. IRC has been introduced in Basel II [43].

Definition 4.2.1 (IRC, [39]) :

IRC represents the potential loss due to the risk of default and the migration risk.

The default is bankruptcy: The issuer cannot fulfil its financial obligations and therefore, will not reimburse the debts that has been issued.

The migration is a change in the rating of companies, which has been defined in Section 2.2.1. The consequence of a downgrade is a decrease in the value of bonds.

In this thesis the following rating system is assumed: R_1, R_2, \dots, R_{20} . R_1 is the best possible rating and R_{20} the worst rating. To these 20 ratings the default rating D and the risk free rating R_0 are added. More insight about these ratings is given in Chapter 7.

4.2.2 Overview and inputs of the IRC model

In this section, we give a high level description of the methodology employed for IRC. In this thesis, we are particularly interested in the way the inputs of IRC are computed, which is detailed in Chapter 7. The general idea is a Monte Carlo simulation where scenarios are simulated. In these scenarios, an issuer may or may not go bankrupt or may migrate to a better or worse rating. This is done for each issuer. In a scenario, the default of an issuer leads to a loss. A migration of a counter-party leads to a loss, due to a decrease in the value of their bonds or to an increase in case of a positive migration. The mathematical model behind is the use of a Merton credit model (see [21] for detail about this model).

This Monte Carlo method relies on three main inputs:

- The current bond and credit derivative positions of the bank.
- Correlations between the asset values of the issuers.
- A 3-month migration matrix and the 3-month probabilities of default. A migration matrix is a 2-dimensional matrix (20 by 20 in our case) containing the probability that R_i migrates to R_j over a 3-month period. The 3-month Probabilities of Default vector (PDs) contains the probability that an entity with rating R_i goes into default over a 3-month period.

The PDs and the migration have to be estimated in a consistent way, following requirements detailed in Chapter 7. With a large number of scenarios generated, the output is a loss distribution from which the quantile (99.99% for the Economic Capital and 99.9% for the Regulatory Capital) is taken as shown in Figure 4.5. This figure provides the output of the IRC. After running the Monte Carlo simulation, the result is a distribution that represents possible scenarios for the IRC. The VaR needs to be taken with a threshold of 99.9% for the Regulatory Capital and 99.99% for the Economic Capital.

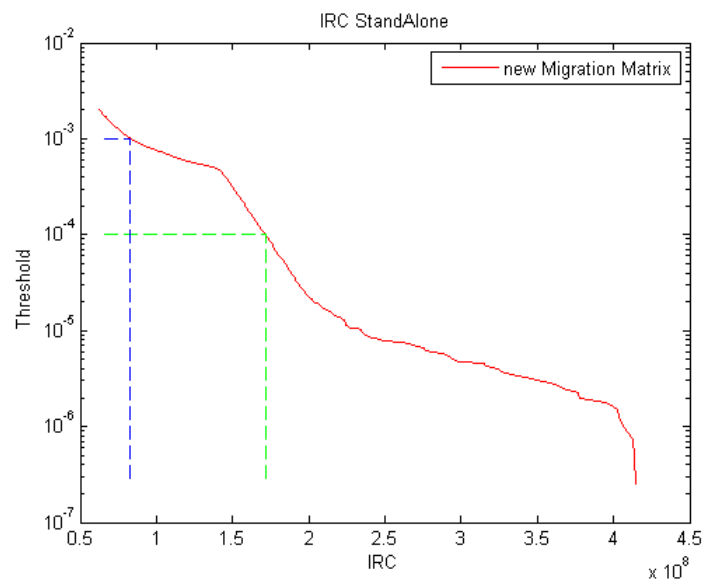


Figure 4.5: Tail IRC distribution.

Chapter 5

Estimation of Value at Risk and Expected Shortfall

This chapter focuses on the estimation of VaR and ES. Data is usually restricted, so discrete estimators are needed. The objective of this chapter is to quantify the error that is made because of discretization, the lack of data and the autocorrelation (overlap) of the PLs. It is important to know whether the results given by the risk measures are accurate or not. For this purpose, three quantifiers of the error are introduced in Section 5.1. Then, in Sections 5.2 and 5.3, classical non-parametric estimators for VaR and ES are introduced. We give a number of properties and relevant theorems. They are concerned with the asymptotic normality of the estimators and their variance.

In Sections 5.4 and 5.5, numerical experiments are performed on non-autocorrelated time series first and, subsequently, on autocorrelated time series because, as mentioned in Chapter 3, data is overlapping and therefore, autocorrelated. Numerical experiments focus on VaR and ES estimators for small samples.

The analysis is first performed on well known distributions.

First of all, the reader is invited to read Appendix B for some background information on statistics. It provides some definitions (stationary time series, α -mixing...) and theorems such as Ibragimov's theorem (Theorem B.1.1) needed to understand this chapter. We define criteria that are used in this chapter to assess the quality of the estimators to compute the scaling factor.

5.1 Definition of Errors

The following defines bias, standard deviation and Mean Square Error (MSE) of an estimator. These are classical error to assess the VaR and ES estimators.

The bias is the mean error to true value. Let X be a PL distribution, ρ be a risk measure and $\hat{\rho}$ be an estimator of this risk measure. Then, the bias is defined as follows:

$$\text{Bias}(\hat{\rho}(X)) = \mathbb{E}[\hat{\rho}(X)] - \rho(X). \quad (5.1)$$

It is often expressed relative to its value: $\text{Bias}(\hat{\rho}(X)) = \frac{\mathbb{E}[\hat{\rho}(X)] - \rho(X)}{\rho(X)}$. We use the relative version of the bias.

The interpretation generally done is that the bias represents the inaccuracy of the estimator for a given sample size. This gives the error made on average by using a limited sample and a given estimator. From its value it is possible to know whether, on average, an estimator of a risk measure is underestimating or overestimating the risk.

The standard deviation of an estimator is expressed relative to the value of the risk measure as well:

$$SD(\hat{\rho}(X)) = \frac{\sqrt{\mathbb{E}[(\hat{\rho}(X) - \mathbb{E}[\hat{\rho}(X)])^2]}}{\mathbb{E}[\hat{\rho}(X)]}. \quad (5.2)$$

The standard deviation is the error in one simulation. On average, it is the error to the biased mean value of the estimator.

The MSE is the mean square error given by the following formula:

$$MSE(\hat{\rho}(X)) = \mathbb{E}[(\hat{\rho}(X) - \rho(X))^2]. \quad (5.3)$$

This is the square error to the true risk measure. It is often expressed in relative terms and the square root is taken:

$$MSE(\hat{\rho}(X)) = \frac{\sqrt{\mathbb{E}[(\hat{\rho}(X) - \rho(X))^2]}}{\rho(X)}. \quad (5.4)$$

It is well known that the following holds for the MSE from Equation 5.3:

$$MSE(\hat{\rho}(X)) = \text{var}(\hat{\rho}(X)) + \text{Bias}(\hat{\rho}(X))^2. \quad (5.5)$$

Therefore, the MSE contains the two previous errors: the bias and the standard deviation. Some more information and properties may be found in [34].

In the following, we only use relative errors because we wish to compare risk measures with different values that, in the end, are scaled to measure the same quantity: 1-year VaR.

5.2 Non-Parametric Estimation of Value at Risk

First, two estimators are defined for VaR. These are commonly used estimators because they are the "natural" estimators of the quantiles. They are easy to use as they consider only one point of a sample.

5.2.1 Discrete estimators

In Definition 2.1.4, VaR has been defined for a general random variable where the distribution is known. However, historical simulated PLs consist of only a small number of observations. They contain a maximum of 1000 PLs. Therefore, discrete estimators of VaR need to be defined.

Definition 5.2.1 (VaR Estimators, [22])

Let $X = \{X_1, X_2, \dots, X_n\}$ be a sample (observed time series) and consider the order statistic $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, where $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$. The following estimators are defined:

$$\widehat{\text{VaR}}_{\alpha}^L(X) = -X_{[n \cdot (1-\alpha)]:n}, \quad (5.6)$$

$$\widehat{\text{VaR}}_\alpha^U(X) = -X_{[n \cdot (1-\alpha)] : n}, \quad (5.7)$$

$\lfloor \cdot \rfloor$ is the rounding to the lower integer and $\lceil \cdot \rceil$ is the rounding to the upper integer. If $(1-\alpha) \cdot n$ is an integer, the two estimators are equal. The following relation holds:

$$\widehat{\text{VaR}}_\alpha^U(X) \leq \widehat{\text{VaR}}_\alpha^L(X).$$

The notations are not intuitive: the lower estimator $\widehat{\text{VaR}}_\alpha^L$ is greater than the upper estimator $\widehat{\text{VaR}}_\alpha^U$. The $\widehat{\text{VaR}}_\alpha^U$ estimator is the upper estimator of the VaR in the sense that the threshold is rounded to the upper integer. The threshold of $\widehat{\text{VaR}}_\alpha^L$ is floored.

This chapter is quite theoretical so the VaR is not corrected by the mean in order to keep simple notations.¹ The analysis is the same with the mean correction and the results do not change.

An alternative is to consider the following estimator:

$$\widehat{\text{VaR}}_\alpha^M(X) = (\lceil n(1-\alpha) \rceil - n(1-\alpha)) \cdot \widehat{\text{VaR}}_\alpha^U(X) + (n(1-\alpha) - \lfloor n(1-\alpha) \rfloor) \cdot \widehat{\text{VaR}}_\alpha^L(X). \quad (5.8)$$

This is the estimator used by the Matlab function `quantile` for instance. Instead of a rounding, it uses a linear interpolation if $n(1-\alpha)$ is not an integer between the two VaR estimators previously defined. This estimator has nicer properties in case $n(1-\alpha) \notin \mathbb{N}$ because there is not "jump" in the threshold (for 100 PLs the 99.1% VaR is given by the 1 or the 2 worst observation with Definition 5.2.1). This estimator is not investigated in the numerical experiments. Its behaviour is linear between the points where $n(1-\alpha) \in \mathbb{N}$. In fact, all the error estimators of $\widehat{\text{VaR}}_\alpha^M$ are bounded by the errors of the estimators $\widehat{\text{VaR}}_\alpha^U$ and $\widehat{\text{VaR}}_\alpha^L$.

5.2.2 Properties of the Value at Risk estimators

Theoretical properties of the estimators from Definition 5.2.1 are derived. The properties are only valid if $n \cdot (1-\alpha) \in \mathbb{N}$ so that all the estimators previously defined are the same. In a first part, the bias of the estimator is investigated, we show that the estimators are biased. Then, a central limit theorem and the asymptotic variance are derived analytically.

The estimators of VaR are biased. The main problem is that for general distributions the bias is unknown. To show that the estimator is biased, a classical distribution is assumed and the bias is derived.

To illustrate the bias, we consider the random variable $X \sim U(0,1)$. It is obvious that $\text{VaR}_\alpha(X) = \alpha - 1$. Note that, because X is positive, the VaR is always negative. Suppose that a sample of n independent observations is drawn and that $i = (1-\alpha) \cdot n$ is an integer. Then, using Equation (B.11) the following holds $\forall x \in [0,1]$:

$$P(X_{i:n} \leq x) = \sum_{k=i}^n C_n^k F_X(x)^k (1 - F_X(x))^{n-k},$$

i.e.,

$$P(X_{i:n} \leq x) = \sum_{k=i}^n C_n^k x^k (1-x)^{n-k},$$

¹In Chapter 2.1, we mentioned that in risk management the mean should be added but this chapter is about mathematical properties of the risk measures.

where C_n^k is the k -combination in a set of n elements. The density can be deduced,

$$f_{X_{i:n}}(x) = i \cdot C_n^i x^{i-1} (1-x)^{n-i}.$$

Then, the expectation is taken on both sides and successive integration-by-parts gives:

$$\begin{aligned} \mathbb{E}[X_{i:n}] &= \int_0^1 x \cdot f_{X_{i:n}}(x) dx = i \cdot C_n^i \cdot \int_0^1 x \cdot x^{i-1} (1-x)^{n-i} dx = i \cdot \frac{n!}{i!(n-i)!} \frac{i!(n-i)!}{(n+1)!}, \\ \mathbb{E}[X_{i:n}] &= \frac{i}{n+1}. \end{aligned}$$

The two VaR estimators are equivalent and are biased with bias $\frac{\alpha-1}{n+1}$. Relatively to the VaR, of the true VaR the bias is constant and equals $\frac{1}{n+1}$. The bias reduces linearly with the sample size.

For usual financial distributions such as normal or Student's t -distributions, we need to investigate the bias, because it is biased for uniform random variables, it is probably the case also for other distributions.

The next property provides a central limit theorem for quantile estimators (so VaR estimators) for autocorrelated time series. First, two conditions are explained. They are technical and refer to concepts explained in Appendix B. Given a time series $X = \{X_1, \dots, X_n\}$, the following two conditions are defined:

Condition 1:

The process $\{X_i\}_{i \in \{1 \dots n\}}$ is strictly stationary, α -mixing and there exists a $\xi \in (0, 1)$ such that $\alpha(k) \leq C\xi^k$, for all k , where $\alpha(k)$ are the α -mixing coefficients (see Appendix B.1.3) and C is a positive constant. Furthermore, X_1 should be continuously distributed with f and F as density and cumulative distribution function, respectively.

This condition states that the distribution is sufficiently smooth and the α -mixing coefficients (a type of autocorrelation) of the time series reduces to 0 with the lag, at (at least) an exponential rate. This condition is stronger than just α -mixing which just states that the coefficients tend to 0 as $k \rightarrow \infty$.

Condition 2:

f has continuous second derivatives and the distribution of (X_1, X_k) has continuous second order partial derivatives and they are uniformly bounded with respect to k .²

These two conditions are rather technical. They ensure existence of the variance and that the autocorrelation is not too strong.

Theorem 5.2.1 (Central limit theorem for Value at Risk, [10])

Under Conditions 1 and 2 we have the following:

$$\sqrt{n}(\widehat{\text{VaR}}_\alpha(X) - \text{VaR}_\alpha(X)) \longrightarrow N(0, \nu^2), \quad (5.9)$$

where,

$$\nu^2 = \frac{1}{f^2(-\text{VaR}_\alpha(X))} \left((1-\alpha) \cdot \alpha + 2 \sum_{h=1}^{\infty} \text{cov}(\mathbb{I}_{X_1 \leq -\text{VaR}_\alpha(X)}, \mathbb{I}_{X_{1+h} \leq -\text{VaR}_\alpha(X)}) \right). \quad (5.10)$$

²Because the time series is strictly stationary the distributions of (X_1, X_k) and (X_{i+1}, X_{i+k}) are the same.

This theorem is the mathematical explanation of the difficulty of estimating the extreme quantiles with a limited number of correlated observations. As α tends to 1 or 0 if the term $f^2(-\text{VaR}_\alpha(X))$ tends to 0 faster than $(1 - \alpha) \cdot \alpha$, then the variance tends to infinity. We do not prove the theorem, however, it is explained and linked to more classical theorems.

This theorem is similar to Theorem B.1.2, which states the same for an empirical cumulative distribution function. The estimation of a VaR and a point of the cumulative distribution function behave in the same way. This is not surprising because the VaR is a point of the distribution.

However, it cannot be proven from it. Theorem 5.2.1 originates from a stronger result: Yoshihara proved in [55] that, under the conditions of Theorem 5.2.1, we have:

$$\widehat{\text{VaR}}_\alpha(X) - \text{VaR}_\alpha(X) = \frac{\hat{F}_n(\text{VaR}_\alpha(X)) - (1 - \alpha)}{f(\text{VaR}_\alpha(X))} + O(n^{-\frac{3}{4}} \log n), \quad (5.11)$$

where \hat{F}_n is the empirical distribution function, as defined in Appendix B.1.1, and f is the probability density function of random variable X .

Equation (5.11) is called a Bahadur representation (see [55]). From it, the variances of VaR estimators may be derived:

Property 5.2.1 *For the trivial VaR estimators and under the conditions of Theorem 5.2.1, the variance of the estimators is the following:*

$$\begin{aligned} \text{var}[\widehat{\text{VaR}}_\alpha(X)] &= \frac{1}{n \cdot f^2(-\text{VaR}_\alpha(X))} ((1 - \alpha)(\alpha) \\ &\quad + 2 \sum_{h=1}^{n-1} (1 - \frac{h}{n}) \text{cov}(\mathbb{I}_{X_1 \leq -\text{VaR}_\alpha(X)}, \mathbb{I}_{X_{1+h} \leq -\text{VaR}_\alpha(X)})), \end{aligned} \quad (5.12)$$

Proof:

Using Bahadur representation (5.11) and taking the variance at both sides gives:

$$\text{var}(\widehat{\text{VaR}}_\alpha(X) - \text{VaR}_\alpha(X)) = \text{var} \left(\frac{\hat{F}_n(-\text{VaR}_\alpha(X)) - (1 - \alpha)}{f(-\text{VaR}_\alpha(X))} + O(n^{-\frac{3}{4}} \log n) \right).$$

Then all constant terms are withdrawn and the terms $O(n^{-\frac{3}{4}} \log n)$ are neglected, i.e.

$$\text{var}(\widehat{\text{VaR}}_\alpha(X)) = \text{var} \left(\frac{\hat{F}_n(\text{VaR}_\alpha(X))}{f(-\text{VaR}_\alpha(X))} \right) + O(1),$$

$$\text{var}(\widehat{\text{VaR}}_\alpha(X)) = \frac{1}{f^2(-\text{VaR}_\alpha(X))} \cdot \text{var}(\hat{F}_n(-\text{VaR}_\alpha(X))) + O(1).$$

The only term to compute is $\text{var}(\hat{F}_n(-\text{VaR}_\alpha(X))) = \text{var}(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq -\text{VaR}_\alpha(X)})$. Such a quantity has already been computed, in Theorem B.1.1, with general random variables. We obtain:

$$\text{var}(\widehat{\text{VaR}}_\alpha(X)) = \frac{1}{n \cdot f^2(-\text{VaR}_\alpha(X))} \left((1 - \alpha)(\alpha) + 2 \sum_{h=1}^{n-1} (1 - \frac{h}{n}) \text{cov}(\mathbb{I}_{X_1 \leq -\text{VaR}_\alpha(X)}, \mathbb{I}_{X_{1+h} \leq -\text{VaR}_\alpha(X)}) \right),$$

where f is the probability density function of X . This proves Property 5.2.1. The asymptotic variance can be computed by taking the limit and using the dominated convergence theorem with the bound of Condition 1 to ensure existence.

□

The variance of Property 5.2.1 is used in the numerical implementation to compute the variance of the simulated PLs. For these PLs, Monte Carlo cannot be used to compute the variance as it is the case when drawing observations from classical distributions.

This theorem only provides a theoretical framework, because the distribution of the time series in the variance and the density are unknown for historical time series. Therefore, exact confidence intervals cannot be build from it. However, if normality or any other distribution is assumed, such an estimation is possible. For PLs this formula is used with an approximation of the density with Matlab function `ksdensity`.

Compared to the central limit theorem for non-autocorrelated time series (with asymptotic variance $\frac{(1-\alpha)(\alpha)}{f^2(-\text{VaR}_\alpha(X))}$), the asymptotic variance is likely to increase because the autocorrelation in the tail is expected to be positive. In our case, the PLs are overlapping and the autocorrelation with lag one is between 85% and 90%. Therefore, the increase of the estimator variance may be significant. This is investigated numerically in Section 5.5.

5.3 Non-Parametric Estimation of Expected Shortfall

This section follows the same scheme as the previous section. Discrete estimators are introduced for ES. Then, the bias and the variance are investigated.

5.3.1 Discrete estimators

As we did for VaR, we define discrete estimators for ES (consistently with the estimators of [9]). Although, many estimators has been proposed in the literature, in this thesis, we only use the trivial estimators.

Definition 5.3.1 (Estimators of ES)

Let $X = \{X_1, X_2, \dots, X_n\}$ be a time series, then, the following estimators for ES are defined:

$$\widehat{\text{ES}}_\alpha^L(X) = -\frac{1}{\lfloor (1-\alpha).n \rfloor} \sum_{i=1}^n X_i \mathbb{I}_{X_i \leq -\widehat{\text{VaR}}_\alpha^L(X)}, \quad (5.13)$$

$$\widehat{\text{ES}}_\alpha^U(X) = -\frac{1}{\lceil (1-\alpha).n \rceil} \sum_{i=1}^n X_i \mathbb{I}_{X_i \leq -\widehat{\text{VaR}}_\alpha^U(X)}. \quad (5.14)$$

The difference between the lower ($\widehat{\text{ES}}_\alpha^L$) and upper ($\widehat{\text{ES}}_\alpha^U$) estimators is only one term of the sum at maximum. For the lower estimator, if the threshold level is not an integer then the number of terms in the sum is floored. For the upper estimator, one additional term is in the sum. It is easy to see that :

$$\widehat{\text{ES}}_\alpha^U(X) \leq \widehat{\text{ES}}_\alpha^L(X).$$

In fact, these are the estimators of Tail Conditional Expectation (Definition 2.1.6) and not of ES (Definition 2.1.5). An estimator of ES should be more precise when treating with the last

term of the sum. This may be done as follows:

$$\widehat{\text{ES}}_{\alpha}^M(X) = \frac{\lceil (1-\alpha) \cdot n \rceil}{(1-\alpha) \cdot n} \widehat{\text{ES}}_{\alpha}^U(X) - \frac{\widehat{\text{VaR}}_{\alpha}^U(X)}{(1-\alpha) \cdot n} \cdot (\lceil (1-\alpha) \cdot n \rceil - (1-\alpha) \cdot n), \quad (5.15)$$

if the upper ES estimator is used. The idea behind this correction term is rather simple: the part exceeding the $1-\alpha$ quantile is cut off. However, if $n \rightarrow +\infty$ the tail conditional estimators and the ES estimator are converging. This is coherent with the fact that for continuous random variables TCE and ES are equal. The way this estimator is written may seem unconventional. However, it is a convenient notation when doing simulation because it is related to a TCE estimator and a VaR estimator. Another way to do it is to add the part missing in the lower estimation:

$$\widehat{\text{ES}}_{\alpha}^V(X) = \frac{\lfloor (1-\alpha) \cdot n \rfloor}{(1-\alpha) \cdot n} \widehat{\text{ES}}_{\alpha}^L(X) + \frac{\widehat{\text{VaR}}_{\alpha}^U(X)}{(1-\alpha) \cdot n} \cdot ((1-\alpha) \cdot n - \lfloor (1-\alpha) \cdot n \rfloor). \quad (5.16)$$

In Sections 5.4 and 5.5, the error made by using TCE estimators for ES is evaluated. These two estimators are equivalent so $\widehat{\text{ES}}^M$ is kept.

5.3.2 Properties of the estimators

The bias of the estimators is first investigated. After that, the central limit theorem is established for ES estimators. In this part, it is assumed that $(1-\alpha) \cdot n$ is an integer.

Similar as VaR estimators, the ES estimators are biased. This is shown by taking a uniform random variable and computing its bias. We take the same example as for VaR: We consider the random variable $X \sim U(0, 1)$. It is obvious that $\text{ES}_{\alpha}(X) = -\frac{1-\alpha}{2}$. It is always negative because X is always positive. Suppose that a sample of n independent observations is drawn and that $(1-\alpha) \cdot n \in \mathbb{N}$. The ES estimators are equivalent and are biased, i.e.

$$\begin{aligned} \mathbb{E}(\widehat{\text{ES}}_{\alpha}) &= \frac{-1}{\lfloor (1-\alpha) \cdot n \rfloor} \mathbb{E}\left[\sum_{i=1}^n X_i \mathbb{I}_{X_i \leq -\widehat{\text{VaR}}_{\alpha}^L(X)}\right], \\ \mathbb{E}(\widehat{\text{ES}}_{\alpha}) &= \frac{-1}{(1-\alpha) \cdot n \cdot (n+1)} \sum_{i=1}^{(1-\alpha) \cdot n} i = -\frac{((1-\alpha) \cdot n + 1)}{2 \cdot (n+1)}. \end{aligned}$$

Therefore, the bias of ES estimator is equal to:

$$\text{Bias}(\widehat{\text{ES}}_{\alpha}) = -\frac{((1-\alpha) \cdot n + 1)}{2 \cdot (n+1)} + \frac{1-\alpha}{2} = \frac{-\alpha}{2 \cdot (n+1)}$$

The bias is larger for more extreme quantiles (as $\alpha \rightarrow 1$).

The bias diminishes with sample size at a linear rate. This formula is consistent with the fact that the mean estimator is not biased. If $\alpha = 0$, then $\text{Bias}(\widehat{\text{ES}}_{\alpha}) = 0$. The relative bias is given by $\frac{\alpha}{(1-\alpha) \cdot (n+1)}$

The estimators are biased for a uniform distribution and numerical experiments need to be done on financial distributions because it is likely that the ES estimators are also biased.

Similar to VaR, we investigate the asymptotic properties of the ES estimators. A central limit theorem for ES for autocorrelated time series is introduced. First, two conditions are defined for this theorem:

Given a time series $X = \{X_1, \dots, X_n\}$ the following conditions are defined:

Condition 1b:

The process $\{X_i\}_{i \in \{1 \dots n\}}$ is strictly stationary, α -mixing (see Definition B.1.3) and there exists a $\xi \in (0, 1)$ such that $\alpha(k) \leq C\xi^k$, for all k and a constant C . Furthermore, X_i should be continuously distributed with f and F as density and cumulative distribution functions, respectively. This condition just states that the distribution is sufficiently smooth and the α -mixing coefficients (a type of autocorrelation) of the time series shrinks to 0 at an exponential rate. This condition is stronger than just α -mixing.

Condition 2b:

Density f has continuous second order derivatives and the distribution of (X_1, X_k) has continuous second order partial derivatives and they are uniformly bounded with respect to k . X_1 is square integrable (the variance exists).³

This condition is technical but ensures convergence and existence of the variance.

The theorem states that if the autocorrelation decays fast enough then the asymptotic variance exists and a central limit theorem holds. Having defined the conditions, we can state the central limit theorem for ES estimator:

Theorem 5.3.1 (Central limit theorem for ES, [9])

Under Conditions 1b and 2b, it holds that:

$$\frac{\sqrt{n \cdot (1 - \alpha)}}{\sigma} (\widehat{\text{ES}}_\alpha(X) - \text{ES}_\alpha(X)) \longrightarrow N(0, 1), \quad \text{as } n \rightarrow \infty, \quad (5.17)$$

where

$$\sigma^2 = \mathbb{E}(g(X_1)^2) + 2 \sum_{i=1}^{n-1} \mathbb{E}(g(X_1) \cdot g(X_{1+i})) - \mathbb{E}(g(X_1)) \cdot \mathbb{E}(g(X_{1+i})),$$

and $g(X_i) = \mathbb{I}_{X_i < -\text{VaR}_\alpha(X)} \cdot (X_i + \text{VaR}_\alpha(X))$.

This theorem is a bit simpler than Theorem 5.2.1 for VaR because the density of the random variables does not explicitly appear in the variance. It may be just an application of Theorem B.1.1 to evaluate a mean on a restricted sample of $(1 - \alpha) \cdot n$ elements. However, the proof is more complex than that because the conditions of the theorems. Ibragimov's theorem assumes that the conditions are valid for the random variables themselves. To directly apply the theorem Conditions 1b and 2b should concern the tail of the distribution and not the entire distribution. So the time series to be considered should be $\{(X_i + \text{VaR}_\alpha(X)) \mathbb{I}_{X_i < -\text{VaR}_\alpha(X)}, 1 \leq i \leq n\}$, which is a new time series with only the terms in the tail of original time series. Therefore, Ibragimov's theorem cannot be used to prove Theorem 5.3.1.

An complete proof is given in [9] and relies on the same type of argument as in the proof of Theorem 5.2.1. The following Bahadur representation is proven under the same conditions as

³Because the time series is strictly stationary the distributions of (X_1, X_k) and (X_{i+1}, X_{i+k}) are the same

Theorem 5.3.1:

$$\begin{aligned} \widehat{\text{ES}}_\alpha(X) - \text{ES}_\alpha(X) &= \frac{1}{n \cdot (1 - \alpha)} \cdot \sum_{i=1}^n (X_i - \text{VaR}_\alpha(X)) \mathbb{I}_{X_i \leq -\text{VaR}_\alpha(X)} \\ &\quad - (1 - \alpha) \cdot (\text{ES}_\alpha(X) - \text{VaR}_\alpha(X)) + O(n^{-\frac{1}{2}}). \end{aligned} \quad (5.18)$$

Property 5.3.1 *The variance of the ES estimators is given by:*

$$\text{var}(\widehat{\text{ES}}_\alpha) = \frac{1}{n \cdot (1 - \alpha)} \left(\mathbb{E}(g(X_1)^2) + 2 \sum_{i=1}^n \left(1 - \frac{i}{n}\right) \cdot \text{cov}(g(X_1), g(X_{i+1})) \right), \quad (5.19)$$

where $g(X_i) = \mathbb{I}_{X_i < -\text{VaR}_\alpha(X)} \cdot (X_i + \text{VaR}_\alpha(X))$.

The proof is identical as for VaR estimators (Property 5.2.1) and is given in [9].

5.4 Application to Independent Distributions

After stating various theorems to assess the bias and the asymptotic normality of the estimators, we test the estimators numerically. Three classical indicators are considered to assess the quality of an estimator: the bias, the Standard Deviation (SD) and the Mean Square Error (MSE). They have been defined in Section 5.1.

5.4.1 Definition and methodology

Two of these tests require to know the true value of the estimator. Therefore, the tests are performed on well known distributions: Normal distribution and Student's t-distributions with 4, 5 and 6 degrees of freedom. A low degree of freedom indicates a fat tail.

The objective of this analysis is to give an insight in of the quality of VaR and ES estimators.

For time series without autocorrelation, the implementation is straightforward: n points of a distribution are drawn and the VaR and ES are extracted. To compute the MSE, SD and the bias, this operation is repeated 10000 times. The exact quantiles are determined by the inverse cumulative distribution functions given in Appendix B.3. For ES, the distributions used all have a closed form formula for the quantile functions which are given in Appendix B.3. Therefore, it is subject to a small numerical integration error. All values are expressed in percentages in order to be comparable for all thresholds. In this implementation, the sizes of the samples are taken in accordance with the samples that are available in practice. Therefore, sample sizes of 250, 500 and 1000 are used. They represent, respectively, 1, 2 and 4 years of trading. Appendix B contains all the tables used for the analysis in this section.

5.4.2 Numerical results

Bias analysis

The bias of the different estimators proposed in Sections 5.2 and 5.3 are analysed. Five estimators have been defined: two for VaR and three for ES. Results depend on the estimator and on the sample size. First, the biases of the two VaR estimators are analysed and then, the

ES estimators. Figure 5.1 presents the bias for the two VaR estimators with $n=250$ and for Student's t-distribution with 5 degrees of freedom.

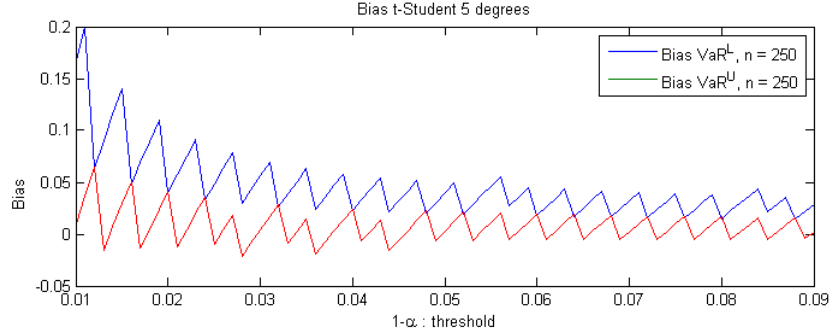


Figure 5.1: Bias of the VaR estimators.

The two estimators of VaR do not have the same bias. The lower estimator ($\widehat{\text{VaR}}^L$) has a positive bias, which means that, for small samples it overestimates the risk. For extreme quantiles, this overestimation can be as high as 20%. This behaviour is observed for all distributions (see Tables C.1, C.2, C.3, C.4, C.5, C.6, C.7 and C.8). The upper estimator ($\widehat{\text{VaR}}^U$) appears to be more accurate if a rounding of the threshold is necessary.

Oscillations are caused by the rounding of the threshold. As the sample is too small, the rounding makes that the estimator does not estimate the threshold because there is not sufficient data beyond the threshold. For instance, the 99% VaR is estimated by the third worst observation for the upper estimator and by the second for the lower whereas it should be estimated by the 2.5th worst observation. This oscillations phenomena is mitigated as n increases. With $n=1000$, the bias is always lower than 2% for all distributions and all thresholds.

When $n \cdot (1 - \alpha)$ is an integer (the two estimators are the same), the estimators are still biased and have positive bias for all distributions investigated here (see Tables previously mentioned). Figure 5.1, the bias without rounding is 5% for the 98% threshold. The positive bias represents an overestimation of the risk. VaR estimators typically over-estimate the risk.

The two functions should intersect when $n(1 - \alpha)$ is an integer. This is not the case in Figure 5.1 because of Monte Carlo noise.

The conclusion is that the estimators should be used with a sufficient amount of data, so that if rounding is used, it does not generate a large bias. The bias is significant even without rounding for extreme quantiles and with $n = 250$. A larger sample size should be used to have a smaller bias.

The case of ES leads to similar results for small sample sizes. However, it is the opposite which is observed: the upper estimator leads to a large negative bias (underestimation) and the lower estimator leads to an oscillating bias around 0%. Figure 5.2 presents the bias of the ES estimators for the Student's t-distribution with 5 degrees of freedom for $n=250$.

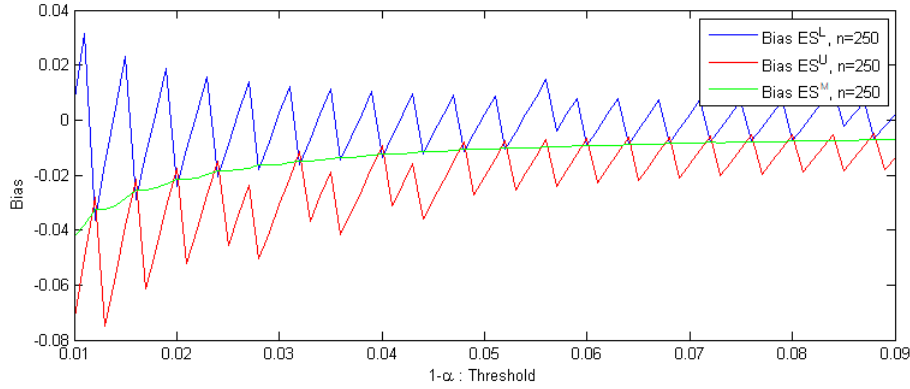


Figure 5.2: Bias for ES, Student's t-distribution.

Results for all the distributions for ES are given in Tables C.9, C.10, C.11, C.12, C.13, C.14, C.15 and C.16. An increase of n leads to a consequent reduction of the bias. With $n = 1000$ the biases of these three estimators are no longer significant ($< 1\%$) for all distributions and all thresholds.

When the value of $n(1 - \alpha)$ is an integer, the ES estimators exhibit a negative bias so they underestimate the risk. The \widehat{ES}^M estimator has the same properties as the lower estimator \widehat{ES}^L . The additional term limits the oscillations and the bias is lower than 4%, but it makes the estimation less intuitive and does not reduce the bias if $n(1 - \alpha)$ is an integer.

A comparison between the bias of VaR and ES estimators shows that the ES estimators are less biased (in absolute value) than the VaR estimators for each threshold level. However, they underestimate the risk which may be a problem in risk management.

Mean Square Error

A very commonly used error to assess the quality of an estimator is the Mean Square Error (MSE) as defined in Section 5.1. We follow the same structure as for the bias.

For the VaR estimators, the results are as it follows: The upper estimator has a lower error than the lower estimator. It was to be expected, because it has a lower bias and it estimates a less extreme thresholds level which leads to a lower variance for the upper estimator. The MSE is high for $n = 250$, and therefore more data has to be used. MSE decreases at a rate of, roughly, \sqrt{n} as n increases (this is coherent with the central limit theorem). These results are shown in Figure 5.3.

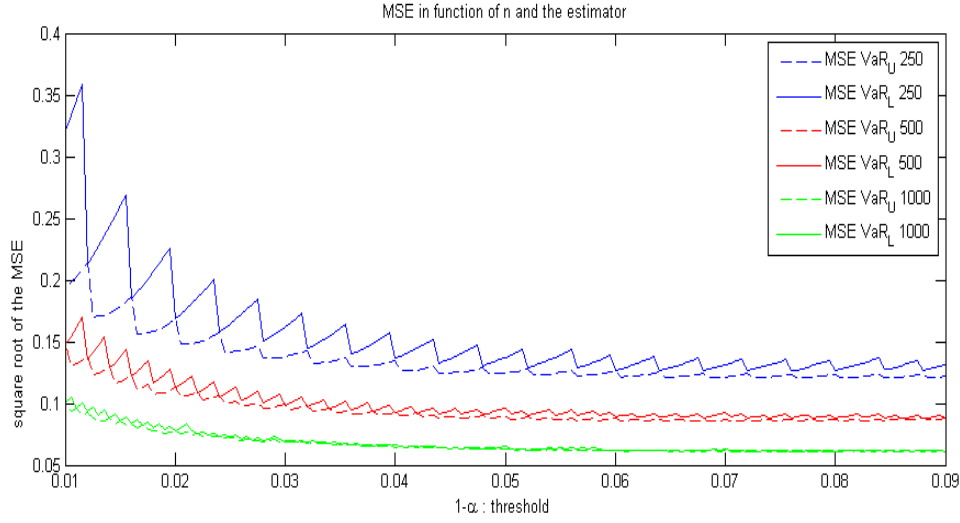


Figure 5.3: MSE for VaR, Student's t-distribution with 5 degrees of freedom.

For ES, the upper estimator (\widehat{ES}^U), that leads a high negative bias, has a lower MSE than the other estimators. Although the upper estimator is more biased, the additional term in the sum of ES brings a higher stability and decreases the standard deviation of the estimator. The last estimator (\widehat{ES}^M) reduces the MSE especially when the threshold is rounded compared to the lower estimator. This is shown in Figure 5.4.

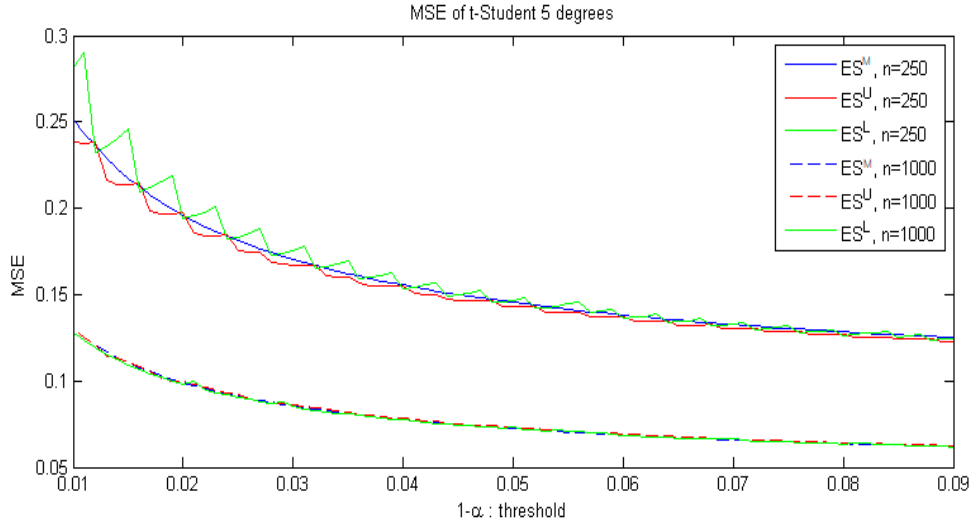


Figure 5.4: MSE for ES.

Although it is difficult to compare MSE of ES and VaR estimators because the thresholds taken in practice are different, we make a comparison for the thresholds 95% and 97%, for $\widehat{ES}^U/\widehat{ES}^L$ on one hand and, for 97.5% and 99%, for \widehat{VaR}^U on the other hand with Student's t-distribution 5 degrees of freedom. These are thresholds commonly used in the banking industry and proposed by the regulator in [43] and [38]. The conclusion is that the MSE are low for time series with $n = 1000$. The differences of MSE between VaR and ES are not very large. The explanation is that two effects are compensating: On one hand, ES is a mean and is, therefore, more stable

than a point estimation and the threshold is lower. On the other hand, extreme quantiles are in the sum which make ES estimators more volatile. All these factors together make that the MSE of ES and VaR are approximately the same.

Table 5.1: MSE for Student's t-distribution with 5 degrees of freedom for VaR and ES, $n=1000$.

$\widehat{\text{VaR}}^U$	t-5	99.0% 9.9%	97.5% 7.3%
$\widehat{\text{ES}}^U$	t-5	97% 8.5%	95% 7.2%
$\widehat{\text{ES}}^L$	t-5	97% 9.8%	95% 7.3%

Accuracy of theoretical variance and application to the portfolio

The two standard deviation formulas in Theorems 5.3.1 and 5.2.1 are assessed with independent samples. Therefore, the sum of covariances equals 0.

First of all, the theoretical variance (σ^2) is analysed given a threshold level. In Figure 5.5, we present the standard deviation versus the threshold for a normal distribution and a Student's t-distribution with 5 degrees of freedom. They are normalized by the true VaR.

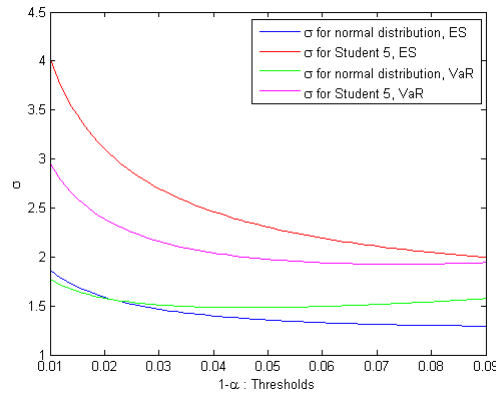


Figure 5.5: Asymptotic normalized standard deviation.

As the standard deviation is normalized, it increases when the VaR tends to 0. This point is at 0 for these examples (the mean is 0 and the distributions are symmetric so $\text{VaR}_{50\%} = 0$). On the other hand, as the probability density function, $f(\text{VaR}_\alpha(X))^2$, decreases in the tail, the standard deviation increases too. The interpretation is that there exists a point where, asymptotically, the standard deviation of the estimator is minimal. In practice, this point is often too far from the tail and does not help to approximate the risk as it does not accurately capture the tail risk.

In general, the theoretical standard deviation gives an accurate approximation of the true standard deviation.

The example is the Student's t-distribution with 5 degrees of freedom. One may remark that this approximation is accurate if the threshold is not rounded. Theoretical standard deviations (σ) and empirical standard deviations (SD) are given for a $n = 500$ in table 5.2.

Table 5.2: Theoretical SD and empirical SD for Student's t distribution with 5 degrees of freedom, $n=500$.

$\widehat{\text{VaR}}^U$	t-5	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
	SD (%)	14.4%	11.7%	11.1%	10.1%	9.5%	9.4%	9.3%	9.0%	9.0%
	Theo std (σ)	13.2%	11.6%	10.7%	10.1%	9.7%	9.3%	9.1%	9.0%	8.8%
$\widehat{\text{VaR}}^L$	t-5	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
	SD (%)	14.5%	12.8%	11.1%	10.6%	9.9%	9.6%	9.2%	9.1%	8.9%
	Theo std (σ)	13.2%	11.6%	10.7%	10.1%	9.7%	9.3%	9.1%	9.0%	8.8%
$\widehat{\text{ES}}^L$	t-5	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
	SD (%)	18.1%	14.0%	12.1%	11.0%	10.6%	9.8%	9.5%	9.1%	8.8%
	Theo std (σ)	18.0%	13.9%	12.1%	11.0%	10.3%	9.8%	9.4%	9.1%	8.9%
$\widehat{\text{ES}}^U$	t-5	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
	SD (%)	17.8%	13.7%	11.9%	10.9%	10.2%	9.6%	9.3%	9.0%	8.7%
	Theo std (σ)	18.0%	13.9%	12.1%	11.0%	10.3%	9.8%	9.4%	9.1%	8.9%

For the historical PLs the procedure is more complex. The probability density function and the exact VaR are not available, so this standard deviation can be difficult to estimate. For these two quantities, there is a need of an estimator. For VaR, the $\widehat{\text{VaR}}^U$ is used and for the probability density function, a Matlab kernel smoother is employed (function `ksdensity`). However, the error due to all these approximations is unknown. Therefore, the results may seem realistic, but they should be interpreted with caution. This is shown in Figure 5.6 with 1000 PLs:

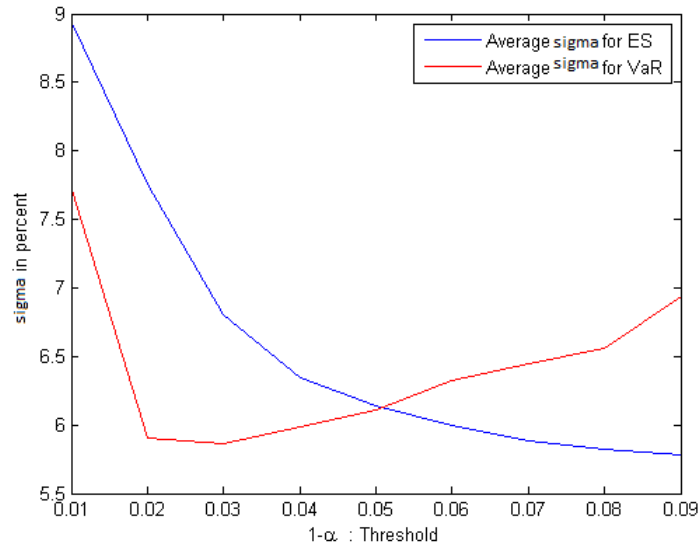


Figure 5.6: Normalized theoretical standard deviation of the estimators.

Conclusion

Several conclusions may be drawn from the analysis performed in this section: For independent samples, the results are not surprising: The estimation of risk measures with extreme thresholds leads to a higher error. To minimize the error the maximum amount of data that is available should be used. If $n \cdot (1 - \alpha) \notin \mathbb{N}$ then $\widehat{\text{VaR}}^U$ should be used for VaR to minimize the effect of the rounding. About ES, both estimators have positive and negative aspects. $\widehat{\text{ES}}^L$ has a lower bias (in absolute value) but $\widehat{\text{ES}}^U$ has a lower MSE. If $n \cdot (1 - \alpha) \in \mathbb{N}$ then VaR tends to overestimate the risk and ES tends to underestimate the risk, however the biases are not significant for 1000 PLs. A more extreme threshold leads to a worse estimation (higher MSE).

The estimator $\widehat{\text{ES}}^M$ is not used in the following parts of the analysis as it adds complexity to the model without any significant effect if there is no rounding. For the purpose of Economic Capital modelling, it is common to take the 25th worst PL of the sample instead of the VaR 97.5%. For ES, it is common to take the average of the 50 worst values.

5.5 Application to Autocorrelated Time Series

The application to highly autocorrelated time series is more complex as it requires more advanced methods to generate time series. The bias, standard deviation and the MSE, as defined in Section 5.1, are used. This section has the same organisation as the previous section: First some processes are introduced. Then, numerical experiments are performed on the processes and the historical PLs. The objective is to measure the impact of having highly autocorrelated (because overlapping) PLs compared to the independent case. Autocorrelation should diminish the quality of the data and should make VaR and ES estimation more difficult.

5.5.1 Processes and methodology

In this section, two processes are described, they generate autocorrelated time series that may be used to mimic the behaviour of the PL distribution which is described in Section 3.2. We use only two basic processes: an AR(1) process and a time series from correlated Student's t-distributions with 5 degrees of freedom.

Autocorrelation

In Section 3.2.3, the autocorrelation of non-overlapping PLs has been discussed. However, for the analysis of the estimators, the autocorrelation of the overlapping PLs seems more important. Therefore, we start by investigating the autocorrelation of overlapping PLs to understand which autocorrelation structure the model should reflect. The results are given in Figure 5.7, the autocorrelation of the historical PLs is presented at three different dates. For small lags⁴ (less than ten), the autocorrelation is very high due to overlapping PLs, then, we see is a cyclic behaviour. It is alternatively positive or negative. However, for a lag larger than ten, data is not sufficient to draw any conclusion about autocorrelation.

⁴For the autocorrelation the lag the index h such that $AC(h) = \frac{\mathbb{E}[(X_i - \mu)(X_{i+h} - \mu)]}{\sigma^2}$, where X_1, \dots, X_n is a time series with mean μ and variance σ^2

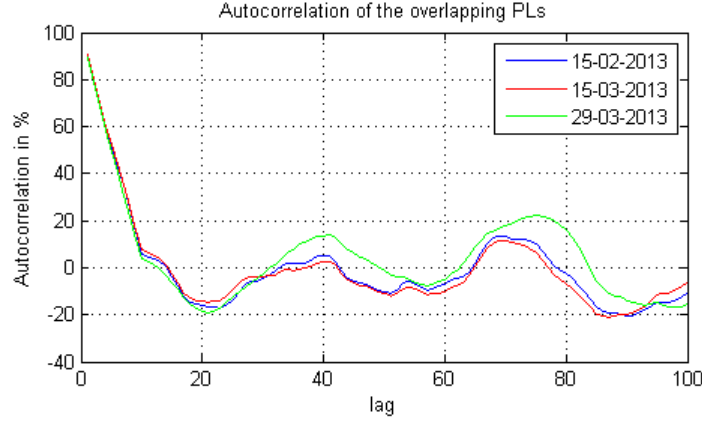


Figure 5.7: Autocorrelation of the PLs.

We assume that the PLs satisfy the conditions of Theorems 5.3.1 and 5.2.1. The following should hold for the α -mixing coefficients: $\exists \xi \in (0, 1)$ such that $\alpha(k) \leq K\xi^k$. However, when only small samples are used, a large K is sufficient. As the sample sizes are always small the conditions of the theorems may be assumed to be satisfied. Economically, this assumption makes sense: The 10-day PLs are not likely to be influenced by the 10-day PLs from 3 years ago.

Processes

To assess the quality of the estimators for autocorrelated time series, two processes are generated. They are defined as follows:

- AR(1) process. The time series is modelled using an autoregressive process.

$$X_1 = \epsilon_1, \quad X_t = cX_{t-1} + \epsilon_t \quad \text{for } t > 1,$$

where $\epsilon_t \sim N(0, \sigma^2)$, c and σ are constant.

This process is basic and models an exponentially decreasing autocorrelation. If the size of the sample is large, the autocorrelation decreases quickly to 0 and it is given by $g(i) = c^i$, where i is the lag.

- Correlated Student's t random variable with 5 degrees of freedom. A Gaussian copula is used to model the correlation as it is shown in Algorithm 2.

```

rand =  $F^{-1}(\text{random}(0,1))$  ;
for period = 2 : n do
    Z = random number from a N(0,1) distribution ;
    rand =  $c \cdot \text{rand} + \sqrt{1 - c^2} \cdot Z$  ;
    D(period) =  $F^{-1}(\text{rand})$  ;
end

```

Algorithm 2: Algorithm for Student's correlated time series.

There, F is the cumulative distribution function of the normalized Student's t-distribution, c is the Gaussian copula correlation.

This process is named T5 in the rest of this analysis.

The implementation of Algorithm 2 is straightforward but the computational time is by far higher than in the independent case. The number of simulations to compute the MSE, bias and standard deviation is reduced to 1000 for the two processes. As a consequence, a larger error for the values of these indicators is expected.

Autocorrelation of the time series

To compute the theoretical standard deviation from Theorems 5.2.1 and 5.3.1, we need to verify that the autocorrelation resembles the desired autocorrelation and we need to compute the sum of the covariances (sum in Theorems 5.3.1 and 5.2.1). These sums reflect the auto-covariance in the tail because the other terms are set to 0. First of all, the autocorrelations of the two models are checked. In Figure 5.8, the correlation parameters (c) are set 85%. The autocorrelation structures from T5 and AR processes are rather simple: They are just decaying, but for small sample sizes the behaviour may be different. In Figure 5.8, the AR and T5 autocorrelations are compared. The autocorrelations of the processes and for the historical PLs are very similar. There are quite some oscillations around 0 for larger lags due to a lack of data.

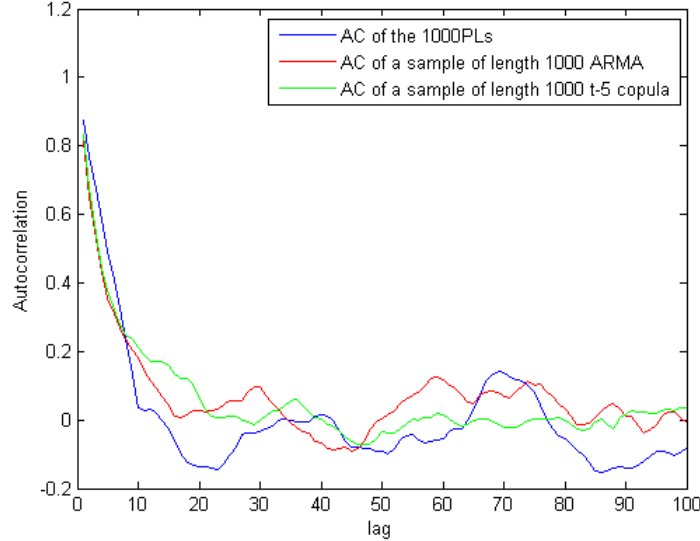


Figure 5.8: Autocorrelation of AR and T5 processes.

However, as it may be seen in Theorems 5.3.1 and 5.2.1, the increase of the variances of the estimators is due to the autocorrelation in the tail. The terms we need to evaluate are:

$$\sigma^2 = (1 - \alpha) \cdot \alpha + 2 \sum_{h=1}^{\infty} \text{cov}(\mathbb{I}_{X_1 \leq -\text{VaR}_{\alpha}(X)}, \mathbb{I}_{X_{1+h} \leq -\text{VaR}_{\alpha}(X)}), \quad (5.20)$$

for the VaR and:

$$\sigma^2 = \mathbb{E}(g(X_1)^2) + 2 \sum_{i=1}^{n-1} \text{cov}(g(X_1), g(X_{1+i})), \quad (5.21)$$

where $g(X_i) = \mathbb{I}_{X_i < -\text{VaR}_\alpha(X)} \cdot (X_i + \text{VaR}_\alpha(X))$, for ES. These two quantities are related to the autocorrelation in the tail. For historical PLs, this autocorrelation is difficult to estimate because there are only few values in the tail. For processes such as T5 and AR, it can be computed more easily using a numerical simulation. As an equivalent quantity, Figure 5.9 gives the classical autocorrelation for lags 1 and 5 of the time series: $\{\mathbb{I}_{X_1 \leq -\text{VaR}_\alpha(X)}, \mathbb{I}_{X_2 \leq -\text{VaR}_\alpha(X)}, \dots, \mathbb{I}_{X_n \leq -\text{VaR}_\alpha(X)}\}$.

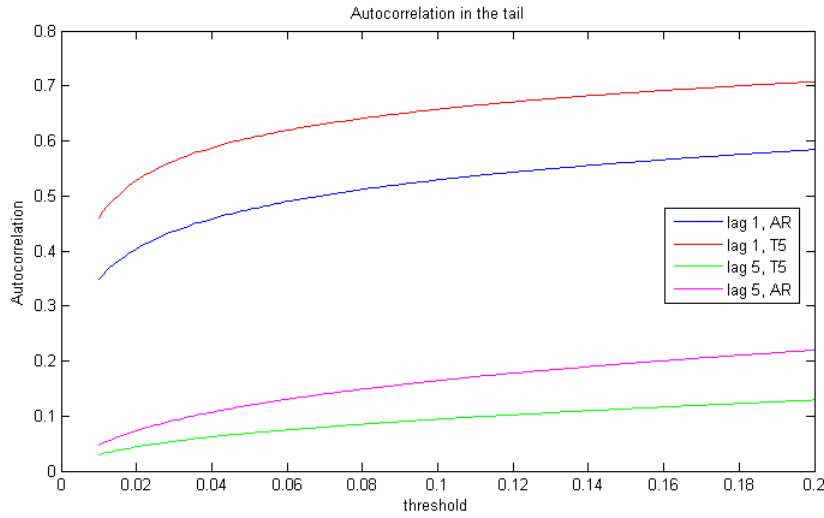


Figure 5.9: Autocorrelation in the tail of AR and T5 processes.

We see that for extreme thresholds the autocorrelation is lower. The implication is that the variances of the estimators of VaR and ES are likely to increase more for less extreme thresholds compared to the independent case. The impact on the MSE is that it may be more threshold independent (constant). This is due to the fact that as seen in Figures 5.4 and 5.3, the MSE curves are rather steep (MSE is tightly connected to the SD not to the bias). The autocorrelation is likely to increase the MSE for non-extreme thresholds. Numerical experimentations are performed in Section 5.5.2, they globally confirm this reasoning.

5.5.2 Numerical results

The same indicators as in the independent tests are presented and the results are analysed. The reference value, when analysing the bias or the MSE, is the average of the risk measure taken with $n = 10000$ (length of the time series) on 1000 simulations for T5 and using the long-term distribution of an AR process. The bias and MSE are investigated.

Bias analysis

For autocorrelated time series, the bias is different than for the independent case. In fact, the behaviour is more erratic, less intuitive, and process dependent. There are many parameters to take into account and it is not easy to understand the behaviour a priori. First, the analysis is

performed for the two estimators of VaR.

The bias increases compared to the independent case. However, in the independent case, we saw that the bias was homogeneous and was of the same sign for all the distributions tested for a given estimator. For autocorrelated time series, the lower VaR estimator has a different bias depending on whether it is an AR process or T5 process. The bias is around 0 for the former and positive for the latter.

Concerning ES, the bias is negative for both estimators (underestimation) and larger (absolutely) than in the independent case. In the independent case, \widehat{ES}^U has a bias lower than -8% (in absolute value) and \widehat{ES}^L has a bias lower than -3% for the α investigated. With autocorrelation, the biases are more than -10% when $n(1 - \alpha) \notin \mathbb{N}$. This is a consequence of the autocorrelation: For small sample sizes, estimation of ES is more difficult. Longer time series are less subject to biases as shown in Tables C.20, C.21, C.22 and C.23, .

When $n(1 - \alpha)$ is an integer, VaR estimators have a positive bias for T5 (up to 3%) and negative for AR (up to -3%). ES estimators are underestimating the risk by more than 8%. This has to be taken in consideration when computing ES for autocorrelated time series. ES estimators tend to underestimate significantly the risk when used on small sample sizes.

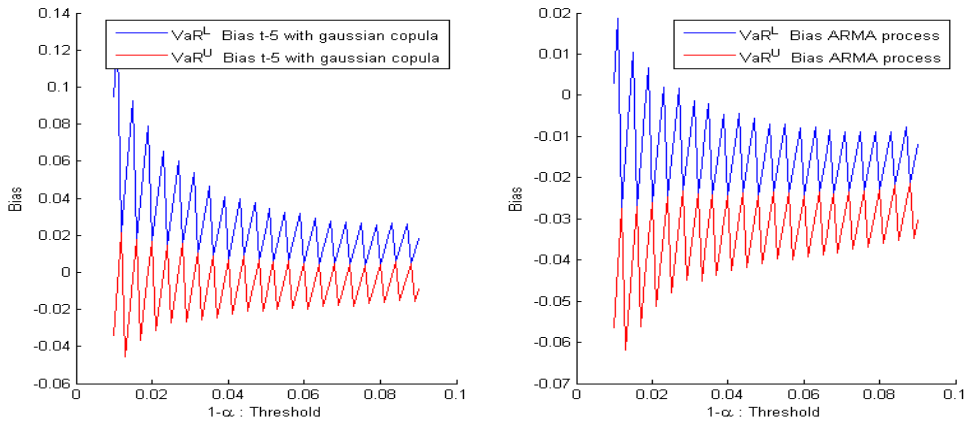


Figure 5.10: Bias for VaR estimators, $n=250$.

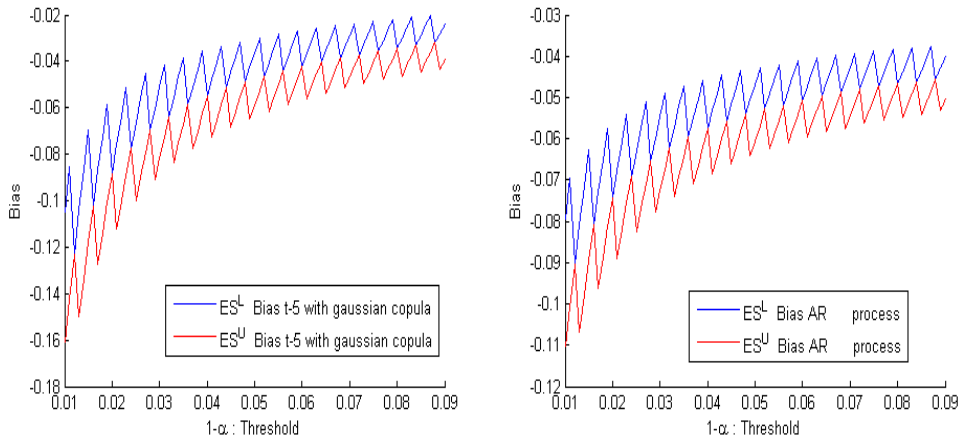


Figure 5.11: Bias for ES estimators, $n=250$.

Mean square error

In this section, the MSE is analysed for autocorrelated time series. It is the main error quantifier when analysing the quality of an estimator. As for the independent case, we consider the cases $n = 250$ and $n = 1000$ for $\widehat{\text{VaR}}^U$, $\widehat{\text{VaR}}^L$ in Figure 5.12 and $\widehat{\text{ES}}^U$, $\widehat{\text{ES}}^L$ in Figure 5.13.

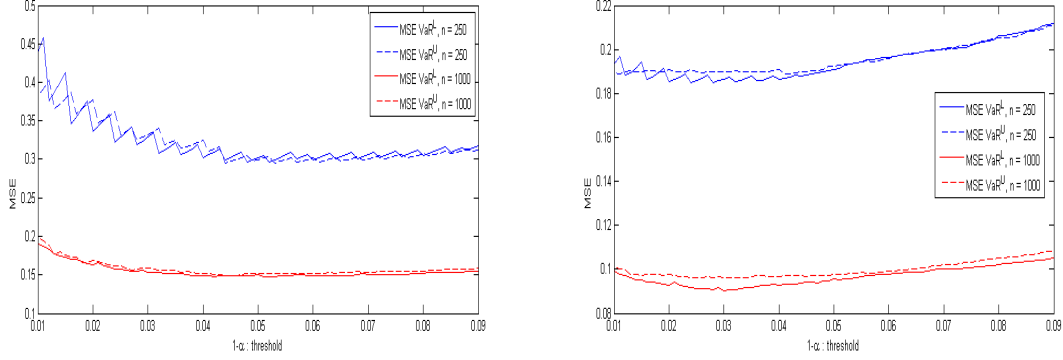


Figure 5.12: MSE VaR for T5 (left) and AR (right).

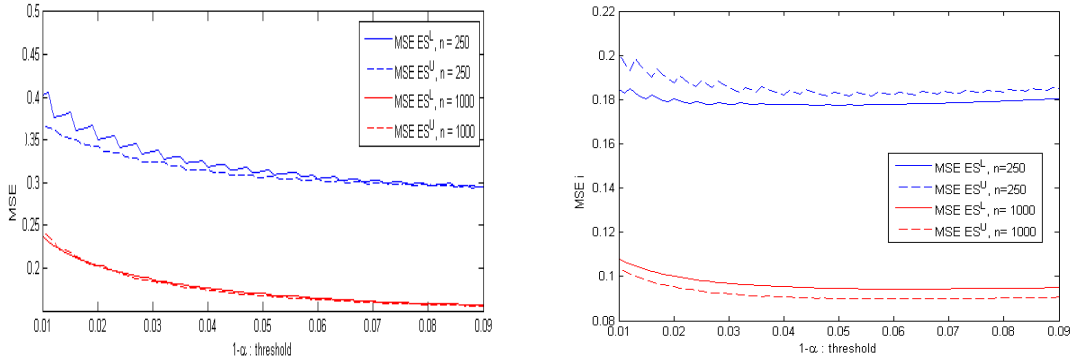


Figure 5.13: MSE ES for T5 (left) and AR (right).

As expected the MSE are higher than in the corresponding independent cases (Figures 5.4 and 5.3). It makes sense as the autocorrelation diminishes the quality of the data and makes the estimation of VaR or ES more difficult. The AR and T5 processes lead to different results for VaR. The AR process has a decreasing error as the threshold is further in the tail. This result is surprising because we expect that the higher the threshold the more difficult the estimation of VaR is. Therefore, the MSE of VaR should be higher when the threshold level is further in the tail. In fact, this is a consequence of the high autocorrelations of the processes: The autocorrelation beyond the threshold increases when the threshold is far from the tail. This feature is in accordance with the explanation of Figure 5.9.

For ES, this behaviour is less obvious although for $n=250$, the MSE is almost constant for the AR process (18%, Figure 5.13). For the T5 process, the increase compared to the Student's t-distribution is quite high, for $n=1000$ and $\alpha = 0.95$, the MSE increases from 5% to 10% which is a significant increase.

Theoretical standard deviation

The theoretical asymptotic standard deviations (σ) of Theorems 5.3.1 and 5.2.1 do not give an accurate approximation as it was the case for non-autocorrelated time series.

A first look at the variance formulas gives a general overview of the behaviour: If an observation in the tail is very likely to be followed by another observation in the tail then the covariance sum is likely to be high. Therefore, the sum increases and so does the variance. Then, a problem is that this sum also depends on the number of observations (the use of the infinite sum is not realistic) and the covariance may change depending on the thresholds of the risk measure being estimated.

The covariance for a less extreme threshold is likely to be higher because the probability to stay in the tail is higher. This creates a standard deviation less dependent on the threshold because the standard deviation is increased by the covariance sum more for non extreme thresholds than for the extreme ones. The variance becomes less dependent on the threshold being estimated. This fact has already been seen in Figure 5.13. The MSE of time series without autocorrelation was dependent on the threshold whereas for the AR process it is almost constant (see Figure 5.4).

Some approximations are done in the computation such as for the density function (approximation by a Matlab function `ksdensity`) and the covariance sum (truncated to 15 terms). This leads to results for VaR that should be interpreted with caution (see Figure 5.14). Results are in line with [9] where the author also found unreliable results for the theoretical variance on small sample sizes for VaR.

However, results for ES are more reliable because there is no need to compute the probability density function explicitly. Furthermore, they may be interpreted as confidence intervals as follows:

$$\frac{\sqrt{n \cdot (1 - \alpha)}}{\sigma} (\widehat{\text{ES}}_{\alpha}(X) - \text{ES}_{\alpha}(X)) \rightarrow N(0, 1), \quad (5.22)$$

which implies than a confidence interval is given by:

$$[\widehat{\text{ES}}_{\alpha}(X) + \frac{\sigma}{\sqrt{n}} \psi^{-1}(q) ; \widehat{\text{ES}}_{\alpha}(X) - \frac{\sigma}{\sqrt{n}} \psi^{-1}(q)], \quad (5.23)$$

where ψ is the cumulative density function of the standard normal distribution and q is the confidence level.

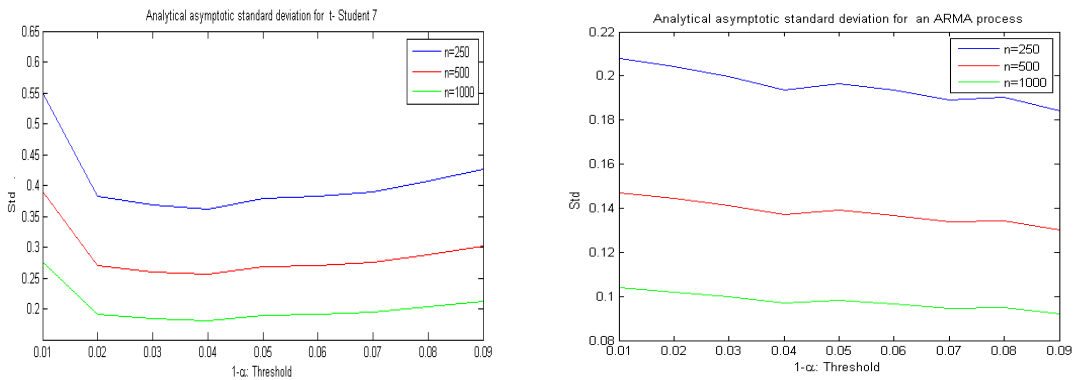


Figure 5.14: Standard deviations for VaR based on the analytical formula.

The standard deviation of ES estimator for Student's t-distribution has been increased by more than 50% compared to the independent case (see table 5.2).

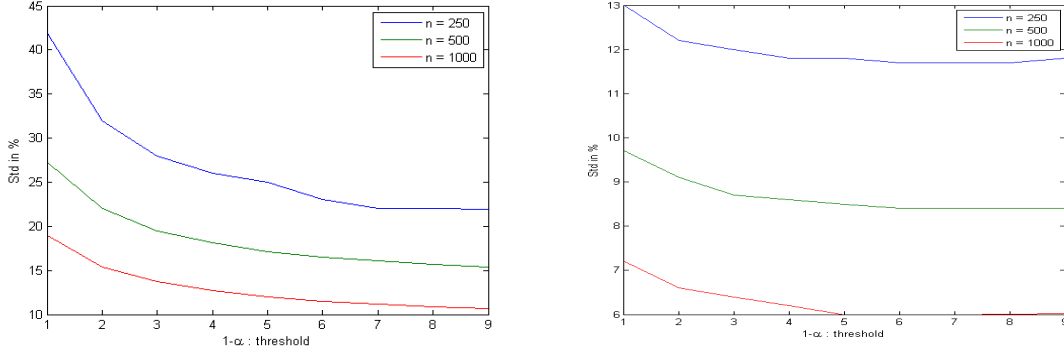


Figure 5.15: Standard deviations for ES based on the analytical formula.

Then, for ES, the theoretical standard deviation is computed for 1000 historical PLs. For more reliability, the results are averaged over 15 dates on February and April 2013 (Figure 5.16).

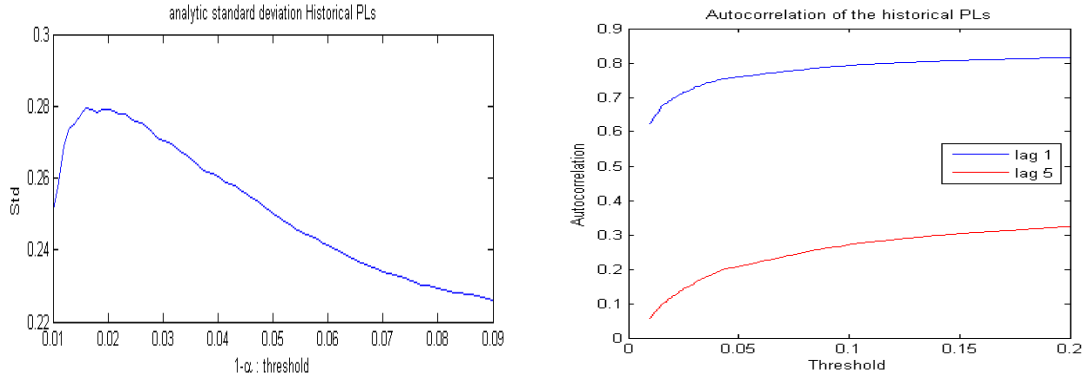


Figure 5.16: Theoretical standard deviation of the historical PLs (left) and autocorrelation in the tail (right).

Figure 5.16 is not usual because the theoretical standard deviation for extreme thresholds ($\geq 99\%$) is a lot lower than the standard deviation of the lower thresholds (lower than 98%). In fact, this behaviour is typical when working with time series without a lot of observations. In Formula 5.19, the covariance in the tail is difficult to estimate for extreme quantiles if the distribution is unknown. This sum tends to 0 when the sample size is small. If only one value is in the tail then the variance will be 0.

Therefore, the results should be analysed with a certain error. For extreme thresholds, the standard deviation is most probably underestimated. However, compared to Figure 5.5, the standard deviation increases drastically reflecting the importance of autocorrelation when estimating ES.

The increases are larger than expected (in comparison to the AR and T5 processes) are due to a higher and steeper autocorrelation in the tail (left graph of Figure 5.16 and Figure 5.9 for

$\alpha = 0.95$ then for historical PLs the autocorrelation is 20% whereas it is 5% and 10% for T5 and AR, respectively).

5.6 Conclusion and Further Research

Time series' autocorrelation was a major problem when estimating VaR or ES. As it was shown in this chapter, the bias was not negligible for small samples. Time series shorter than 1000 PLs should not be used to avoid a significant bias even without a rounding ($n \cdot (1 - \alpha) \in \mathbb{N}$). For ES, the bias was negative for extreme thresholds, so it may underestimate the Economic Capital. The MSE lead to the same analysis. Autocorrelation increased the MSE and it remained high even for 1000 PLs (more than 10% for both ES and VaR). Last but not least, the implementation should avoid rounding ($n \cdot (1 - \alpha) \in \mathbb{N}$) to prevent from additional noise and keep the estimators simple (no use of \widehat{ES}^M).

From this chapter, we saw that 1000 PLs should be used in order to minimize the bias and the MSE. For ES, a threshold lower than 97% had to be used to avoid any significant bias (less than 2%) and a MSE lower than 20% for autocorrelated time series. VaR had an almost constant MSE (but which is high) for autocorrelated samples.

In the literature (see [9]), other estimators are proposed to reduce the MSE. They usually add complexity because kernel estimators (smoothers) are used. The reduction of the MSE is usually rather small (less than 2% of reduction for both VaR and ES).

Other methods are used to reduce the bias such as bootstrapping techniques. These methods are very efficient as shown in [27]. However, it does not make sense to use them for our purposes because we are looking for an approximation of Algorithm 1. The methods proposed are often more complex than Algorithm 1.

Chapter 6

Market Fluctuation Risk

In Chapter 5, we analysed the properties of various estimators of VaR and ES. In this chapter, these estimators are applied to historical PLs. The outcome is analysed and the results are applied to compute the scaling factor. The scaling factor, defined in Chapter 4, extends the VaR to a 1-year capital horizon and increases the threshold to 99.99%. Both VaR and ES are employed in this analysis.

In Section 6.1, the general framework is introduced and the requirements for the scaling factor are explained. In Section 6.2, VaR and ES are compared, then, a linear relation is investigated. One of the components of the scaling factor is given by the VaR of a sum of random variables. Therefore, properties of the VaR of the sum are investigated for classical distributions and historical PLs in Section 6.3. In Section 6.4, criteria used to assess VaR and ES, and also to choose a threshold level are defined. This section also provides numerical results. In Section 6.5, a last requirement is investigated: this is that the scaling factor should be rounded. Finally, the impact on the market fluctuation risk is analysed in Section 6.6 for historical PLs.

The objectives are to determine which of VaR or ES is more suitable to model the market fluctuation risk and to determine an appropriate threshold. The VaR which is obtained from the sampling Algorithm 1, is called sampled VaR.

6.1 Introduction and Model Requirements

The market fluctuation risk is the main component of the Economic Capital for the trading book. The scaling factor is one of the main parameters and therefore, having a reliable number is crucial. The scaling factor is given by:

$$SF = \frac{\widetilde{\text{VaR}}_{\alpha}(X, n\Delta t)}{\rho(X, \Delta t)},$$

where $\widetilde{\text{VaR}}_{\alpha}(X, n\Delta t)$ is computed following Algorithm 1, ρ is a risk measure (VaR or ES), X is a PL distribution (a random variable), Δt is the liquidity horizon, $T = n\Delta t$ is the capital horizon and $\alpha = 99.99\%$. The objectives of this chapter are to select an adequate risk measure and a threshold to fulfil in the best way as possible the requirements of the Economic Capital defined in Section 2.2.3. Next to these general requirements, there are requirements the scaling

factor should comply with.

The Economic Capital should cover 99.99% of the yearly possible scenarios that may occur in the market. Therefore, the model employed should cover the tail risk accurately because 99.99% is an extreme threshold. So, the scaling factor has to be based on an extreme threshold. This requirement is in contradiction with results from Chapter 5 because we have shown that for accurate VaR and ES estimations, the threshold should not be too extreme to avoid a significant bias and a large MSE.

The scaling factor is fixed and should be reviewed regularly to take in consideration major changes in the portfolio structure (major change of pricing function, change of strategy, change of scope...). The following three requirements are given for the scaling factor:

- It should be stable (relatively portfolio or data independent) to always reflect the correct scaling factor and avoid multiple reviews (see Section 6.4.1). It means that, despite some changes in the products, in the PLs or exposures to a new risks, the scaling factor should be rather stable over short periods of time.
- It should be conservatively rounded (see Section 6.5).
- The final result $SF \cdot \rho(X, 10d)$ should have a low error and a high correlation with the sampled VaR (see Sections 6.4.2 and 6.4.3).

The general idea is that the scaling factor must be fixed over time and accurate (at least over short periods of time).

6.2 Expected Shortfall versus Value at Risk

We start by analysing how VaR relates to ES. For a bank, there is no reason to move from VaR to ES if the relationship between them is one to one. First, the behaviour of each risk measure is analysed based of 1000 PLs (4 years) for several portfolio snapshots. We use 1000 PLs because we saw in Chapter 5 that the maximum number of PLs that is available should be used to minimize the bias and the MSE of the VaR and ES estimators with highly autocorrelated time series. The surfaces of the ES and VaR of 10-day historical PLs for 14 dates and various thresholds are provided in Figure 6.1. The main conclusion is that ES is smoother for a given date. This may imply that ES is less dependent on the threshold used.

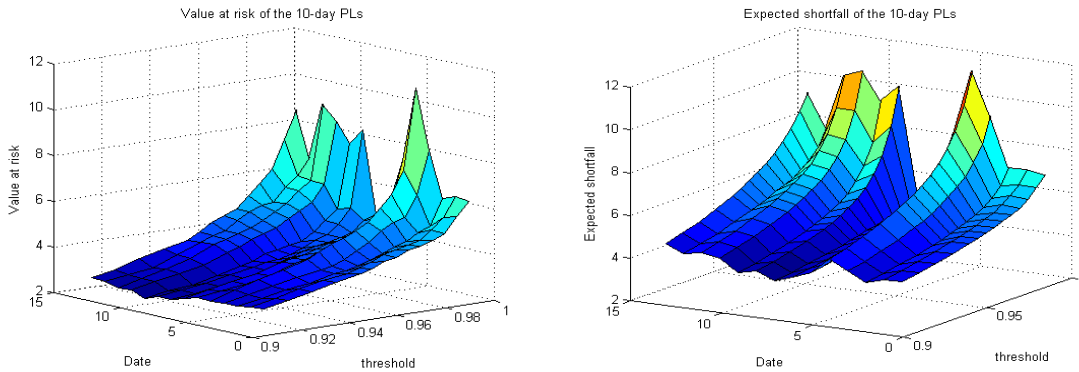


Figure 6.1: Surfaces of VaR (left) and ES (right).

We need to analyse the joined behaviour of the two risk measures. One of the main questions

is whether it is possible to find a constant k such that,

$$\text{ES}_\alpha(X_i) \approx k \times \text{VaR}_{\alpha'}(X_i), \quad \text{for } i = 1, \dots, D,$$

where α and α' are threshold levels, X_i is the PL distribution at date i and D is the number of dates. This analysis is performed, by convenience, for thresholds for which the ratio is around 1 in Figure 6.2. The couples $(\text{ES}_\alpha, \text{VaR}_{\alpha'})$ used for this figure are $(\text{ES}_{93\%}, \text{VaR}_{97.5\%})$ and $(\text{ES}_{97.5\%}, \text{VaR}_{99\%})$. The conclusion is that, even on a short period of time (three months, one date per week), the ratio is not very stable. The ratio fluctuates between 0.9 and 1.1 which may lead to different estimations of the Economic Capital of hundreds of millions Euros for a bank. Therefore, more analysis should be done in order to understand the empirical differences between VaR and ES.

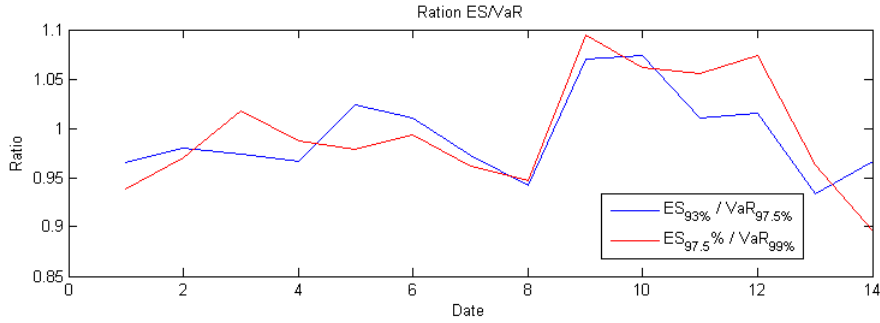


Figure 6.2: Ratio ES/VaR.

An important requirement is that the risk measure should reflect the tail risk. We measure this by computing the kurtosis of the PLs. Over time, the kurtosis of the portfolio is changing, we wish to analyse if the risk measures vary accordingly with the kurtosis. If the 10-day PL distribution has a fat tail, then the kurtosis is higher and the value of the risk measure should generally increase. This is the case if we assume that the amount of money invested in the portfolio is the same at every dates. For a random variable X , if it exists, the kurtosis is defined as follows:

$$K(X) = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right], \quad (6.1)$$

where μ is the mean of X and σ is the standard deviation. The correlations over time between the kurtosis and the ES/VaR are presented in Figure 6.3.

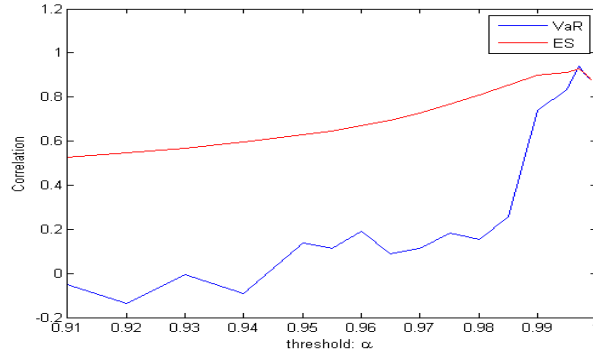


Figure 6.3: Correlation kurtosis/risk measures.

The result is that ES seems more suitable to reflect the tail risk: For thresholds 95% and 96% the correlation is more than 60%, whereas for VaR the correlation is high only for extreme thresholds. Thresholds below 99% have a rather low correlation with the kurtosis, for instance, the 97.5% threshold has a correlation of less than 20% with the kurtosis. Smoothness of the ES correlation comes from the fact that ES is an average of several terms.

6.3 Theoretical Analysis of the Sampled VaR

$\widetilde{\text{VaR}}_{99.99\%}(X, T)$ computed using the sampling algorithm may be very much dependent on large losses of the 1-period PLs because the threshold (99.99%) is extreme. In this section, we analyse this behaviour. Specifically, we want to model and analyse the probability that the i^{th} worst 1-period PLs has to be one (of the 25) composing the $\widetilde{\text{VaR}}_{99.99\%}(X, T)$ when using the sampling algorithm. This is a reliable indicator of the impact that a 1-period PL may have on the sampled VaR.

Note that this problem is different from knowing which PLs compose the scenario of the true $\text{VaR}_{99.99\%}(X, T)$ because for discrete random variables this problem has a deterministic solution: There is only a finite number of possible values for the true $\text{VaR}_{99.99\%}(X, 1y)$ so only one set of 25 PLs composing it. However, this solution is computationally too expensive to determine because there are 10^{50} possible outcomes for the sum of 25 discrete random variables with 1000 possible values. Therefore, we will resort to a numerical approximation.¹ For the historical PLs, this is unlikely because, most of the time, only one combination can give one outcome for the sum (plus its permutations).

Let $X = \{X_1, X_2, \dots, X_n\}$ be identically distributed discrete random variables defined on Ω the discrete set of the PLs. Let $y \in \Omega$, the probability we wish to compute is the following:

$$P(\exists j, -X_j = y \mid -\sum_{i=1}^n X_i = \widetilde{\text{VaR}}_{\alpha}(X, T)), \quad (6.2)$$

¹This problem is equivalent to pick 25 balls without order in a urn containing 1000 balls and put back the balls between the draws. This is because we assume that the PLs cannot permute as in the following example: If $\Omega = \{0, 1, 2, 3\}$, if X and Y are two uniform discrete random variables on Ω then the sum is 4 if $\{X = 1, Y = 3\}$ or $\{X = 2, Y = 2\}$ or $\{X = 3, Y = 1\}$.

where $\widetilde{\text{VaR}}_\alpha(X, T)$ is the sampled VaR returned by the sampling algorithm with 1.000.000 simulations. This is the probability that, when running the sampling algorithm, the PL y appears in a sampled VaR. We simplify the problem by fixing $j = j_0$ and taking $n = 25$, fixing $j = j_0$ is convenient for the continuous case:

$$P(-X_{j_0} = y \mid -\sum_{i=1}^{25} X_i = \widetilde{\text{VaR}}_\alpha(X, T)). \quad (6.3)$$

It represents the probability that a given PL (the one equal to y) has to be the j_0 -draw from the scenario giving the sampled VaR. The expected result is that the larger losses have a higher probability than the higher profits to be part of the sampled VaR. This is due to the fact that the threshold of the sampled VaR is extreme and is equal to 99.99%.

6.3.1 VaR of a sum of continuous random variables

For continuous random variables the setting is different because there are infinitely many PLs and, therefore, the sum of identically distributed random variables is also a continuous random variables. In some cases, the true 1-year VaR can be computed analytically. In that case, the sampling algorithm is not needed because the probabilities of occurrence may be computed. Let X_1, \dots, X_n be n identically distributed continuous random variables. Define $Y = (-X_1 \mid -\sum_{i=1}^n X_i = \text{VaR}_\alpha(X, T))$ and $C = \text{VaR}_\alpha(X, T)$, C is the true 1-year VaR then, using Bayes formula, we have for the continuous case:

$$f_Y(y) = f_{-\sum_{i=1}^n X_i \mid -X_1=y}(C) \cdot \frac{f_{-X_1}(y)}{f_{-\sum_{i=1}^n X_i}(C)}. \quad (6.4)$$

If the random variables X_1, \dots, X_D are independent we obtain:

$$f_Y(y) = f_{-\sum_{i=2}^n X_i}(C + y) \cdot \frac{f_{-X_1}(y)}{f_{-\sum_{i=1}^n X_i}(C)}, \quad (6.5)$$

Therefore, in case of independent and continuous random variables, there is an analytical formula for the density function of X_1 knowing that the sum is the true 1-year VaR.

Example 6.3.1 *For a normal random variable, an analytical solution may be computed, $\text{VaR}_\alpha(X, T)$ is known because the distribution of a sum of independent normal random variables is known: it is again a normal distribution. Therefore, Formula 6.5 may be used directly. The resulting distribution is a density with $\mathbb{E}(Y) = \frac{\text{VaR}_\alpha(X, T)}{25}$, as shown in Figure 6.4.*

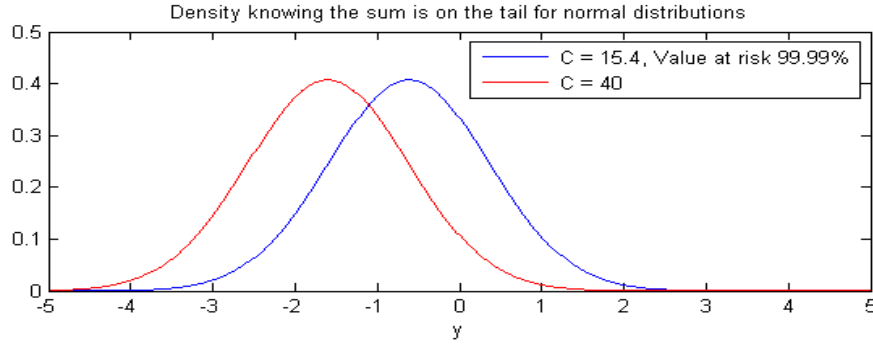


Figure 6.4: Density of X_1 knowing the value of the sum for densities , $C=\text{VaR}_\alpha(X, T)$.

6.3.2 VaR of a sum of discrete random variables

The discrete case and the continuous case are very different due to the drawing effects. Intuitively in the discrete case, if the sample is small enough, the worst PL should be drawn more often on the sampled VaR because the threshold is far in the tail (99.99%).

Example 6.3.2 *In this example, a Bernoulli random variable for which it is possible to compute the distribution of the sum is used. $P(X_1 = 0)$ is plotted versus the value of the sum of 25 Bernoulli random variables (which is a Binomial).*

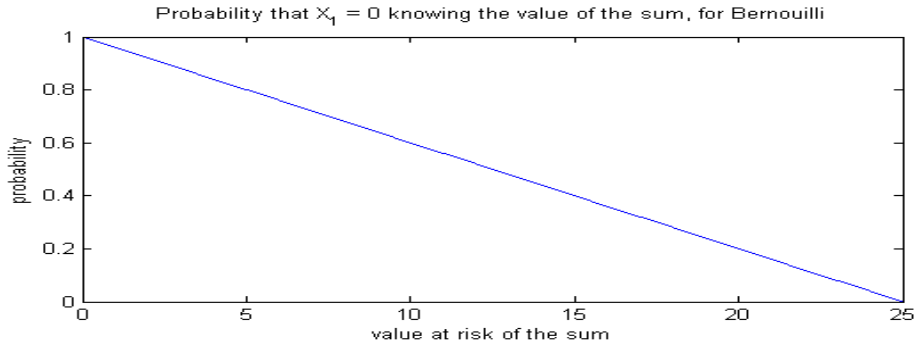


Figure 6.5: Probability of $X_1 = 0$ knowing the value of the sum for Bernoulli random variables.

The result makes sense: To have the sum of 25 Bernoulli random variables equals to 25, only ones should occur. Therefore, the probability that X_1 is 0 has to be 0.

In Figure 6.6, we run the sampling algorithm many times and keep track of the PLs that are in a sampled VaR. This gives the probability that the i^{th} PL has to be in the sampled VaR.

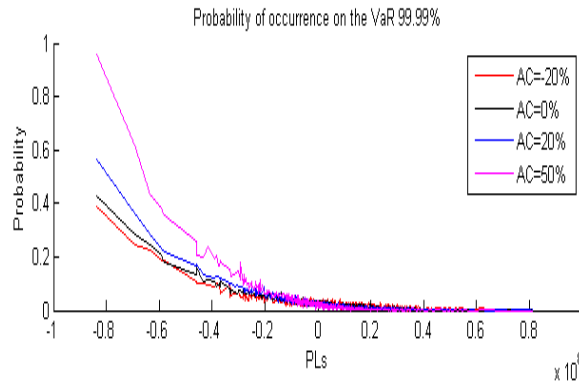


Figure 6.6: Probability of occurrence of a given PLs on the $\widetilde{\text{VaR}}_{99.99\%}(X, T)$.

As expected, the larger losses occur more frequently on the sampled VaR. In Figure 6.6, the worst PL is one of the terms of the sampled VaR with a probability of 50%. Therefore, this PL has a higher impact on the sampled VaR than the 100^{th} worst PL which has a probability of occurrence of about 5%. Some more simulations are done to determine the impact of the autocorrelation on these probabilities. Figure 6.6 shows that a higher autocorrelation in the simulation increases the probability of occurrence of larger losses on the sampled VaR. This is consistent with the fact that the autocorrelation is positively correlated with VaR. Therefore, more occurrences of large losses are needed in order to equal to the sampled VaR. So, the autocorrelation plays an important role in the probabilities of occurrence mainly because it increases the sampled VaR. In case of a lower threshold α the probabilities would be different and it may give a function like in Figure 6.4 (not only decreasing with the PLs).

A conclusion of this section is that, because all the PLs do not contribute equally to the sampled VaR, the risk measure used to compute it should also take in consideration this behaviour. Because ES represents the whole tail, it is more suitable to fulfil this condition.

6.4 Multi Criteria Analysis of the Scaling Factor

This section presents in a mathematical way the criteria that are used to assess the potential risk measures (VaR and ES) and thresholds for the scaling factor. Then, results are explained and analysed in Sections 6.4.1, 6.4.2 and 6.4.3. The first one is about the stability whereas the two others assess the quality of the approximation of the sampled VaR.

In this section, two different portfolios are tested. We call them Portfolio 1 (P1) and Portfolio 2 (P2). P1 is included in P2. P1 includes additional products compared to P2: mainly linear interest rate products. In following pictures, the plain lines are P1 and the dotted lines often represent P2.

6.4.1 Stability of the scaling factor and the Economic Capital

Stability is an important requirement for the scaling factor. If it is known that the scaling factor is relatively portfolio independent, the review of the scaling factor does not need to be done very often. On the contrary, a volatile parameter would lead to unrealistic Economic

Capital figures and frequent reviews. Two types of stabilities through-the-cycle are required: Stability of the market fluctuation risk and stability of the scaling factor. Stability of the market fluctuation risk is a difficult requirement to achieve because, on the other hand, there is the requirement to reflect the tail risk and the tail is likely to be volatile. However for the business, it is convenient to have a relatively stable Economic Capital on short periods of time. Mathematically, the stability is defined as the normalized standard deviation for a given set of observation $Y = (Y_1, \dots, Y_D)$ where Y_i may be the scaling factor or the market fluctuation risk at a given date i and D is the number of observed dates:

$$\text{NSTD}(Y) = \frac{\sqrt{\frac{1}{D-1} \sum_{i=1}^D (Y_i - \bar{Y})^2}}{\bar{Y}}, \quad (6.6)$$

where $\bar{Y} = \frac{1}{D} \sum_{i=1}^D Y_i$. In Figure 6.7, the graph at the top takes Y to be the scaling factor and at the bottom it is the risk measure through-the-cycle.

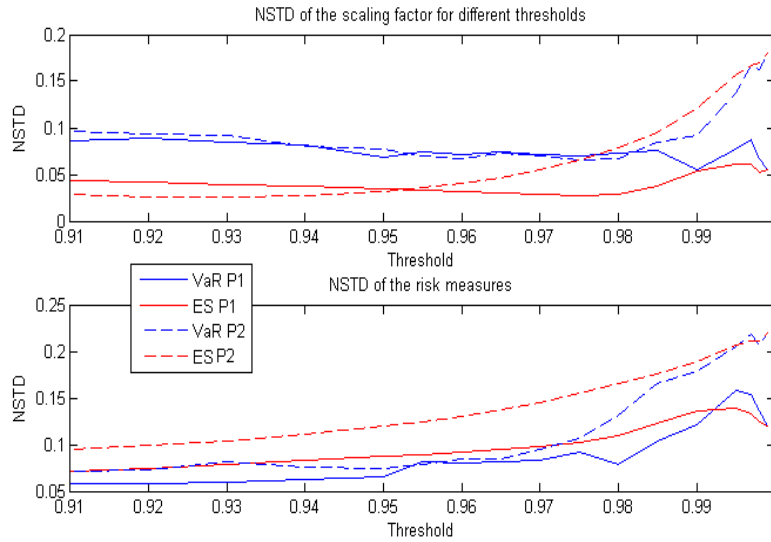


Figure 6.7: NSTD of the scaling factor(top) and of the Market fluctuation risk (bottom) over time.

The NSTD of the scaling factor is influenced by several factors: An appropriated risk measure (that reflects the tail risk) has a lower NSTD for the scaling factor because it implies that the scaling factor is less volatile. A risk measure with a threshold too extreme tends to be more volatile over time because it is significantly affected by small changes in the portfolio. These two factors are opposite, they lead to a minimum at 99% for VaR and 97% for ES for P1 for NSTD of the scaling factor. For P2 the results are different, they lead to a minimum at 93% for ES and 98% for VaR.

For the risk measures (VaR and ES), the NSTD increases with the threshold.

6.4.2 Correlation with the sampled VaR

In this section, an indicator of the quality of the approximation of the sampled VaR is analysed. This is the correlation over time between the sampled VaR and the risk measure of the scaling

factor $\rho(X_k)$ where X_k is a discrete uniformly distributed random variable with support Ω_k containing the 1000 PLs at date k . The scaling factor is not needed here because it does not impact the correlation. We are aiming for a risk measure with a high correlation with the sampled VaR in order to reflect the tail risk accurately. The correlation used is the usual Pearson correlation defined as:

$$r_\rho = \frac{\sum_{k=1}^D (\rho(X_k) - \overline{\rho(X)}) \cdot (\text{VaR}_\alpha(X_k, T) - \overline{\text{VaR}_\alpha(X, T)})}{\sqrt{\sum_{k=1}^D (\rho(X_k) - \overline{\rho(X)})^2 \cdot \sum_{k=1}^D (\text{VaR}_\alpha(X_k, T) - \overline{\text{VaR}_\alpha(X, T)})^2}}, \quad (6.7)$$

where $\overline{\text{VaR}_\alpha(X, T)} = \frac{1}{D} \sum_{k=1}^D \text{VaR}_\alpha(X_k, T)$, $\overline{\rho(X)} = \frac{1}{D} \sum_{k=1}^D \rho(X_k)$ and D is the number of dates.

Simulations are performed using historical PLs of 15 dates (one date per week). The simulations with a normal distribution and a Student's t-distribution use 200 dates with 1000 points randomly drawn. It means that we draw "artificial PLs" with 1000 points from a normal distribution 200 times. Then, the sampling algorithm is used with 1,000,000 simulations. Figure 6.8 presents the results for historical PLs on the left and artificial (from classical distributions) PLs on the right. The correlation is computed for both VaR and ES.

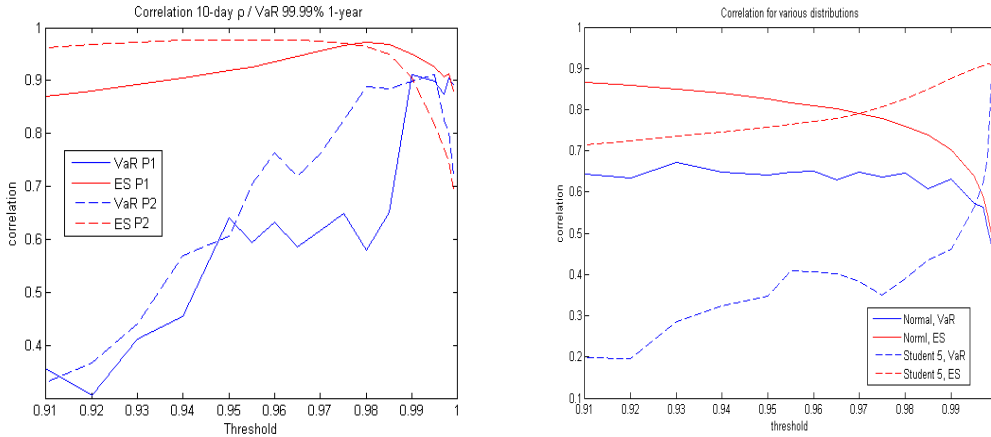


Figure 6.8: Correlation between ρ and VaR 99.99% 1-year for historical PLs (left) and standard distributions (right).

The two portfolios lead to different results, the correlation increases with the threshold, then decreases for extreme thresholds. These maxima are different for the two portfolios. For ES, the correlation decreases for P1 for thresholds greater than 98% and for P2 after 96%. For VaR the correlation increases with the threshold until the 99.5% threshold. It means that larger losses are influencing the sampled VaR more than other PLs, which is coherent with the results from Section 6.3. For a normal distribution, it is different: Severe losses do not impact sampled VaR as much. It confirms that historical PL distribution is closer to a Student's t-distribution. For all probability distributions tested, ES performs better for the correlation criterion than VaR because it always has a higher correlation with the sampled VaR.

The difference observed between P1 and P2 for ES comes from the fact that P1 is included in P2 and the products from P2 that are not in P1 are mainly linear products. Therefore, it

makes the PL distribution to become closer to a normal distribution. As seen on the left of Figure 6.8 for a normal distribution, ES with a high threshold does not have a correlation as high as for a Student's t-distribution.

6.4.3 Error with the sampled VaR

In this section, a final criterion is analysed empirically. It assesses the error to the sampled VaR. To assess the error, we need to fix a scaling factor. Here, the assumption made is that the scaling factor for a risk measure ρ is the average scaling factor over the D dates. Rounding the scaling factor will be discussed in Section 6.5. We are aiming for a risk measure with a low error. Let $(X_k)_{k \in \{1, \dots, D\}}$ be D discrete uniformly distributed random variables with support Ω_k containing the 1000 PLs at date k . Mathematically, the relative error is defined as follows:

$$Er_\rho = \frac{1}{D} \sum_{k=1}^D \frac{|\overline{SF} \cdot \rho(X_k) - \text{VaR}_\alpha(X_k, T)|}{\text{VaR}_\alpha(X_k, T)}, \quad (6.8)$$

where $\overline{SF} = \frac{1}{D} \sum_{i=1}^D SF^k$ and SF^k is the scaling factor at date k .

For the simulation, the same inputs are used as in Section 6.4.2.

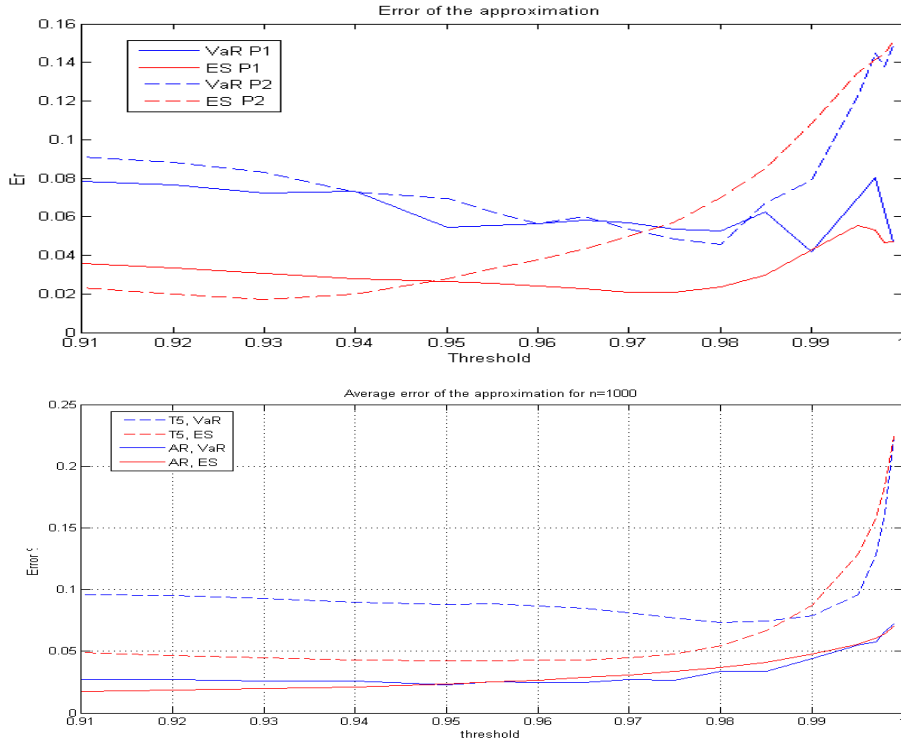


Figure 6.9: Error of the simulation with the sampled VaR.

For usual thresholds (between 95% and 97% for ES and greater than 97% for VaR), ES approximates better the sampled VaR than the 10-day VaR. This is clear for P1, but this also holds for P2 if the threshold of ES is not too extreme.

For historical data, some peaks are in the characterising error functions. They seem rather artificial and data related. For more regular processes, AR and T5, this behaviour is smoother

(see the bottom graph of Figure 6.9). There is also a minimum close to 98% / 99% for VaR. For ES, the minimum is not as obvious. These minima are the consequences of two effects: On one hand, there is a high correlation for extreme thresholds between ρ and the sampled VaR (Figure 6.8). On the other hand, a threshold too high implies an inaccurate estimation of the scaling factor (Figure 6.7). Figure 6.9 is a trade off between the volatility of the risk measure and the accuracy to reflect the tail risk.

After analysing the three quantitative criteria, it is clear that ES is more appropriated to approximate the sampled VaR. A scaling factor based on ES is more stable and reflects better the sampled VaR.

6.5 Rounding and Analysis of the Scaling Factor

In the previous sections of this chapter, the scaling factor and the Economic Capital have been analysed. However, the value of the scaling factor should be rounded (or rounded to nearest half, for instance). For the sake of risk management, this rounding must be conservative to avoid underestimating the Economic Capital.

6.5.1 Analysis of the scaling factor

We start by analysing the values of the scaling factors for both the historical PLs and some standard processes (Figure 6.10). They are scaling factors for the sum of 25 correlated variables. The first remark is that for a normal random variable, with autocorrelation $c\%$, the scaling factor can be computed analytically as it follows: Suppose $(X_i)_{i \in \{1, \dots, n\}}$ are identically normally distributed but not independent with variance σ^2 . Let $\rho_{\alpha'}$ be a risk measure: Either VaR or ES with threshold α' and assume that:

$$\text{cov}(X_i, X_j) = \begin{cases} c \cdot \sigma^2 & \text{if } |i - j| = 1, \\ \sigma^2 & \text{if } |i - j| = 0. \end{cases}$$

Then we can compute the variance:

$$\begin{aligned} \text{var}\left(\sum_{i=1}^n X_i\right) &= \text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right), \\ \text{var}\left(\sum_{i=1}^n X_i\right) &= n\sigma^2 + \sum_{i \neq j}^n \text{cov}(X_i, X_j). \end{aligned}$$

We change the order of summation and then compute the sum., as

$$\begin{aligned} \text{var}\left(\sum_{i=1}^n X_i\right) &= n\sigma^2 + 2\sigma^2 \cdot \sum_{i=1}^n \sum_{j=i+1}^n c^{j-i}, \\ \text{var}\left(\sum_{i=1}^n X_i\right) &= n\sigma^2 + 2\sigma^2 \cdot n \cdot c \sum_{i=1}^n \frac{1 - c^{n-i}}{1 - c}, \\ \text{var}\left(\sum_{i=1}^n X_i\right) &= n\sigma^2 + \sigma^2 \frac{2 \cdot c}{1 - c} \left(n + \frac{1 - c^n}{1 - c}\right). \end{aligned}$$

Therefore, the standard deviation is given by:

$$STD(\sum_{i=1}^n X_i) = \sigma \sqrt{n + \frac{2.c}{1-c} (n + \frac{1-c^n}{1-c})}.$$

Then, we compute the 1-year VaR using Property 4.1.1 and:

$$\text{VaR}_\alpha(X, n\Delta t) = \frac{\text{VaR}_\alpha(X, n\Delta t)}{\rho_{\alpha'}(X_1)} \rho_{\alpha'}(X_1),$$

$$\text{VaR}_\alpha(X, n\Delta t) = \sigma \sqrt{n + \frac{2.c}{1-c} (n + \frac{1-c^n}{1-c})} \frac{N^{-1}(\alpha)}{\rho_{\alpha'}(X_1)} \rho_{\alpha'}(X_1).$$

In case $\rho_{\alpha'} = \text{VaR}_{\alpha'}$ then the scaling factor is given by:

$$SF = \sqrt{n + \frac{2.c}{1-c} (n + \frac{1-c^n}{1-c})} \frac{N^{-1}(\alpha)}{N^{-1}(\alpha')}. \quad (6.9)$$

If $\rho_{\alpha'} = \text{ES}_{\alpha'}$, we have:

$$SF = \sigma \sqrt{n + \frac{2.c}{1-c} (n + \frac{1-c^n}{1-c})} \frac{N^{-1}(\alpha)}{\rho_{\alpha'}(X_1)}. \quad (6.10)$$

Using the fact that $\text{ES}_{\alpha'}(X_1) = \mathbb{E}(X_1) + \frac{f(\text{VaR}_{\alpha'}(X_1))}{1-\alpha'} \sigma$, where f is the density of a normal distribution, gives an analytical formula for the scaling factor of ES for a normal random variable. Then, numerical experiments are performed on historical PLs, normal random variables and Student's t random variables, see Figure 6.10. The computations for the known distributions are done analytically. For the historical PLs of P1 and P2, the range(min-max) is also provided.

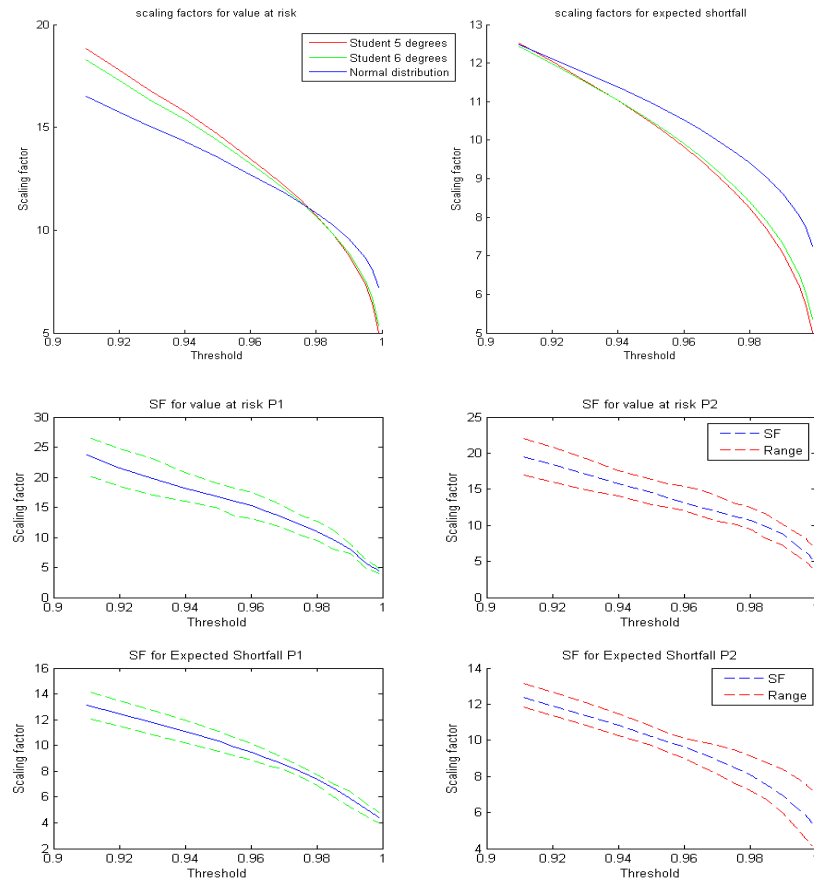


Figure 6.10: Scaling factors for historical PLs(bottom) and for some distributions(top).

The larger portfolio P2 has a scaling factor less dependent on the threshold (for P1 the scaling factor for the VaR is between 25 and 5 for P2 it lies between 20 and 5). It is because the products added are rather linear and create more normality in the PLs. Therefore, the same occurs as when comparing the scaling factors for a Student's t-distribution and a normal distribution.

A scaling factor does not necessarily represent the fatness of the tail. It is linked to the fat-tailness beyond a given threshold. The historical PLs show a higher scaling factor than the classical distributions. This is caused by some very large losses that according to Figure 6.6 have a large impact on the sampling algorithm but are not modelled by classical distributions because these events may be seen as "black swans" (very rare events but with major consequences).

6.5.2 Requirements

This section focuses on a practical requirement of the scaling factor: the rounding. Setting a scaling factor to 12.23 does not really make sense. It may just create a feeling that the scaling factor is very precise for users of the Economic Capital. Whereas, as we saw in Figure 6.10 the scaling factor lies in a certain range. An appropriate risk measure is not volatile, so the rounding should not have a large impact. The rounding should be done conservatively but it should also minimize the error. These two requirements could be in contradiction. In

this analysis, the scaling factor is rounded to the nearest half and to the point. Formally, the following requirements are given:

Let SF^k be the scaling factor at date k , $k \in \{1, \dots, D\}$. Then, in order to be conservative the scaling factor SF should satisfy the following condition:

$$SF \geq SF^{z:D}, \quad (6.11)$$

where $SF^{z:D}$ is the classical order statistics on the scaling factors and $z = \lceil \beta \cdot D \rceil$.

This condition states that the scaling factor should be conservative and be greater than $\beta\%$ of the scaling factors observed.

The second requirement is that the outcome should minimize the error and reflect the sampled VaR. It means that the criteria defined in Section 6.4 should be used again. However, the correlation (Section 6.4.2) is not affected by the scaling factor because it is concerned with the risk measure directly. The two other criteria may be adapted as follows:

$$NSTD(Y) = \frac{\sqrt{\frac{1}{D} \sum_{i=1}^D (SF^i - \widehat{SF})^2}}{\widehat{SF}}, \quad (6.12)$$

where $\widehat{SF} = g(SF^1, \dots, SF^D)$ and g is a function to be determined. However, the function should be determined a priori. As Equation (6.11) must be satisfied we take $\widehat{SF} = SF^{z:D}$, with $z = \lceil 0.9 \cdot D \rceil$. This criteria represents the stability of the scaling factor with respect to a quantile of the historical scaling factor for different time series.

The other criterion, defined in Section 6.4.3, is the error to the sampled VaR defined as follows:

$$Er_\rho = \frac{1}{D} \sum_{k=1}^D \frac{|\widehat{SF} \cdot \rho(X_k) - \text{VaR}_\alpha(X_k, T)|}{\text{VaR}_\alpha(X_k, T)}, \quad (6.13)$$

where $\widehat{SF} = SF^{z:D}$. In the next section, these criteria are analysed numerically.

6.5.3 Quantitative analysis

This section analyses the error committed when the scaling factor is rounded. For simplicity, all properties are summarized in Tables 6.1 and 6.2. The error made when using the rounding should be higher than when the mean is used (Section 6.4). The results for the two portfolios P1 and P2 are analysed.

Table 6.1: NSTD and Er with rounding for P1.

		0.5 rounding			1 rounding		
	Threshold	SF	Er	NSTD	SF	Er	NSTD
VaR	95%	18.0	8.9%	13.3%	18.0	8.7%	10.6%
	96%	16.5	7.5%	11.7%	16.0	7.5%	9.0%
	97%	14.5	7.6%	12.6%	14.0	7.6%	9.4%
	97.50%	13.5	8.8%	14.4%	13.0	8.7%	10.7%
	98%	11.5	7.1%	9.2%	11.0	5.3%	7.1%
	98.50%	10.5	10.7%	13.3%	10.0	7.0%	9.2%
	99%	8.5	7.7%	9.0%	8.0	4.4%	5.7%
	99.50%	6.0	7.4%	16.9%	6.0	7.1%	9.9%
	Worst	4.5	5.1%	15.7%	5.0	14.2%	15.6%
ES	93%	12.0	3.6%	4.5%	12.0	3.7%	4.3%
	94%	11.5	5.2%	6.0%	11.0	1.9%	2.7%
	95%	10.5	3.4%	4.2%	10.0	2.6%	3.0%
	95.50%	10.0	2.7%	3.6%	10.0	2.6%	3.3%
	96.00%	9.5	2.3%	7.5%	9.0	3.9%	4.3%
	96.50%	9.0	2.2%	7.4%	9.0	2.0%	2.6%
	97%	8.5	2.2%	7.6%	9.0	7.3%	7.7%
	98%	7.5	3.7%	10.5%	8.0	10.2%	10.7%
	99%	6.0	5.1%	13.0%	6.0	5.4%	6.8%

Table 6.2: NSTD and Er with rounding for P2.

	Threshold	0.5 rounding			1 rounding		
		SF	Er	NSTD	SF	Er	NSTD
VaR	95%	16.0	13.4%	13.4%	16.0	13.3%	13.3%
	96%	14.0	9.6%	9.6%	14.0	9.5%	9.5%
	97%	12.5	8.6%	6.8%	13.0	11.7%	11.7%
	97.50%	11.5	8.9%	8.9%	12.0	8.9%	8.9%
	98%	11.0	7.4%	8.4%	11.0	10.7%	7.5%
	98.50%	10.5	11.7%	9.1%	10.0	11.9%	9.4%
	99%	9.5	18.3%	18.3%	10.0	18.4%	18.4%
	99.50%	8.0	29.3%	22.9%	8.0	29.5%	23.1%
	Worst	7.0	40.2%	24.3%	7.0	40.3%	40.3%
ES	93%	11.5	6.2%	6.2%	11.0	3.1%	4.3%
	94%	11.0	3.4%	8.0%	11.0	7.1%	3.6%
	95%	10.5	8.4%	4.0%	11.0	8.5%	8.5%
	95.50%	10.5	7.1%	3.9%	10.0	7.3%	4.2%
	96.00%	10.0	10.4%	6.0%	10.0	10.6%	6.3%
	96.50%	10.0	9.5%	9.5%	10.0	9.7%	9.7%
	97%	9.5	13.8%	13.8%	10.0	14.0%	14.0%
	98%	9.0	14.7%	7.8%	9.0	15.0%	15.0%
	99%	8.0	21.8%	21.8%	8.0	28.5%	21.9%

Tables 6.1 and 6.2 show the NSTD and Er, as defined above, for the P1 and P2, respectively, depending on the threshold, the risk measure and the rounding (nearest half or integer). The resulting scaling factor is also given.

For P1, results show that ES performs better on these two criteria. The scaling factor is more stable and the outcome is closer to the sampled VaR. VaR is a very volatile measure of risk and therefore, the quantile of the scaling factor taken on many portfolio snapshots is far from the rounded scaling factor. Because there is a higher dispersion in the scaling factors for VaR, the scaling factors are high.

P2 shows more homogeneous results between VaR and ES, but ES still performs better than VaR. However, due to a larger range of scaling factors for ES (see Figure 6.10), the error and the NSTD are higher.

Overall, VaR is likely to overestimate the risk, otherwise it is not conservative enough. It is also the case that the nearest integer rounding gives a lower error and volatility than the half point rounding for certain thresholds. This is caused by a more beneficial rounding.

For instance, if the un-rounded scaling factor is 13.4 then the half point rounding is 13.5 but the nearest integer is 13.

In case VaR is kept as risk measure, the choice for the threshold would be difficult because for P1 extreme quantiles perform better with this analysis but for P2, it is the opposite.

6.6 Impact Analysis

The impact analysis of the market fluctuation risk is quite straightforward. The scaling factors from Tables 6.1 and 6.2 are used for thresholds that are giving low errors and that are meaningful according to analysis of Section 6.4. The threshold levels are 97.5% and 99% for VaR, and, 95% and 96% for ES. These thresholds are selected because for ES they are close to the point where the error is minimal in the previous sections and the correlation is high. For VaR, 97.5% is a classical threshold and it has a rather low error for P2 (see Table 6.2). For P1, 99% seems appropriate for VaR. A scaling factor based on the worst PLs is also investigated.

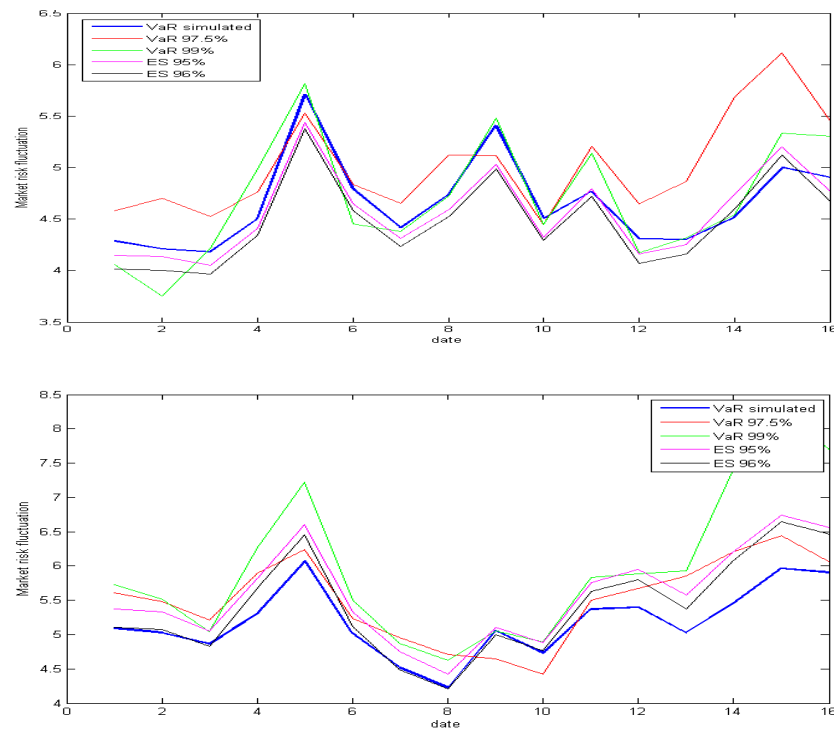


Figure 6.11: Market fluctuation risk for several risk measures with nearest half rounding of the scaling factor for P1 (top) and P2 (bottom).

A market fluctuation risk based on the worst PL (equivalent to a scaling factor based on $\text{VaR}_{99.9\%}$) has a higher amplitude and is more volatile. Due to the conservative rounding, the approximations are often overestimating the risk. Furthermore, it implies a very strong dependency on data and possibly data errors. Figure 6.12, also shows that P2 has a very volatile tail and, therefore, a scaling factor based on it would be extremely volatile.

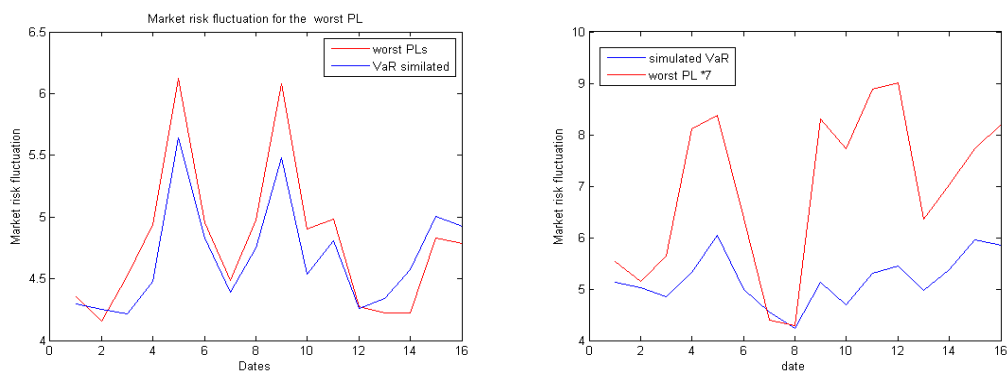


Figure 6.12: Market fluctuation risk for several risk measures based on the worst PL for P1(left) and P2 (right).

6.7 Conclusion and Further Research

This chapter provided a detailed analysis of the method employed to simplify the sampling algorithm. Results showed that ES was more appropriate to scale up to the 1-year sampled VaR. Indeed, it reflected a larger part of the tail so it was likely to reflect a larger part of the risk. Also, ES was more stable because the threshold was lower and we saw that it was the mean of several PLs. About the threshold, we showed that for ES thresholds of 95% or 96% were appropriate and for VaR the choice was more difficult but thresholds from 97.5% to 99% were acceptable. This chapter showed that a change to ES was possible and relevant to capture the risk better.

More changes and advice are expected in the next regulation. In [43], the Basel committee proposes that the use of a unique 10-day liquidity horizon should be changed. The idea is that each book would receive a different liquidity horizon depending on the products involved and on the asset classes of the products in a book. The problem of computing the Economic Capital becomes more difficult because summing 10-day losses with 30-day losses may not really makes sense in certain cases. Furthermore, the correlations between the different asset classes are also harder to compute. In case of such a change the regulator does not consider a capital horizon of one year but a capital horizon which is different for each book and equals the liquidity horizon. In Section 4.1.2, we saw that applying the square root rule directly on the risk measure was not giving similar results because it was underestimating the risk due to financial non-linear derivatives and the fact that the 10-day PLs were overlapping. This new regulation may lead to new research about the market fluctuation risk.

Chapter 7

Incremental Risk Charge and Migration Matrix

In this chapter, some of the most important inputs of the IRC methodology are investigated (see [44]). These are the migration matrix and the probabilities of defaults. A general overview of the IRC was given in Section 4.2. The IRC capitalises for the default risk and the migration risk (credit risks) of non-securitized¹ credit products that are in the trading book. These products may be bonds, structured notes or credit derivatives such as CDSs (Credit Default Swaps). A CDS is an insurance on the default of a non-securitized bond of an issuer. It may be noted that securitized products such as CDOs (Collateralized Debt Obligations²) are excluded from the IRC because they are included in the "SR" of the Economic Capital (see Equation 4.1) that is not in scope of this thesis. The risk of this type of products is analysed employing another methodology according to the regulations for the Regulatory Capital.

The migration matrix and the probabilities of default should satisfy some requirements in order to be robust, well defined and manageable.

First, background information about the IRC methodology is provided in Section 7.1. Then, in the following sections, the migration matrix and the probabilities of default are computed satisfying the requirements of Section 7.1 and the outcome is analysed employing confidence intervals and sensitivity tests on the IRC. A comparison between several methods is also performed and the qualitative advantages of each method are presented. Finally, in Section 7.7, some analysis is performed to obtain a more realistic framework to model the migration and default risks taking in consideration the non-Markovity of the rating process. The idea is that the previous rating should be taken into account.

7.1 Problem Description and Definitions

This section provides some general background information about the rating systems that are employed in this thesis and the methodology that models the IRC.

¹A non-securitized product may be seen as a product with only one underlying asset.

²The formal definition is that a CDO is an investment-grade security backed by a pool of bonds, loans and other assets.

7.1.1 Definitions and rating systems

First, the notion of rating is defined, then, some explanations are given.

Definition 7.1.1 (Rating, [52])

A rating expresses an opinion about the ability and the willingness of an issuer to meet its financial obligations.

A rating is assigned by a rating agency or a bank to companies, countries or products. Since the 2008 crisis, banks are asked to develop and use their own credit rating methodology for the IRC (see [39]). Credit managers of banks often use ratings from several sources: Internal and external (from different agencies).

A first remark is that credit rating systems employed by banks are conceptually different from the ones that are created by rating agencies. While the latter provide a rating to issuers, banks usually assign an expected probability of default. They often develop rating systems based on fixed probabilities of default for the credit risk. These are called internal rating based probabilities of default.

The two systems are very different. In fact, banks assign a probability of default to an issuer and rating agencies provide an opinion about the quality of the debt of the issuer. For instance, a AA rating means "Very strong capacity to meet financial commitments" [52]. Assigning probabilities of default is more convenient for risk management. We now formally define the concept of probability of default:

Definition 7.1.2 (Probability of Default (PD), [28])

A probability of default is the probability that a borrower/issuer will fail to service obligation, leading to bankruptcy.

A parameter that is missing in this definition, is the time horizon of the PD. A 1-day time horizon gives PDs of almost 0 because it is very unlikely that an issuer goes bankrupt within one day. An infinite time horizon gives, very likely, PDs of 1. Therefore, a time horizon has to be set so that the above definition of a PD makes sense.

In the same way, a migration probability is the probability that an issuer changes rating within a certain time horizon. A complete migration matrix is a matrix representation of the migration probabilities and the PDs. It is commonly computed for a 1-year time horizon. However for IRC, the migration matrix and the PDs need to be computed for a 3-month time horizon. This corresponds to the liquidity horizon the is assigned to the type of products (credit and credit derivatives) in the IRC framework. As for the market fluctuation risk, the capital horizon of the IRC is one year.

In this thesis, two rating systems are used:

- The internal rating system. There are 21 ratings (R0 to R20) and a default rating (D). R0 is the best rating and R20 the worst before default.
- Standard and Poors (S&P) rating system. S&P is one of the largest rating agencies and its rating system is widely used. It provides ratings to bond issuers and individual bonds.

Table 7.1 provides these two rating systems, it also gives the 1-year internal rating based PDs.

Table 7.1: Rating systems and internal rating based PDs.

Number	Ratings S&P	Number	Ratings RRR	1-year internal PDs (in Basis points (bp) 1bp=0.01%=0.0001)
1	AAA	0	R0	0
2	AA+	1	R1	1.4
3	AA	2	R2	1.96
4	AA-	3	R3	2.74
5	A+	4	R4	3.84
6	A	5	R5	5.37
7	A-	6	R6	7.52
8	BBB+	7	R7	10.54
9	BBB	8	R8	14.75
10	BBB-	9	R9	21.75
11	BB+	10	R10	33.20
12	BB	11	R11	49.80
13	BB-	12	R12	74.71
14	B+	13	R13	112.06
15	B	14	R14	168.10
16	B-	15	R15	252.15
17	CCC+	16	R16	378.22
18	CCC	17	R17	567.34
19	CCC-	18	R18	851.01
20	CC	19	R19	1,276.52
21	C/CI/R	20	R20	1,914.78
22	SD	21	D	Default
23	D			

By convenience, the ratings are very often numbered as in Table 7.1. Some remarks may be made concerning these ratings: SD is the Selective Default, it is the default on a part of the debt. We assimilate it to a regular default. Ratings from AAA to BBB- are said to be investment ratings. From BB+ to C, they are speculative ratings (ratings worse than CCC+ are highly speculative). R0 is a risk free rating, it is a technical rating that is not used for the IRC and without any impact on the migration matrix. Some notations are introduced, which will be used in the next sections:

- $p_{Ri,Rj}(\Delta t)$ is the migration probability from rating Ri to Rj within a period of time Δt . The simplified notation $p_{i,j}$ is also used when the time horizon and the rating space are immediate from the context.
- $PD_{Ri}(\Delta t)$ is the probability of default of the rating Ri within a time horizon Δt . In the same way, the notation PD_i is also used if the time horizon is implicit. The default state is said to be absorbing because no migration can occur from it.

A complete migration matrix is given by the migration probabilities and the PDs. Consider a rating system with n states plus the default state (D), then, the migration matrix is given by:

$$M = \left(\begin{array}{ccccc|c} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,n} & PD_1 \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,n} & PD_2 \\ p_{3,1} & p_{3,2} & p_{3,3} & \cdots & p_{3,n} & PD_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{n,1} & p_{n,2} & p_{n,3} & \cdots & p_{n,n} & PD_n \\ \hline 0 & 0 & 0 & \cdots & 0 & 1 \end{array} \right)$$

The size of the matrix is $n+1$ columns and $n+1$ rows. The last column contains the PDs and the rest of the matrix contains the migration probabilities (see M). The last row contains the migration probabilities from the default state and, therefore, is filled with 0, except for the probability of staying in the default state. This is because the default state is absorbing. Finally, the diagonal elements represent the probabilities that the rating does not change over the time horizon. These probabilities are usually rather high for the 3-month migration matrix (over 90%). This is due to the fact that ratings are rather stable and are not updated very frequently. The matrix may be separated in four parts: the diagonal elements, the upper triangular matrix (downgrades), the lower triangular matrix (upgrades) and the PDs.

7.1.2 Mapping to external ratings

One major issue when computing PDs or the migration matrix for internal ratings is the limited amount of data that is available. Banks do not own a long history of their internal ratings because the regulation is quite recent. Therefore, it is common to buy the database of a rating agency (like S&P). This database contains the rating history of 16117 issuers (corporates, financial institutions but no sovereign) rated from 1980 until the first quarter of 2013. It implies that a mapping between the two rating systems is needed to compute the internal migration matrix and PDs. The development of this mapping is part of another project and is given for this thesis by [26]. This mapping is provided in Table 7.2.

Table 7.2: Mapping between the S&P and the internal rating systems.

S&P Rating	Internal Rating
AAA	R1
AA+	R2
AA	R3
AA-	R4
A+	R5
A	R6
A-	R7
BBB+	R8
BBB	R9
BBB-	R10
BB+	R11
BB	R13
BB-	R14
B+	R15
B	R17
B-	R18
CCC+	R20
CCC	R20
CCC-	R20
CC	R20
C/CI/R	D
SD	D
D	D

The internal ratings from R1 to R10 are investment ratings and R11 to R20 are speculative ratings. The rating C is assimilated to a default.

7.1.3 Requirements and algorithm of the IRC

This section gives general requirements and assumptions for the IRC, after which, the methodology is provided. Most of the requirements are taken from the Basel document [39]. Because the IRC is used for Regulatory Capital, it has to satisfy many regulatory requirements and it has to be approved by the regulators.

- The IRC must cover the same level of risk as the market fluctuation risk. In our case, the threshold level is set to 99.99% for the Economic Capital. For the Regulatory Capital, the threshold is 99.9%.
- The model should satisfy the constant risk assumption (see Section 2.3.3) with a 1-year capital horizon. At the beginning of each liquidity horizon, the same risk is assumed.
- The liquidity horizon for products included in the IRC is assumed to be 3 months.
- The model should reflect the correlation between the migrations and the defaults.

- The IRC may be diversified with the banking book³ for the Economic Capital (internal requirement allowed by the regulator). This requirement is not discussed in this thesis, as we are interested in the standalone IRC.

Figure 7.1 provides a high level description of the algorithm employed for the IRC.

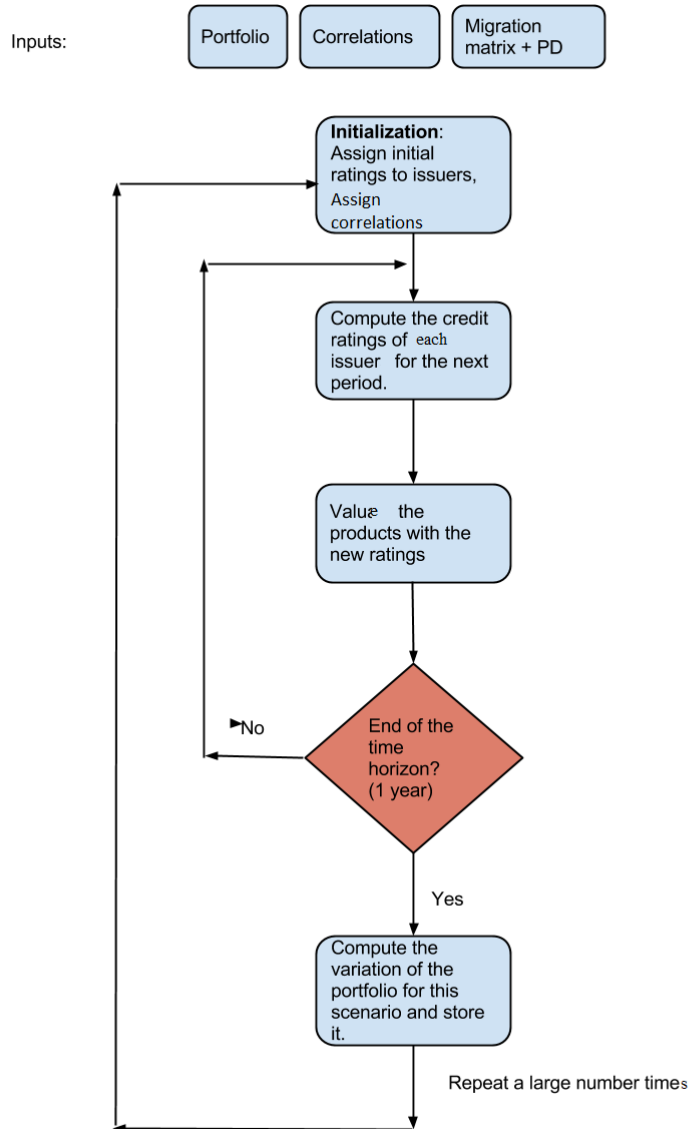


Figure 7.1: Algorithm for the IRC.

As mentioned previously (in Section 4.2), the IRC is a Monte Carlo simulation that employs a Merton credit model (see [21]). For each Monte Carlo scenario, a process describes the net worth of each issuer. At the end of the period, a new rating is assigned depending on the credit value of the issuer. A valuation of the financial products of this issuer is performed. This leads to a loss or a gain. This step is repeated for all issuers. This procedure is repeated for four liquidity horizons, but at the beginning of each period, the risk of the current portfolio is

³The banking book has been defined in Chapter 1 as the products that are placed in the credit risk framework.

assumed to always be the same, and not the risk of the end of the previous period. Then, after four liquidity horizons (one year) the final variation of the portfolio is computed and this gives one scenario. This step is repeated many times.

7.2 Estimation of Migration Matrices

This section provides mathematical background of the methods and concepts that are employed to compute a migration matrix in a financial context. First, the definition of a Markov chain and some properties are provided (Section 7.2.1). Then, the simplest method to compute migration probabilities is investigated (Section 7.2.2). Finally, more advanced methods are introduced: The generator matrix estimation (Section 7.2.3), which estimates continuous-time transitions, and the Aalen-Johansen method (Section 7.2.4), which does not assume time-homogeneity of the rating process. It should be noted that the word "transition" is equivalent to the word "migration". Transition is used in a mathematical context and migration in finance.

7.2.1 Markov chains and definitions

To model a rating process, it is common to assume that a rating process is a Markov chain. Subsequently, some Markov chain theory is explained in this section. Limitations of this theoretical framework for credit ratings are also discussed. First, a definition of a discrete Markov chain is given:

Definition 7.2.1 (Markov chain, [29])

A Markov chain is given by the following elements:

- A finite set of states $S = \{1, 2, \dots, n\}$.
- An initial probability distribution.
- A family of matrices $M(t, t + \Delta t) = (p_{i,j}(t, t + \Delta t))_{i,j \in \{1, \dots, n\}}$, where $p_{i,j}(t, t + \Delta t)$ is the probability that given that the state at time t is i , the state at time $t + \Delta t$ will be j with $\sum_{j=1}^n p_{i,j}(t, t + \Delta t) = 1$, $\forall i \in \{1, \dots, n\}$.

Then, we say that process $R(t)$, defined as the state at time t with migration matrix family $M(t, t + \Delta t)$ where $t \in \mathbb{N}$ is a Markov chain.

The condition that the sum of a row is one ensures that the matrix is well defined. A Markov chain is a process that does not depend on past states but only on the current state. If there is more than one matrix for the whole time-dependent process, this is not convenient for risk management. Therefore, it is common to assume that the rating process is driven by a single migration matrix. This property is called time-homogeneity, which means that matrix M is time independent. A Markov chain $R(t)$ is time-homogeneous if the family of migration matrices associated to this process satisfies: $M(0, \Delta t) = M(t, t + \Delta t) \quad \forall t \in \mathbb{N}$. For classical migration matrices, the set of states is simply the rating space. It implies that the information that is available in the migration matrix only depends on the current rating. In case of a time-homogeneity, the notation of the migration matrix is simplified to $M(\Delta t)$, $M(1)$.

A transition matrix M modelling a time-homogeneous Markov process has the semigroup property (see [29]):

$$M(\Delta t) = M(1)^{\Delta t}. \quad (7.1)$$

We say that G is a generator for $M(1)$ if the following holds:

$$M(1) = \exp(G), \quad (7.2)$$

where $\exp(\cdot)$ is the exponential of a matrix defined as follows:

$$\exp(G) = \sum_{k=0}^{\infty} \frac{G^k}{k!}.$$

Combining Equations (7.1) and (7.2) we obtain for the Δt -time horizon:

$$M(\Delta t) = \exp(\Delta t \cdot G). \quad (7.3)$$

A generator matrix G with elements $(g_{i,j})$ satisfies the following conditions:

$$\begin{aligned} \sum_{j=1}^n g_{i,j} &= 0 \quad \text{for } i = 1, \dots, n, \\ g_{i,j} &\geq 0 \quad \text{for } i \neq j. \end{aligned} \quad (7.4)$$

An important question is to know whether taking a power of the migration matrix with an exponent smaller than 1 makes sense or not. As is it the case for the real number, the square root is not well defined on the complete set of the real numbers. This is also the case for matrices. In fact, the question of existence is almost equivalent to the one to know whether a generator exists. For migration matrix for credit risk, we have a simple criterion to ensure that Equation (7.1) is well defined. A complete analysis of this question and proofs may be found in [23]. This analysis is mainly based on two theorems.

First, we define the following quantity:

$$Z = \max\{(a-1)^2 + b^2 : a + bi \text{ is an eigenvalue of a matrix } M, \text{ and } a, b \in \mathbb{R}\}. \quad (7.5)$$

An eigenvalue λ , associated to an eigenvector v , satisfies $Mv = \lambda v$. This quantity is very similar to the well know spectral radius except that the real part of the eigenvalues is taken minus 1. The maximum is taken over all the eigenvalues of the migration matrix M . We can now state the following theorem:

Theorem 7.2.1 (Convergence of the logarithm series, [23])

Let M be an $n \times n$ Markov migration matrix and I be the identity matrix, and suppose that $Z < 1$. Then, the series

$$\tilde{G} = (M - I) - (M - I)^2/2 + (M - I)^3/3 - (M - I)^4/4 + \dots, \quad (7.6)$$

converges and gives rise to an $n \times n$ matrix \tilde{G} having rows sums 0, such that $\exp(\tilde{G}) = M$ exactly.

A first remark is that Series (7.6) is the Taylor expansion of the natural logarithm applied to the matrix M (i.e. $\log(M)$). However, this theorem does not state that the off-diagonal elements are all positives (as in the definition of a generator in Equation (7.4)). But, it is empirically shown in [23] and [19] that for credit migration matrices if any off-diagonal element is negative then setting is to 0 is an option and it does not significantly change the migration matrix. A second theorem gives a more practical condition on the migration matrix that implies that $Z < 1$, subsequently, Theorem 7.2.1 holds.

Theorem 7.2.2 (Convergence criterion, [23])

Suppose the diagonal entries of a migration matrix M are all greater than 0.5.⁴ Then, $Z < 1$; that is the convergence of Series (7.6) is guaranteed.

In practice, for migration matrices with time horizons short enough (typically less than 1 year), the diagonal elements are greater than 0.5 and the theorem holds. Therefore, the logarithm of matrix M is well defined because Series (7.6) converges. A way to overcome this problem of uniqueness and existence, is to estimate the generator directly and not the migration matrix (see [24]).

We wish to have one migration matrix that is related the other migration matrices (with other time horizons). In Markov theory, it is possible by taking the exponential with a certain coefficient. However, on main assumption should be satisfied: The rating of an issuer should be a Markov chain. It means that the rating of an issuer should depend only on the current rating.

7.2.2 The cohort method

The cohort method is the first method that is presented to compute migration matrices (with PDs included). This method is quite straightforward and is easily understandable. For this method, the set of states is the rating space. The principle is that for each period of time (3-month or 1-year, for instance), the number of migrations from R_i to R_j between the time $t \rightarrow t + \Delta t$ is counted and divided by the number of issuers at rating R_i at time t . Mathematically, we define:

- $N_{i,j}(t)$: The number of companies migrating that are rated R_i at time t and R_j at time $t + \Delta t$.
- $N_i(t)$: The number of companies rated R_i at time t .
- $p_{i,j}(t, t + \Delta t)$: The probability that a company rated R_i at time t is rated R_j at time $t + \Delta t$.

The estimator of the migration probability at time t is given by:

$$\hat{p}_{i,j}(t, t + \Delta t) = \frac{N_{i,j}(t)}{N_i(t)}. \quad (7.7)$$

Under the assumption that a rating process may be represented by a time-homogeneous Markov chain, the maximum likelihood estimator of the migration probabilities is given by:

$$\hat{p}_{i,j} = \frac{\sum_{t=0}^{T-\Delta t} N_{i,j}(t)}{\sum_{t=0}^{T-\Delta t} N_i(t)}, \quad \forall i, h \in \{1, 2, \dots, r\}. \quad (7.8)$$

When summing, it is usual to use overlapping data, which means that the unit of t is in days and that we have $\Delta t = 90$ days (or 91 depending on the day counting convention). Then, n is the total number of days that the sample covers. Therefore, the sum is computed for each day and one migration may be counted several times. The main drawback of this method is that many migration probabilities are 0 whereas it does not mean that such a migration will never happen. This method is used mainly for analysis purposes because it represents the historically observed migration probabilities (see [33]).

⁴This condition is equivalent to strictly diagonally dominant.

7.2.3 Estimation of the generator

Direct estimation of the generator (the GM method) is more convenient from a theoretical point of view. As mentioned in Section 7.2.1, estimating it directly ensures its existence. This method is based on continuous time observations and was introduced in [24]. More theory about this method is provided in [33]. The principle is to estimate the generator matrix of a time-homogeneous Markov chain. The main advantage of this method is that it infers non-zero probabilities for the entire matrix because it does not only use the observed migrations, but every migrations.

The estimation of the generator assumes that the Markov rating process may be modelled by a time-homogeneous Markov chain. Homogeneity means that the transition probabilities depends on the time horizon only and not on the date on which they are estimated. The generator should satisfy the definition of a generator from Equation (7.4). The probability of migration from state Ri to Rj knowing that a migration occurs, is given by $\frac{g_{i,j}}{g_{i,i}}$. Under the assumption that the rating process is time-homogeneous, the maximum likelihood estimator of the generator is given by (see [31]):

$$\hat{g}_{i,j} = \frac{N_{i,j}(T)}{\int_0^T Y_i(s)ds}, \quad \text{for } i \neq j, \quad (7.9)$$

where $N_{i,j}(T)$ is the number of transitions from i to j over the period of observation T , $Y_i(s)$ is the number of issuers with rating i at time s . The values of s and T determine the time horizon of the resulting generator. The denominator may be interpreted as "issuers times time horizon" quantity. If there are two companies rated i for three months and the time horizon is three months then this number is 2.

This method is clearly more efficient than the cohort method in the sense that it uses all data that is available and not just the observed transitions. This method is analysed in Section 7.2.5 and an example is provided.

7.2.4 The Aalen-Johansen method

The Aalen-Johansen Method (AJ) is an extension of the GM method for a time-inhomogeneous rating process. This is necessary because time-homogeneity is hard to prove on the long term. The method has been first introduced in [1]. The default and migration probabilities are clearly affected by economic cycles, for instance. In case of an economic meltdown, an increasing number of negative migrations may happen compared to when the economy is doing well. The same space of states is considered: $S = \{1, 2, \dots, n, D\}$. The previous derivation of the generator does not hold anymore.

The intensities at time t may be written as an integral as follows:

$$g_{i,j}(t) = \int_0^t \theta_{i,j}(s)ds, \quad (7.10)$$

where $\theta_{i,j}(s)$ are the transition intensities. In practice, this integral is computed using a discretization method.

Let $T_{i,j,h}$ be the date when the h^{th} transition from Ri to Rj occurs and $Y_i(s)$ the number

of issuers with rating Ri at time s (just before the migration occurring at time s). Then, we have the following estimator for the intensity matrix elements:

$$\hat{g}_{i,j}(t) = \sum_{h|T_{i,j,h} \leq t} \frac{1}{Y(T_{i,j,h})}, \quad \text{for } i \neq j. \quad (7.11)$$

The family of migration matrix has to be computed based on Equation (7.11). However, it is not possible to compute it using the exponential function because we do not have a generator. Instead, there is a sequence of matrices. However, the exponential function may be approximated using the following definition:

$$\exp(x) = \lim_{n \rightarrow +\infty} \left(1 + \frac{x}{n}\right)^n. \quad (7.12)$$

The limit cannot be taken but we can approximate the exponential. In our case, n is the number of dates where at least one transition happens and $\frac{x}{n}$ represents the generator matrix at a particular date. Therefore, we may define the following estimator for the transition matrix at time t :

$$\hat{M}(0, \Delta t) = \prod_{i=1}^n (I + \Delta \hat{A}(T_i)). \quad (7.13)$$

Here T_i is a transition time of any rating occurring between the times 0 and Δt . $\Delta \hat{A}(T_i)$ is an instant intensity matrix, which is computed as follows:

$$\Delta \hat{A}(T_i) = \begin{pmatrix} -\frac{\Delta N_1(T_i)}{Y_1(T_i)} & \frac{\Delta N_{1,2}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{1,3}(T_i)}{Y_1(T_i)} & \cdots & \frac{\Delta N_{1,n}(T_i)}{Y_1(T_i)} \\ \frac{\Delta N_{2,1}(T_i)}{Y_2(T_i)} & -\frac{\Delta N_2(T_i)}{Y_2(T_i)} & \frac{\Delta N_{2,3}(T_i)}{Y_2(T_i)} & \cdots & \frac{\Delta N_{2,n}(T_i)}{Y_2(T_i)} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \frac{\Delta N_{n,1}(T_i)}{Y_n(T_i)} & \frac{\Delta N_{n,2}(T_i)}{Y_n(T_i)} & \frac{\Delta N_{n,3}(T_i)}{Y_n(T_i)} & \cdots & -\frac{\Delta N_n(T_i)}{Y_n(T_i)} \end{pmatrix}. \quad (7.14)$$

The notations are the same as in Section 7.2.2. This matrix has non-zero elements if at date T_i there is a transition at a rating. In practice, each element of this matrix represents a "daily" intensity rate. It is a cohort method applied on a daily time horizon. The generator method (Section 7.2.3) is for continuous periods of time (infinitely short periods), the cohort method (Section 7.2.2) is based on one period of time. To obtain a matrix for the whole dataset, it is possible to use a weighted average of the different Δt -matrices obtained by this method. The weights are the square root values of the total number of companies rated at the beginning of the period. So, we compute one matrix per day which has a time horizon of Δt and we average all the matrices.

7.2.5 Examples and Comparison

The following is a basic example of the three methods explained above.

Example 7.2.1 *To illustrate the three methods to estimate a migration matrix, we define a rating system with only three ratings A, B and D (default, absorbing). We have one year of history. Suppose five companies are rated A and four are rated B at the beginning. Assume that one transitions from A to B occurs after 6 months and one transition from B to D occurs after*

9 months (not the one that migrated to B). No transition from A to D occurs, then we obtain by the GM method:

$$g_{A,B} = \frac{1}{5 \times 0.5 + 4 \times 0.5} = 0.22, \quad (7.15)$$

$$g_{B,D} = \frac{1}{4 \times 0.5 + 5 \times 0.25 + 4 \times 0.25} = 0.23, \quad (7.16)$$

$$g_{A,D} = 0. \quad (7.17)$$

The remaining elements of the matrix is 0 except for the diagonal elements which are minus the sum of their row, i.e.

$$\hat{G}_{GM} = \begin{pmatrix} -0.22 & 0.22 & 0 \\ 0 & -0.23 & 0.23 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then, the migration matrix is given by:

$$\hat{M}_{GM} = \begin{pmatrix} 0.80 & 0.17 & 0.02 \\ 0 & 0.79 & 0.21 \\ 0 & 0 & 1 \end{pmatrix}.$$

The cohort method gives the following migration matrix:

$$\hat{M}_{CO} = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.75 & 0.25 \\ 0 & 0 & 1 \end{pmatrix}.$$

Within the AJ method, there are only two matrices to compute because there are only two transitions. One migration occurs after 6 months and the second happens after 9 months.

$$\Delta\hat{A}(1/2) = \begin{pmatrix} -1/5 & 1/5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Delta\hat{A}(3/4) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1/5 & 1/5 \\ 0 & 0 & 0 \end{pmatrix}.$$

The migration matrix for the AJ method is then given by:

$$\hat{M}_{AJ} = (I + \Delta\hat{A}(1/2)) \cdot (I + \Delta\hat{A}(3/4)) = \begin{pmatrix} 4/5 & 1/5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4/5 & 1/5 \\ 0 & 0 & 1 \end{pmatrix}. \quad (7.18)$$

Finally, the three matrices are the following:

$$\hat{M}_{AJ} = \begin{pmatrix} 0.80 & 0.16 & 0.04 \\ 0 & 0.80 & 0.2 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{M}_{GM} = \begin{pmatrix} 0.80 & 0.17 & 0.02 \\ 0 & 0.79 & 0.21 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{M}_{CO} = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.75 & 0.25 \\ 0 & 0 & 1 \end{pmatrix}. \quad (7.19)$$

The main advantage of continuous-time estimation is that it infers non-zeros probabilities where there is no observed migration. There are no transitions from A to D but the corresponding migration probability is not equal to 0. The GM and AJ methods model the fact that, when it

is known that a company migrates from R_i to R_j and another company migrates for R_j to R_k then it infers a non-zero probability of migration for R_i to R_k . In addition, the GM method is the only method that captures the fact that there are five companies at the rating B for 3-months but the rest of the time there are only four companies. However, when there are more transitions and if they are "well distributed" in time, the AJ estimator and the GM method give similar matrices.

7.2.6 The rating dynamics

A main assumption is that the rating process itself is a Markov chain. It means that the process does not depend on previous ratings but only on the current rating. The set of states that is considered is $S = \{R1, R2, \dots, Rn, D\}$, where R_i is a rating. Let M be the migration matrix associated to the rating process modelled as a Markov process. This defines the basic framework for the migration matrix. However, despite the fact that the process is assumed to be Markovian, there are several ways to show empirically that the rating process is not Markov but depends on the past ratings, this is only an assumption for modelling. This is shown empirically in a first part.

In [29], the author shows that a time-inhomogeneous Markov chain may be more appropriated. Time-inhomogeneous means that the generator of the migration matrix is not constant but evolves with the time. This is shown in a second part below.

Markov process

A basic way to show that the rating process of a company is not a Markov process is to verify directly the hypothesis that the rating at time $R(t + \Delta t)$ depends only on the current rating $R(t)$. A simple analysis is that we just analyse the probabilities of migration depending on the previous rating. A time horizon of one year is used in Table 7.3. This table presents the 1-year probability of being downgraded/upgraded/no migration knowing the previous rating change.

Table 7.3: Evolution of the ratings.

past change - current change	Downgrade	No change	Upgrade
Downgrade	29%	65%	6%
No change	14%	77%	9%
Upgrade	7%	77%	16%

The conclusion is that the rating process is not a Markov chain. The three rows should be equal in case of a Markov chain because the past ratings should not influence the future changes. There is a momentum (inertia) effect in our case. A downgrade is likely to be followed by another downgrade and an upgrade is more likely to be followed by another upgrade. However, this effect is not taken into account in the current Markov chain. Methods to solve this issue are investigated in Section 7.7.

Time-inhomogeneity

The time-homogeneity is a convenient property of the Markov chains. It makes the estimation easier. Otherwise, several migration matrices depending of the time should be computed and it cannot be assumed, that

$$M(0, \Delta t) = M(t, t + \Delta t), \quad \forall t \in \mathbb{N}.$$

Therefore, one matrix may be assumed and the migration matrix of a different time horizon may be computed using the product rule. This behaviour needs to be investigated. Intuitively, the economic cycles should influence the rating of a company over time. Negative migrations are expected during economic turmoil and positive migrations are expected when the economy is doing well. To detect these characteristics an approximation is to compute the migration matrix over different time periods. We compute PDs using a 5-year window and we roll the window over time. The result is given in Figure 7.2. The average PDs for investment ratings and speculative ratings are also computed. In Figure 7.2, a rolling window of 5 years is used to compute the PDs. The x-axis represents an the end of this window. The conclusion is that the PDs are not very stable. For instance, the log PDs increased (so also the PDs increased) during the credit crisis (which is between 1000-2000).

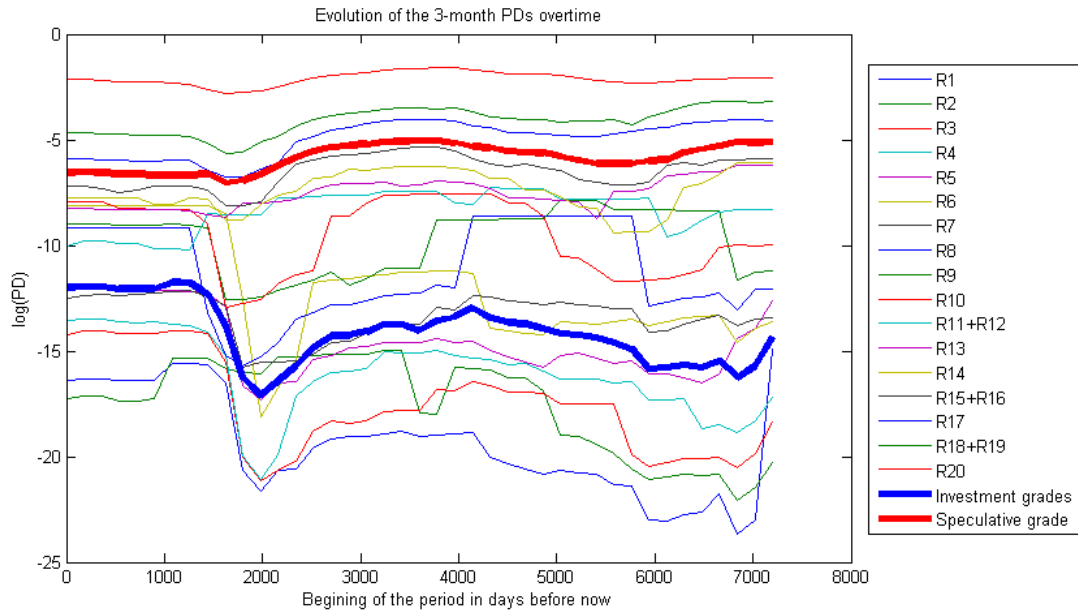


Figure 7.2: Evolution of the PDs.

The characteristics investigated in these sections are not desirable from a modelling point of view. It means that the process we wish to model by a Markov process is not Markovian. Although, most of the systems and the models (Monte Carlo simulations) can only accept a Markovian framework this has to be taken in consideration when modelling. The conclusion of this section is that we showed the existence of a momentum effect. An issuer with a good rating that has been recently downgraded has a type of inertia and, therefore, is likely to be downgraded again. The opposite is also true for highly speculative ratings.

7.3 The Computation of Probabilities of Default

Due to different requirements, the PDs and the rest of the migration matrix are computed in different ways. This section provides the methodology and some analysis for the PDs. First, we discuss the requirements of the 3-month PDs (Section 7.3.1). Then, a methodology is proposed to satisfy these requirements (Section 7.3.2). Finally, numerical experiments are performed (Section 7.3.3).

The specificity of this calibration is that the 1-year PDs are well known (given input) and the methodology to compute the 3-month PDs should be consistent with these 1-year internal PDs.

7.3.1 Requirements

IRC has many specificities because it is a component of the Economic Capital for the trading book capturing credit risk. Therefore, there are two different ways to compute the PDs for IRC, one is market oriented and the second one is based on credit data.

The first method employs market data directly. This approach is discussed in many papers ([53], [18] and [15], for instance) and is very convenient for pricing purposes because it uses real world PDs. Therefore, there is no arbitrage opportunity with the market prices of the products. The general principle is to use products (bonds and CDSs mainly) priced in the market, to imply PDs.

The second alternative consists in using credit data. This method uses the historical migrations to determine the migration probabilities. There are also several methods to estimate them. They have been defined in Section 7.2.1, the cohort, GM and AJ methods.

Although, for the trading book the first option seems more appropriate, it would lead to inconsistency with the Economic Capital of the banking book (credit risk). It means that there would be arbitrage opportunities in terms of capital requirements depending on whether a bond is placed in the market risk (trading book) or credit risk (banking book) framework. Products for which the IRC has to be computed are credit products on the market. Therefore, the book (trading or banking) in which they should be placed is not strict. An approach that does not allow for any capital arbitrage opportunity is preferable. For credit risk, internal rating based PDs are used but they are 1-year PDs and we wish to compute 3-month PDs consistent with them.

Three requirements are given for the 3-month PDs:

- The 3-month PDs should be consistent with the 1-year PDs from the credit risk model (internal rating based PDs). This is the most challenging requirement.
- The PDs should be increasing as the rating becomes worse. Mathematically, this is given by the following conditions:

$$i < j \Rightarrow PD_{\Delta t}(R_i) < PD_{\Delta t}(R_j). \quad (7.20)$$

This requirement makes sense but it cannot be verified historically.

- The last requirement is more abstract: The PDs should be smoothed. This criterion just implies that the increase of the PDs between two consecutive ratings should be rather stable. For the 1-year internal rating based PDs, the ratio $\frac{PD_i}{PD_{i+1}}$ is close to 0.7, $\forall i \in \{1, \dots, 19\}$.

7.3.2 Methodology

Introduction

In this section, the methodology and the implementation of a method satisfying the requirements of Section 7.3.1 are presented. The 1-year internal rating based PDs are given in Table 7.1.

The method uses historical data to extrapolate the 3-month PDs from the 1-year internal rating based PDs.

Assuming that the generator with time horizon one year is called G then, the generator for another time horizon t is given by $t \cdot G$. However, we have reasons to diverge from this equation for our purposes: The most important reason is that we do not have a generator matrix for the 1-year time horizon which is consistent with the internal rating based PDs. The internal rating based PDs are not consistent with historical PDs and there is no internal rating based migration matrix to obtain the generator.

$M(\Delta t) = M(1)^{\Delta t}$, is the migration matrix in the Markov framework (already discussed in Section 7.2.1). However, we do not have a complete migration matrix. Therefore, we need an approximation using only the internal rating based PDs and not the complete migration matrix. First, we consider the simple case where there is no migration but only defaults. Then, the rating process may be seen as many independent life/death processes. For the rating R_i we define the space as $S_i = \{R_i, D\}$, where the two possible states are: stay at the current rating (R_i) or goes into default (D). Let $PD_{R_i}(1y)$ be the probability of going to default over one year. Then, assuming the process is Markovian, the PDs for a different time horizon Δt are given by:

$$PD_{R_i}(\Delta t) = 1 - (1 - PD_{R_i}(1y))^{\Delta t}. \quad (7.21)$$

By Equation (7.21), we wish to take in consideration the migrations that may happen within the period. It means that in a year, an issuer rated R2 may migrate to R10 after 6 months (so its PD increases) and then goes into default 3 months later (so after 9 months). Furthermore, the process is not time-homogeneous. Therefore, the migration probabilities may have a different impact on the 1-year PDs than on the 3-month PDs. The last characteristic that we wish to capture is non-Markovian property. For an investment rating, the momentum effect is likely to lead to an increase of the PDs larger than in the Markovian framework as the time horizon increases. This has consequences for the convexity term structure in Figure 7.3. This figure represents the time term structure of the PDs for all ratings, observed over the last 30 years. For good ratings, the term structure is convex because of the non-Markovity of the process.

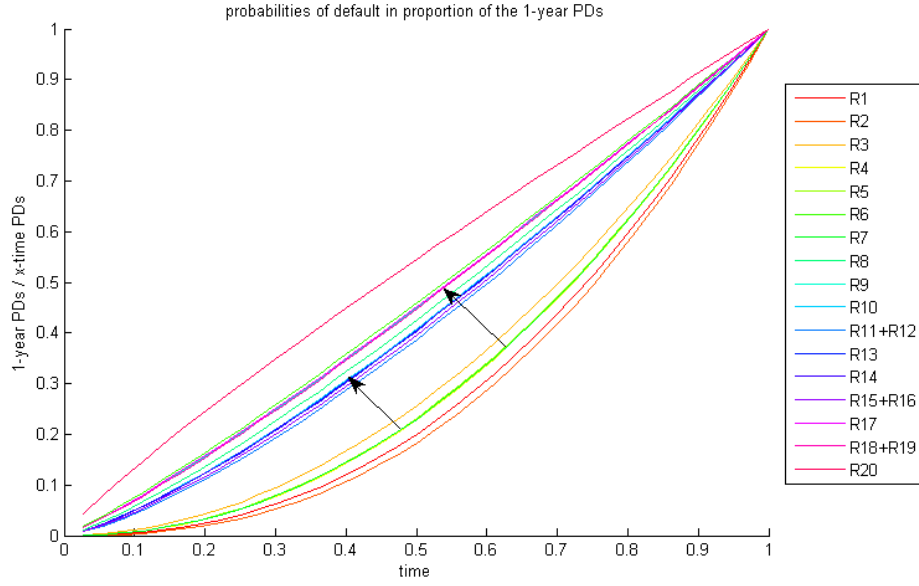


Figure 7.3: Ratio of the $PD_{Ri}(\Delta t)$ to the 1-year PDs.

We now analyse empirically the term structure and ways to take this information into account when extrapolating the 3-month PDs from 1-year PDs. Figure 7.4 compares the two Markov methods we have seen (with just the PDs, Equation (7.21) in red, and with the complete migration matrix, Equation (7.1) in green). The conclusion is that both methods are not accurate to extrapolate the 3-month PDs. Therefore, we need to introduce a parameter to reflect the convexity of the term structure.

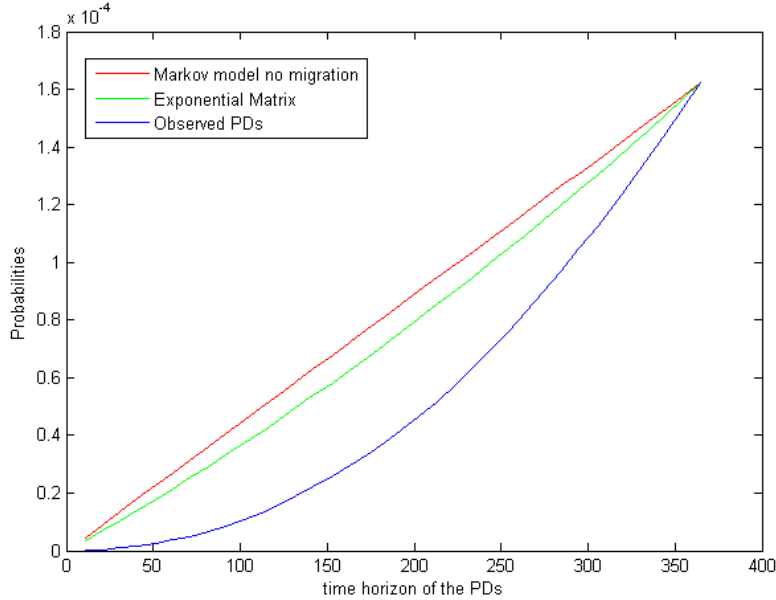


Figure 7.4: R2 term structure.

Subsequently, to reflect this convex term structure, that is not captured by the Markov frame-

works, we introduce a parameter γ_i , for each rating Ri , that take into account the fact that the survival/death PDs do not take in consideration any migrations nor non-Markovity.

$$PD_{Ri}(\Delta t) = 1 - (1 - PD_{Ri}(1y))^{\Delta t^{\gamma_i}},$$

we write it differently, as,

$$PD_{Ri}(\Delta t) = 1 - \exp(\Delta t^{\gamma_i} \log(1 - PD_{Ri}(1y))). \quad (7.22)$$

We modify the equation and then, take the logarithm:

$$\log(1 - PD_{Ri}(\Delta t)) = \Delta t^{\gamma_i} \log(1 - PD_{Ri}(1y)). \quad (7.23)$$

A Taylor expansion is then used on the logarithm because the PDs are usually closed to 0 (at least for ratings that are not highly speculative, but this error is investigated later):

$$PD_{Ri}(\Delta t) \approx \Delta t^{\gamma_i} PD_{Ri}(1y). \quad (7.24)$$

Figure 7.3 also suggests that a scaling by a power of the time function with exponent parameter dependent on the rating may be appropriate because it may be convex or concave, be 0 if the time horizon is 0 and 1 if the time horizon is 1. This figure shows the ratio $\frac{PD(\Delta t)}{PD(1y)}$ when using the AJ method to compute the PDs. With the proposed approximation, it should equal Δt^{γ_i} . Therefore, it measures the impact of the non-Markovity on the PDs. The important fact is that it is characteristic of the PDs that is independent from the value of the PDs. It means that it is possible to extract it from historical data and then, to apply it to the internal rating based PDs.

In most cases, we expect the γ_i to be greater than 1 (except for bad ratings) because the non-Markovity generally decreases the PD of the 3-month time horizon compared to the 3-month PDs extracted from 1-year PDs. The longer the time horizon, the more possibilities an issuer has to migrate to another rating and to go to default.

Example 7.3.1 *We provide an example to show the impact of the migrations. This example does not contains any non-Markovity, it only measures the impact of the migrations on the PDs. Suppose we have a rating system with only three ratings A, B and Default and that the migration matrix M (1-year) is given by:*

$$M = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.75 & 0.15 \\ 0 & 0 & 1 \end{pmatrix}.$$

The 3-month PDs in case of no migrations but, only defaults, are given by $PD_A(3/12) = 1 - (1 - 0.1)^{1/4} = 0.026$ and $PD_B(3/12) = 1 - (1 - 0.15)^{1/4} = 0.0398$.

And the PDs with migrations are given by $PD_A(3/12) = 0.0253$ and $PD_B(3/12) = 0.0405$. The PDs have been computed using the Equation (7.1). The PD of the good rating A increases for the 3-month PD (because the issuer may migrate to B for the 1-year PD) and decreases for B. For B, it is the opposite the PD increases with migration because in case of migration, there is the opportunity to migrate positively to A for 1-year and to decrease the PD.

There is still the question to know whether the approximation of Equation (7.22) by Equation (7.24) is accurate or not. In Table 7.4, the accuracy is tested by computing the error at $t=3/12$ for different 1-year PDs and different γ_i .

Table 7.4: Relative error of the approximation for 3-month PDs.

1-year PD / γ	0.7	1	1.5	2	2.5
0.01%	-0.0031%	- 0.0038 %	-0.0044 %	-0.0047 %	-0.0048%
0.1 %	-0.031%	- 0.038%	-0.044%	-0.047%	-0.048 %
1 %	-0.31%	-0.38%	-0.44%	-0.47%	- 0.49 %
10%	-3.2%	-3.9%	-4.5%	-4.8%	-4.9%
30%	-10.1%	-12.1%	-14.0%	-14.9 %	-15.4 %

Some numbers may seem significant but Equation (7.22) is also an approximation because the corrections with a γ_i terms are not exact. The overall error is investigated in the next section. The relative error is always negative due to the Taylor expansion of order 1. The main task is to determine the γ_i consistently and accurately for all i . They are determined based on the GM method and the AJ method. The historical PDs are computed for various time horizons and, then, a linear regression is performed in the log space.

Equation (7.24) becomes:

$$\log(PD_{Ri}(\Delta t)) = \gamma_i \log(\Delta t) + PD_{Ri}(1y). \quad (7.25)$$

The results for different Δt values are used and, a linear regression is applied. For Equation (7.23) the following is used:

$$\log |\log(1 - PD_{Ri}(\Delta t))| = \gamma_i \log(\Delta t) + \log |\log(1 - PD_{Ri}(1y))|. \quad (7.26)$$

The absolute values are necessary because $\log(1 - PD_{Ri}(1y))$ and $\log(1 - PD_{Ri}(\Delta t))$ are negative quantities.

Computation of γ_i , $i = 1, \dots, n$

In this paragraph, the computation of γ_i , $i = 1, \dots, n$ is explained.

The method involves the use of the S&P dataset. The computation uses a least squares approximation of each γ_i based on historical PDs. The PDs are computed for several time horizons using historical data. Then, γ_i is determined, so that it minimizes the square error of Equations (7.23) and (7.24). We investigate the two (AJ and GM) methods to compute γ_i $i = 1, \dots, n$. The cohort method is not used because it gives many zero PDs. This implies that the corresponding γ_i is not defined.

We analyse several methods in this paragraph: The AJ and GM on one hand, and, on the other hand, which of the two formulations, i.e. Equations (7.23) and (7.24), is the most suitable. We show that all methods give very similar results.

The first point of interest is which method of AJ or the GM should be used to compute γ_i , $i = 1, \dots, n$. The results are very similar due to the large dataset that is available (30

years of S&P rating history). Figure 7.5 gives a scatter plot of γ_i , $i = 1, \dots, n$, given by AJ and the GM methods. The outcome for the PDs is very close to a straight line, so the two methods are almost equivalent. This is not surprising because despite the fact that the process is not time-homogeneous, the dataset is very large and compensates for this feature. There are many defaults that occur at different times, so the two methods are very similar.

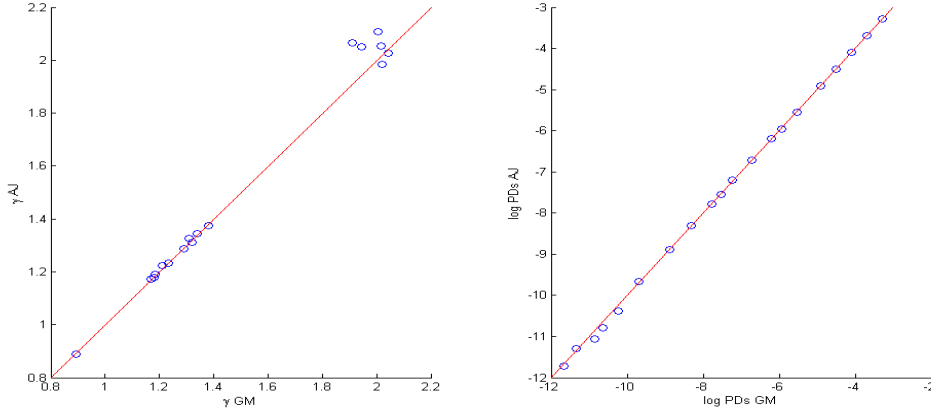


Figure 7.5: Comparison of GM and AJ methods to extract the PDs.

The other point to investigate is whether approximation by a Taylor expansion of Equation (7.24) is accurate or not. To be accurate the 1-year PDs should be close to 0. However, for highly speculative ratings this is not the case as it may be seen in Table 7.1. Some PDs are as high as 20%. Results are analysed in terms of γ_i , $i = 1, \dots, 17$, and of PDs. We test the difference between the two methods to extract the 3-month PDs.

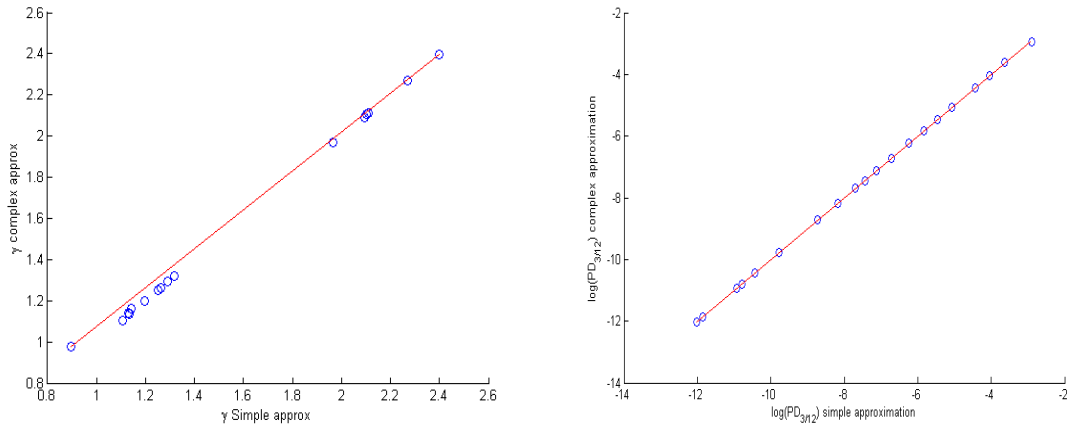


Figure 7.6: Scatter plot γ_i , $i = 1, \dots, 17$, (left) and scatter plot $\log(PD_{3/12})$ (right) computed with Equations (7.23) and (7.24).

Although, the two γ_i from the two approximations are a bit different, see Figure 7.6, the PDs are almost equal. The reason is simply that the two equations are different approximations so they have different γ_i but very close PDs because their, respective, γ_i are optimized by linear regression. Therefore, the simplest method is used, Equation (7.24).

As mentioned previously the γ_i , $i = 1, \dots, 17$, are obtained by employing a linear regression of the log PDs in the log time space of several PDs. Figure 7.7 shows this regression. Results are accurate as the error is rather low: The average error between the 3-month observed log PDs and the regression is only 1.6%.

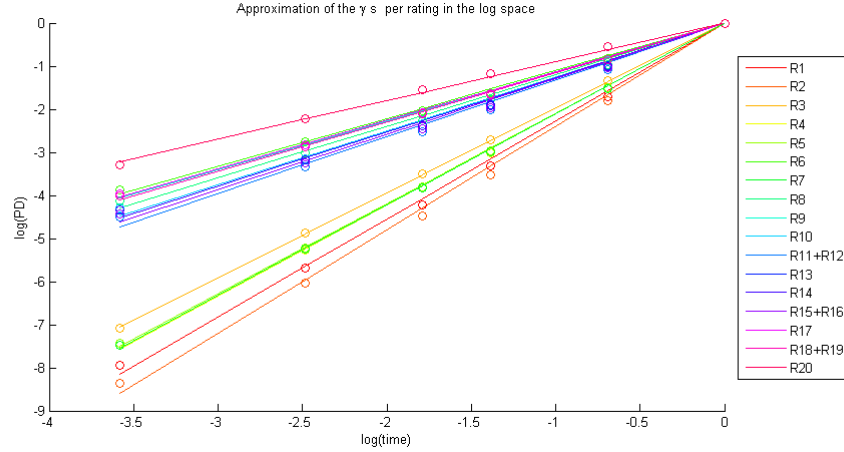


Figure 7.7: Linear regression on the γ_i , $i = 1, \dots, 17$, in log space.

The interpretation is that γ_i represents the convexity/concavity of the term structure may be seen in Figure 7.3. A γ_i far from 1 implies that the non-Markov migrations have a large impact on the PDs for different time horizons. In Figure 7.7, one may remark that the γ_i , $i = 1, \dots, n$ are divided in two groups. One group is around 2 and the other one is closer to 1. The reason is that the parameters γ_i , $i = 1, \dots, n$ are dependent on observed defaults. If a transition occurred in the history then it has a large impact on the PDs for all time horizons used for the linear regression. This case leads to a low corresponding γ_i because the direct default is counted for all time horizons. On the other hand, if no direct migration occurred from R_i but some companies rated R_i migrated to R_j and other companies went to default from R_j , the 10-day PD does not capture this default (for R_i) but the 1-year PD does. In addition, the non-Markovity of the rating process makes that a company with an investment rating that is downgraded is more likely to go to default than a stable company at the downgraded rating. Stated differently, if we assume that we have a vector of PDs, a company rated R1 and downgraded to R6 has a higher probability of default than a company that has been rated R6 for 2 years. As only downgrades may occur from R1, this effect is very significant. Therefore, this effect creates convexity in the term structure and the corresponding γ_i is close to 2.

For rating between the R7 and R15, the non-Markovity is still present, but it may not be captured by the γ_i simply because it may happen for both upgrades and downgrades. Then, the γ_i are usually above 1, $i = 1, \dots, n$. For bad ratings, the effect is opposite and there is a non-Markov effect when being upgraded. This leads to γ_i below 1.

Smoothing the γ_i , $i = 1, \dots, n$

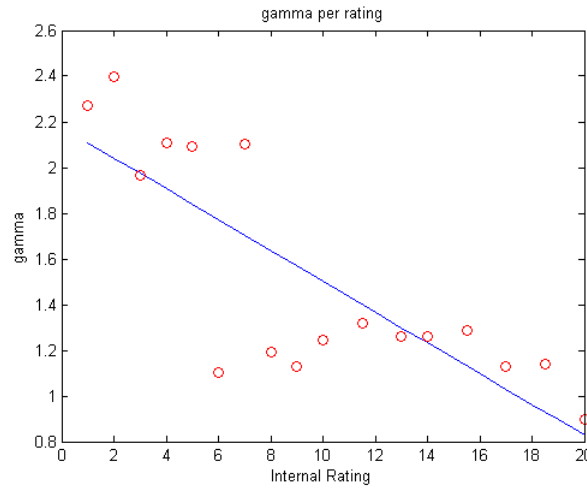
A smoothing between the γ_i , $i = 1, \dots, n$ is performed to avoid large jumps in the PDs. Another economical explanation is that the γ_i is very much driven by some defaults. In the dataset, the A rating has some defaults (see Figure 7.8). This is driving γ_i to a low value. It may be

interpreted that, the fact that there is a direct default at this rating but not at ratings A- of A+ is random. Smoothing diminishes the randomness and the data's imperfections. Although the smoothing does not seem accurate from a mathematical point of view because the error between the observed γ_i and the interpolated one is large, it is meaningful from an economic point of view.

At this stage as shown in Figure 7.3, we only have 17 ratings because several ratings have been merged in order to map the RRR/S&P rating (see Table 7.2). Therefore, the following procedure from Table 7.5 is employed. The first line contains the steps of the procedure and below it is the rating system for this step.

Table 7.5: Table of the mappings during the computation of the γ_i , $i = 1, \dots, n$.

Rating system of the data \rightarrow	Extract according to the following mapping. Then, compute the γ_i , $i = 1, \dots, n \rightarrow$	Map the γ_i , $i = 1, \dots, n$ to the following point then perform a linear regression \rightarrow	From the linear regression obtained, extract the following points \rightarrow	Map them to the internal rating corresponding
AAA	R1	1	1	R1
AA+	R2	2	2	R2
AA	R3	3	3	R3
AA-	R4	4	4	R4
A+	R5	5	5	R5
A	R6	6	6	R6
A-	R7	7	7	R7
BBB+	R8	8	8	R8
BBB	R9	9	9	R9
BBB-	R10	10	10	R10
BB+	R11 + R12	11.5	11	R12
BB	R13	13	12	R13
BB-	R14	14	13	R14
B+	R15 + R16	15.5	14	R15
B	R17	17	15	R16
B-	R18 + R19	18.5	16	R16
CCC+	R20	20	17	R17
CCC	R20		18	R18
CCC-	R20		19	R19
CC	R20		20	R20
C/CI/R	R20			
SD	D			
D	D			

Figure 7.8: Linear interpolated γ_i , $i = 1, \dots, 20$.

The final γ_i , $i = 1, \dots, 20$ resulting from the procedure are given Figure 7.8.

7.3.3 Resulting probabilities of default

After the steps described in the previous sections, we are able to compute the 3-month PDs for each rating Ri extrapolated from the 1-year internal PDs using the term structures $(\gamma_i, 1, \dots, 20)$. The results are presented in Table 7.6. The term structure is also provided.

Table 7.6: Probabilities of defaults (in basis points).

R1	0.07	R11	6.60
R2	0.12	R12	10.79
R3	0.18	R13	17.62
R4	0.28	R14	28.79
R5	0.42	R15	47.02
R6	0.65	R16	76.79
R7	0.99	R17	125.41
R8	1.51	R18	204.82
R9	2.43	R19	334.51
R10	4.04	R20	546.32

Smoothing the γ_i , $i = 1, \dots, 20$ also smooths the term structure. Compared to Figure 7.3, the term structures depending on the rating exhibit more regular behaviours. It implies that there is no jump in the PDs and in the term structures. This is shown in Figure 7.9.

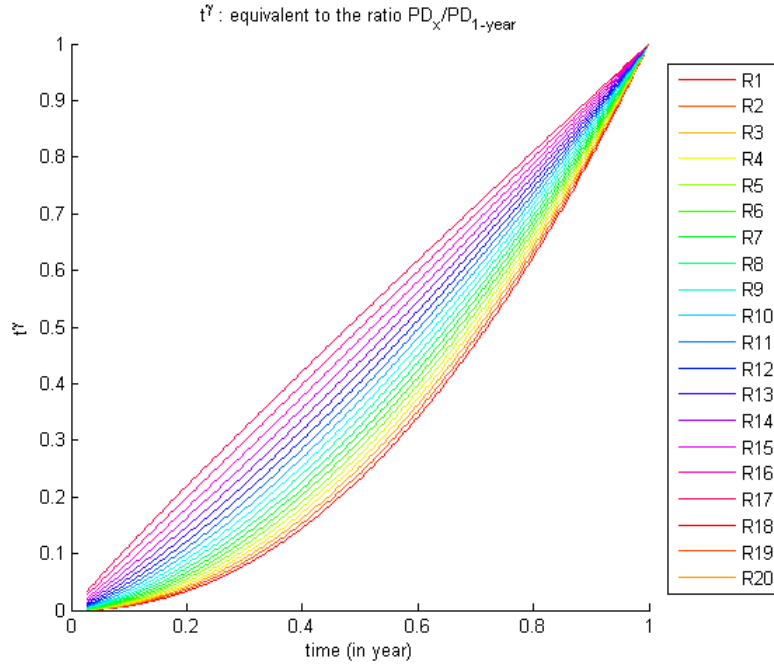


Figure 7.9: Term structures of the probabilities of default.

Ratings better than R17 have convex term structures reflecting the fact that the longer the time horizon the more likely an issuer may move to a bad rating and then go into default.

Therefore, all our requirements are satisfied: Using historical data, we obtain 3-month PDs consistent with the 1-year PDs and smoothed PDs for the internal rating system.

7.3.4 Historical Probabilities of Default

In this section, we analyse whether the internal PDs are conservative compared to historical PDs or not. Internal PDs are given by the rating model and are supposed to be fixed through the cycle. This analysis is also useful for the next section where the rest of the migration matrix is computed with the AJ method and a mismatch between the PDs may have an impact on rest of the migration probabilities. The analysis is performed for both the 1-year PDs and for the 3-month PDs.

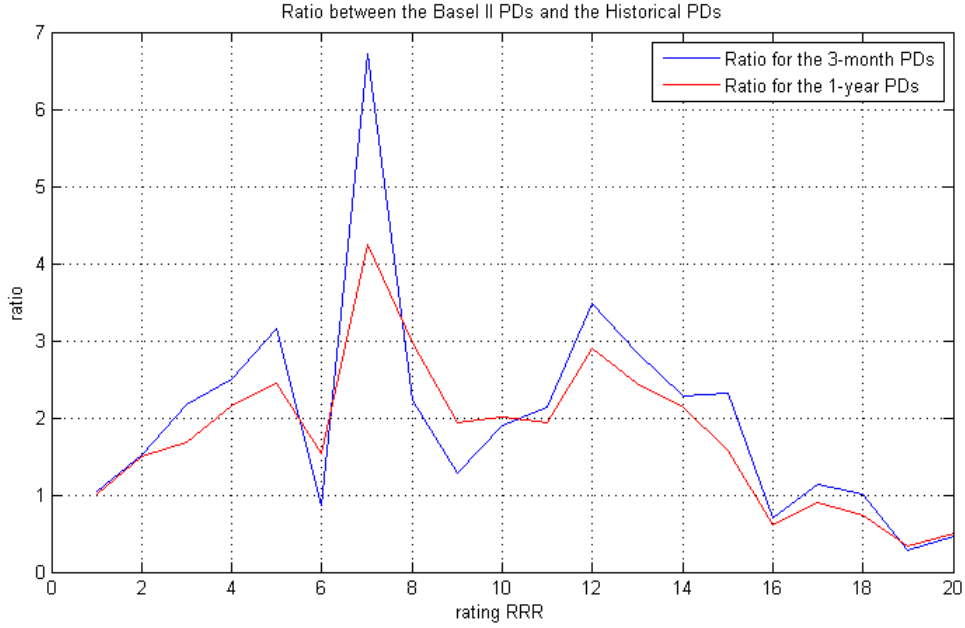


Figure 7.10: Ratio between extracted PDs from internal rating based PDs and historical 3-month PDs.

Figure 7.10 presents the ratio of the internal PDs to the historical PDs for two time horizons. The 3-month internal PDs are computed employing the algorithm described in the previous sections while the 1-year internal PDs are given and the historical PDs are simply the results of the AJ method. The result is that internal rating based PDs are overestimating the historical PDs for ratings from R1 to R15 and underestimating them for the five worst ratings. Exposure to the worst ratings is usually rather small. The final result is an overestimation of the risk. Internal PDs are typically very conservative.

7.4 Computation of the Migration Matrix

In this section, we investigate the computation of the remaining of the migration matrix, the migration probabilities and the probabilities of not migrating (not the PDs). The methodology is decomposed into two main steps: First, a migration matrix is computed using one of the methods of Section 7.2.1. This step is explained in Section 7.4.2. Then, a regularization of the matrix is performed to satisfy the smoothness requirement (Section 7.4.3). First, a general introduction and some requirements are given (Section 7.4.1).

7.4.1 Introduction and requirements

The computation of the migration probabilities is more straightforward than the computation of the PDs. It is mainly due to the fact that we do not need to be consistent with the credit risk framework because the migration matrix is not explicitly used to compute the Economic Capital for the banking book.

Unlike the PDs, it is not possible to compute market implied migration probabilities. This is because ratings are not as objective as a defaults. A downgrade may have no impact on the

face value of a bond. Further, many different rating systems are used by banks and agencies. Therefore, only the credit method using historical migrations is employed.

The matrix is computed with the methods from Section 7.2.1, but we also define some smoothness requirements:

- The migration probabilities should be decreasing from the diagonal, which is equivalent to the following for the matrix $M = (p_{i,j})_{i,j \in \{1, \dots, n\}}$:

$$\begin{aligned} p_{i,j} &\geq p_{i,k}, & \forall i, j, k \text{ such that } i < j < k, \\ p_{i,j} &\leq p_{i,k}, & \forall i, j, k \text{ such that } j < k < i. \end{aligned}$$

This condition means that we expect that a migration of a large number of ratings is not as likely as a small migration.

- The second requirement is that the migration matrix should be smoothed. Similarly to the PDs, this requirement may be vague but it is just to avoid too much reliability on the data. This requirement will make more sense in the next section where Figure 7.11 is analysed, this is the migration matrix obtained directly from the dataset and this is not smoothed due to the lack of data.

7.4.2 Estimation of the matrix

The computation of the migration probabilities is rather simple. The AJ method is employed to use the information in a better way than the cohort method and to take in consideration the time-inhomogeneity. Here, the GM and AJ methods are compared. In this section, the cohort method is not used because it has been proved less efficient as it gives many zero migration probabilities (see [24], for instance). Figure 7.11 presents the resulting matrices when using the two methods (GM and AJ). The difference between the matrices is very small.

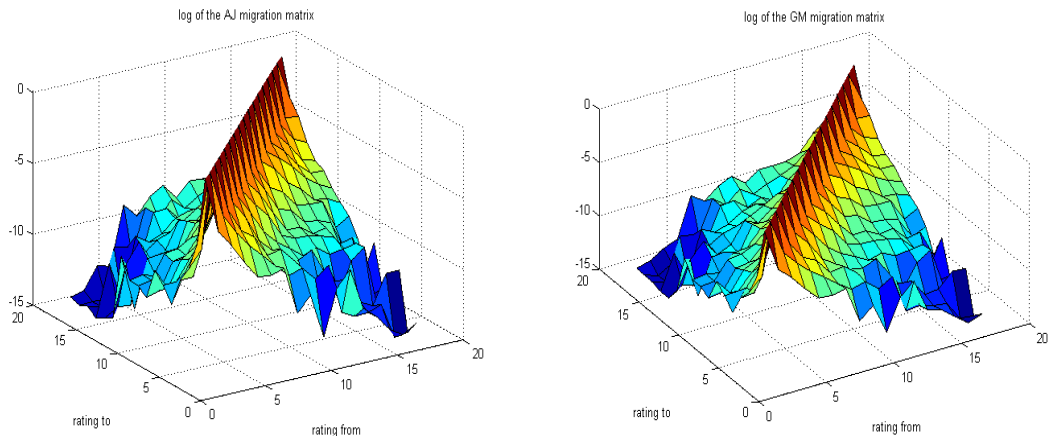


Figure 7.11: Migration matrices computed with AJ and GM methods.

Measuring differences between matrices is not totally trivial. Many norms and measures have been proposed and they are less intuitive than for numbers or vectors. We simply use the

max-norm and the average difference. The max-norm is represented by the largest element of a matrix (in absolute value). The average difference is computing the average of the individual differences of the two matrices. We denote the 3-month AJ migration matrix by M_{AJ} and the GM migration matrix by M_{GM} . The relative maximum difference is given by:

$$\max \left(\frac{M_{AJ} - M_{GM}}{M_{GM}} \right) = 5.8\%,$$

and the relative average error is given by:

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{|(M_{AJ})_{ij} - (M_{GM})_{ij}|}{(M_{GM})_{ij}} \right) = 0.3\%,$$

where n is the number of rows or columns of the matrix. The difference between the two matrices is not significant.

This matrix elements are not sufficiently smooth. The requirements of Section 7.4.1 are not satisfied. We continue the analysis with the AJ method, as mentioned.

7.4.3 Regularization

As mentioned in the previous section, the matrix obtained from any of the method is not smoothed. Therefore, we perform a smoothing of the migration matrix. Some non-parametric methods that may be employed directly on the generator are outlined in [30]. However, in the literature, it is very common to find migration matrices restricted to 6-7 ratings (no +/- variant for S&P). Therefore, there is often enough data and the matrix is easy to smooth non-parametrically. However, we consider the full migration matrix and the regularization step is more complex. For transparency, we wish to implement a parametric smoothing. This section also extends the rating space: For now, there are only 17 ratings and we aim for 20 ratings.

It is decided to employ a parametric smoothing, therefore, the first step is to choose and define a smooth function $f : [1, n]^2 \rightarrow \mathbb{R}$, where n is the number of ratings after the smoothing procedure. The first coordinate is the rating from and the second one the rating to. We now define what a smooth function is, for our purposes. It should hold that $f \in Q$, where Q is the following function space:

$$Q = \begin{cases} f \in C^\infty(I), \\ f \in C^\infty(J), \end{cases} \quad (7.27)$$

where $I = \{(x, y) \mid (x, y) \in [1, n]^2 \text{ and } x \neq y\}$ and $J = \{(x, y) \mid (x, y) \in [1, n]^2 \text{ and } x = y\}$. $C^\infty(\cdot)$ is the space of infinitely differentiable functions on a given domain (here, I or J). The space on which the function f should be defined is $[1 : n]^2$. However, no smoothness condition is required for $x = y$ because these are the probabilities of no migration and we saw, in Section 7.4.1, that the migration matrix should be decreasing from the diagonal. In terms of migration matrix, I represents the off-diagonal elements and J represents the diagonal elements. Furthermore, f should approximate M_{AJ} , where M_{AJ} is the 17×17 migration matrix given by AJ method. We wish to find f such that:

$$f = \arg \min_g \{\|g - M_{AJ}\|, g \in Q\}, \quad (7.28)$$

where $\|\cdot\|$ is a norm to be determined, that evaluates g at the points of column three of Table 7.5 such that it is a 17×17 matrix and the difference $g - M_{AJ}$ makes sense. Let u be the vector of the third column from Table 7.5 and v the ratings where we wish to obtain migration probabilities. In our case, we take $v = (1, 2, \dots, 20)$.

Having defined the general mathematical framework, we divide the smoothing procedure in two steps: First, the function f is defined on J to smooth the diagonal elements. Then, the function f is determined on I , it employs a 2-dimensional polynomial on the upper triangular matrix and the lower triangular matrix in the log-space.

Diagonal elements

The first step of the smoothing procedure is to smooth the diagonal elements (i.e. the probability that the rating does not change). The smoothing is a linear smoothing. For this step, we need to define the function f on J . The general idea is that a linear regression is performed in the log space. Therefore, f may be defined as follows:

$$f(x, y) = \exp(a \cdot x + b), \quad (x, y) \in J, \quad (7.29)$$

where a, b are constant determined by linear regression at the points given by the elements of vector u . The norm employed is the 2-matrix-norm.⁵ Therefore, the minimisation problem for the diagonal is equivalent to

$$(a, b) = \arg \min_{c, d} \{\|c \cdot u + d - \log(\text{diag}(M_{AJ}))\|_2\}, \quad (7.30)$$

where diag of a matrix is the diagonal elements in form of a vector (similar to the Matlab function `diag`). Although in Figure 7.12 the smoothing seems to be a straight line, it is curved due to the fact that the linear regression is performed in the log-probability space.

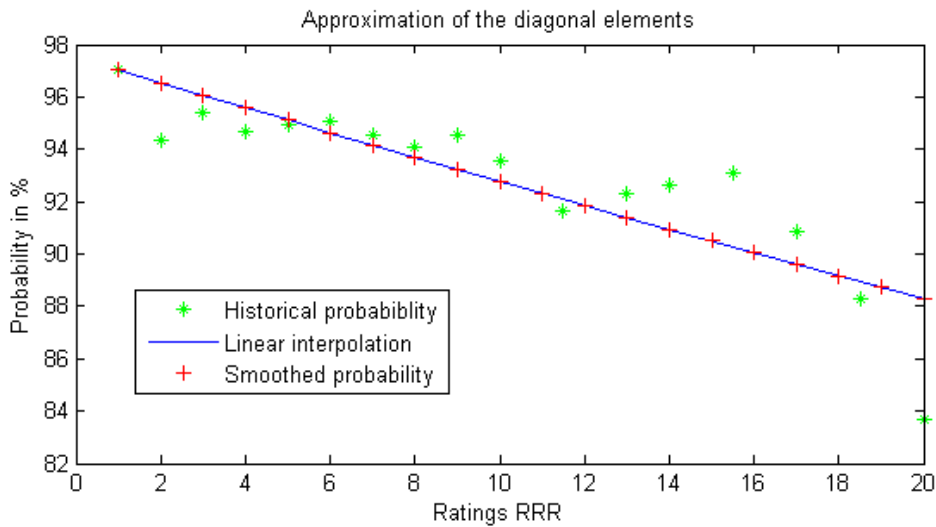


Figure 7.12: Smoothing of the diagonal elements.

⁵The 2-norm of a vector v is defined as $\|v\|_2 = \sqrt{v' \cdot v}$.

The diagonal elements of the final migration matrix are given by $\exp(a \cdot v + b)$, where v is the vector defined above. The average (absolute) error committed is less than 1% and the maximum error only 2% in absolute values, if R20 is excluded. Therefore, the error is acceptable for these diagonal elements.

Off-diagonal elements

The off-diagonal elements are smoothed after the diagonal elements. The matrix is only 17×17 and has to be extended to 20×20 . We also need to satisfy the requirement that the probabilities should be decreasing from the diagonal (see Section 7.4.1). The same strategy is used as for the diagonal elements: The 17 current points are mapped to the elements of vector u , then, the smoothing is performed and the migration matrix for 20 ratings is extracted from the continuous smooth function. The two sides (upper triangular matrix and the lower triangular matrix) of the diagonal are then smoothed independently using two different functions. Some final adjustments are made after this step to have a probability mass equals to 1 (per rows).

The first step is to divide the domain I into two parts I_U and I_L , where $I_U = I \cap \{x < y\}$ and $I_L = I \cap \{x > y\}$. These define the strictly upper and lower triangular matrices, respectively. The smoothing employs a continuous function on each of the two subsets: a 2-dimensional parabola in the log-probability space. A 2-dimensional parabola permits to have a simple and rather accurate calibration and, it is understandable and transparent for the users. We treat the case of I_U , the lower triangular case I_L is similar. We now define the smooth function f on I_U

$$f(x, y) = \exp(Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F), \quad \forall (x, y) \in I_U, \quad (7.31)$$

where A, B, C, D, E , are constant real numbers to be determined. First, we modify the equation such that the diagonal (J) is an axis and is equal to 0. A rotation is performed by changing the x-axis to $y - x$:

$$f(y - x, y) = \exp(A(y - x)^2 + 2B(y - x)y + Cy^2 + D(y - x) + Ey + F), \quad \forall (x, y) \in [1, n]^2. \quad (7.32)$$

As we want $f(x, x) = 0$, we must have $C = 0$ and $Ey + F$ is the equation of the diagonal terms.

$$\frac{f(y - x, y)}{\exp(Ey - F)} = \exp(A(y - x)^2 + 2B(y - x)y + D(y - x)), \quad \forall (x, y) \in [1, n]^2. \quad (7.33)$$

The equation of the parabola reduces to three parameters A, B and D . These parameters are computed employing a regression on the basis vectors $(y - x)^2$, $(y - x)y$ and $(y - x)$, where x and y are the points where the value of f is known, with the components of vector u defined above. The norm of Equation (7.28) is the 2-norm.⁶

f needs to be evaluated at the rating numbers, the lattice $v \times v$, and this gives the migration probabilities. After this step, the log migration matrix is smoothed and is given in Figure 7.13.

⁶The 2-norm of a matrix $A = (a)_{i,j \in \{1, \dots, n\}}$ is defined as $\|A\|_2 = \sum_{i,j} a_{i,j}^2$.

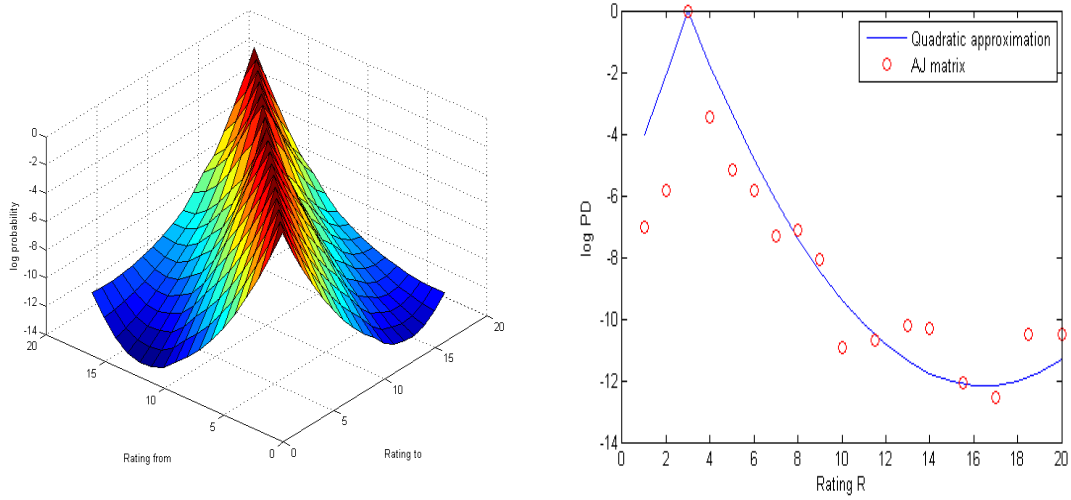


Figure 7.13: Smooth 3-month migration matrix and for R3 rating.

However, these log probabilities do not satisfy all the requirements of Section 7.4.1. For a given rating the probabilities are not always decreasing from the diagonal (the parabola is increasing for good ratings for large migrations). Furthermore, a scaling is needed because the smoothing does not keep the sum of each row (with the PDs) equals to 1. On average, each row has a sum of probabilities equal to 1.43 and not 1. Therefore, two more steps are needed: one to floor to probabilities so that they do not increase from the diagonal and one to scale the probabilities to 1.

We need to floor the probabilities. This is done by flooring to minimum of the corresponding side. In fact, these probabilities are very low and do not influence the IRC. The flooring procedure is performed as follows:

The flooring level for the upper triangular migration matrix is given by

$$m = \min(p_{i,j}, \forall i, j \text{ such that } i \leq j).$$

Then, per current rating (per rows of the matrix) an index j_0^i is defined such that

$$j_0^i = \arg \min_j (p_{i,j}, \forall j \text{ such that } i \leq j).$$

Then, the $p_{i,j}$ with $j \geq j_0^i$ are floored to the probability m . The same is done for the lower triangular migration matrix.

The last step is to scale each row to a total probability mass of 1. This is done when incorporating the PDs. Per current rating (row) a scaling factor is computed as follows:

$$SC_i = \frac{1 - PD(Ri) - p_{i,i}}{\sum_{j=1}^{20} p_{i,j} - p_{i,i}}. \quad (7.34)$$

Then, the new migration probabilities \tilde{p} are given by:

$$\tilde{p}_{i,j} = SC_i \cdot p_{i,j}, \quad \forall i, j \text{ such that } i \neq j. \quad (7.35)$$

The scaling of the probabilities only affects the off-diagonal elements of the matrix because the increase in the probability mass is partly due to the smoothing of the off-diagonal elements.

The outcome of this step is now analysed. In the log space, the migration matrix is given by Figure 7.15, a complete migration matrix is provided in Appendix D.

This analysis is not straightforward because there is no 20×20 reference migration matrix that is available. S&P only publishes a 1-year migration matrix and this is an internal migration matrix, no other bank or agency uses such a matrix (this is an internal rating system). A direct comparison between the unsmoothed matrix and the smoothed is not possible either because the unsmoothed is a 17×17 matrix whereas the final matrix is a 20×20 . However, some indicators may be compared. A classical one is represented by the total migration rates defined as follows:

$$M_i^+ = \sum_{j=1}^{i-1} p_{i,j},$$

and

$$M_i^- = \sum_{j=i+1}^n p_{i,j},$$

$\forall i, 1 \leq i \leq n$, where n the size of the matrix considered (17 or 20) and $p_{i,j}$ are the entries of this matrix. We compare the unsmoothed migration matrix obtained by the AJ method and the smoothed final migration matrix. The results are provided in Figure 7.14. They show an underestimation of the off-diagonal migration probabilities for investment ratings which means that the risk is potentially underestimated. This is due to two factors: The first one is that for ratings R2 to R5 the diagonal elements are overestimated (see Figure 7.12). The second factor is that we are overestimating the PDs (see Figure 7.10) for ratings better than R15. These two factors decrease the scaling factors that are applied only to off-diagonal elements and it decreases the corresponding off-diagonal probabilities.

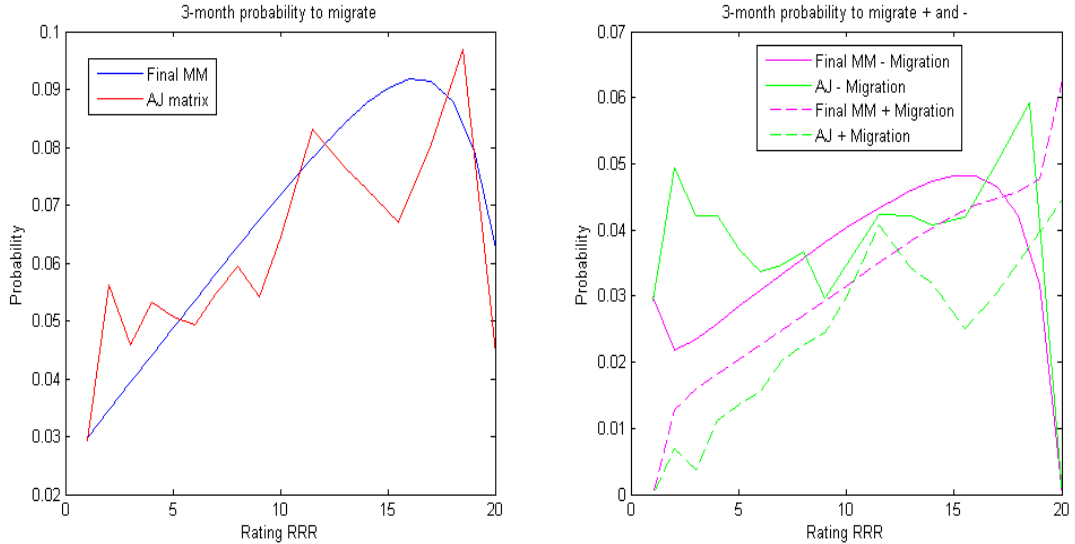


Figure 7.14: On the left: sum of M^+ and M^- , the probability to migrate for the AJ matrix and M . On the right: M^+ and M^- , the probability to be upgraded or downgraded for the AJ matrix and M .

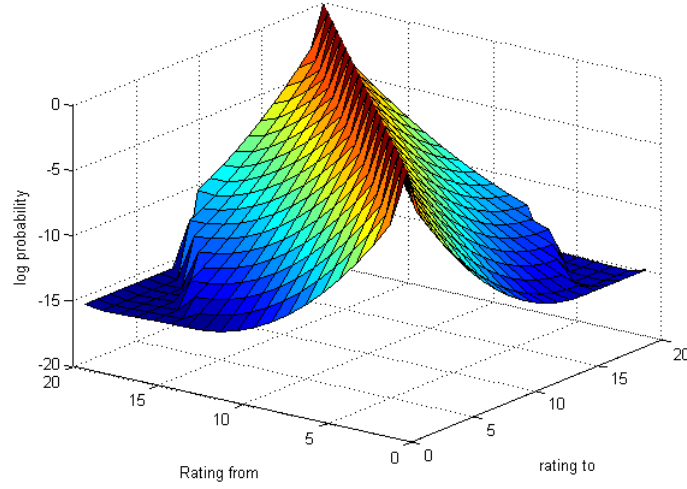


Figure 7.15: Final migration matrix.

7.5 Impact Analysis on the IRC

In this section, the impact of several migration matrices and PDs on the IRC is investigated. The experiments are performed for a single portfolio snapshot of the positions held at a certain date. Various matrices are used for the analysis. Some of them are sensitivity analysis tests and others are to analyse the impact of the smoothing and the internal PDs on the final result. In a first part, the matrices that are employed are explained and then, numerical experiments are performed.

7.5.1 Matrices used for the numerical experiments

As mentioned previously, several matrices are used and they are explained in this section. M is the matrix of reference computed with the complete smoothing procedure described above.

$$M = \left(\begin{array}{cccc|c} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,n} & PD_1 \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,n} & PD_2 \\ p_{3,1} & p_{3,2} & p_{3,3} & \cdots & p_{3,n} & PD_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{n,1} & p_{n,2} & p_{n,3} & \cdots & p_{n,n} & PD_n \\ \hline 0 & 0 & 0 & \cdots & 0 & 1 \end{array} \right)$$

The first type of matrix, consists of shocks applied to the PDs. The PDs are increased or decreased by a certain percent or an absolute shock. The matrices are computed using the complete smoothing method of the previous section (AJ methods and smoothing and the internal based PDs). The rest of the matrix is affected by this change because of the scaling at the end of the procedure. The following matrices are defines: The PDs are increased or decreased relatively by: -10%, -5%, -2.5%, 2.5%, 5%, 10%. The following absolute shocks are also used:

increase of all the PDs of 1bp, 2bp.⁷ These shocks trade the PDs (in green) with the migration probabilities (in red) because of the scaling.

A second type of migration matrix is that the off-diagonal elements (in red) are increased. It means that we trade a part of the probability no migration with the probability to migrate to other ratings. The PDs remain fixed because they are not affected by the scaling. Let $p_{i,j}$ be the elements of M . Before the scaling procedure, the diagonal elements of the new matrix ($\tilde{p}_{i,i}$) are scaled as follows:

$$\tilde{p}_{i,i} = C \cdot p_{i,i}, \quad \forall i.$$

The scaling procedure following this step modifies the off-diagonal elements on a pro-rata basis (relatively).

The following C are used: 90%, 95%, 97.5% and 102.5% (called MR+10%, MR+5%, MR+2.5% and MR-2.5%, respectively). 105% and more cannot be used they lead to negative probabilities. These shocks trade the migration rates (in red) with the diagonal elements (in blue).

A third type of matrix is defined: the γ_i , $i = 1, \dots, n$, from Section 7.3.2 are not smoothed. The rest of the matrix is computed with the same algorithm.

Finally, two other matrices are tested: Only defaults may occur in the simulation, there are no migrations (the migration matrix is an identity matrix minus the PDs) and no defaults, only migrations may occur. The no-migration matrix is given by:

$$M = \left(\begin{array}{cccc|c} 1 - PD_1 & 0 & \cdots & 0 & PD_1 \\ 0 & 1 - PD_2 & \cdots & 0 & PD_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 - PD_n & PD_n \\ \hline 0 & 0 & \cdots & 0 & 1 \end{array} \right).$$

and the zero default matrix is:

$$M = \left(\begin{array}{cccc|c} p_{1,1} & p_{1,2} & \cdots & p_{1,n} & 0 \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} & 0 \\ \hline 0 & 0 & \cdots & 0 & 1 \end{array} \right).$$

In Table 7.7, these shocks are compared because they are very different in terms of magnitude. This table is used in the analysis of the results.

⁷1bp = 0.0001 = 0.01%.

Table 7.7: Comparison of the sensitivity scenarios.

	PDs					MR		Probability no migration	
	Abs (bp)				Relative	Abs(bp) avera	Relative	Abs(bp) average	Relative
M	R2	R6	R10	R15	0	0	0	0	0
PD-10%	-0.01	-0.06	-0.40	-4.70	-10%	140	1%	0	0
PD-5%	ns	-0.03	-0.20	-2.30	-5%	70	0.50%	0	0
PD-2.5%	ns	-0.02	-0.10	-1.20	-3%	35	0.25%	0	0
PD+2.5%	ns	0.02	0.10	1.20	3%	-35	-0.25%	0	0
PD+5%	ns	0.03	0.20	2.30	5%	-70	0.50%	0	0
PD+10%	0.01	0.06	0.40	4.70	10%	-140	1%	0	0
MR-2.5%	0.00	0.00	0.00	0.00	0%	-230	-17.00%	230	-2.50%
MR+2.5%	0.00	0.00	0.00	0.00	0%	230	17.00%	-230	2.50%
MR+5%	0.00	0.00	0.00	0.00	0%	460	34.00%	-460	5%
MR+10%	0.00	0.00	0.00	0.00	0%	920	68.00%	-920	10%
PD+1bp	1.00	1.00	1.00	1.00	Depend on Rating	0	0.00%	-1	<0.1%
PD+2bp	2.00	2.00	2.00	2.00	Depend on Rating	0	0.00%	-2	<0.1%

This table reads as follows: the MR+2.5% does not change the PDs, the migration rate increases by 230bp per row on average which is equivalent to a 17% increase, on average.

7.5.2 Numerical results for the IRC

Numerical experiments are performed in this section. Several characteristics have to be investigated: The sensitivity to a change in the PDs and the off-diagonal migration probabilities should be investigated. The error committed is also analysed.

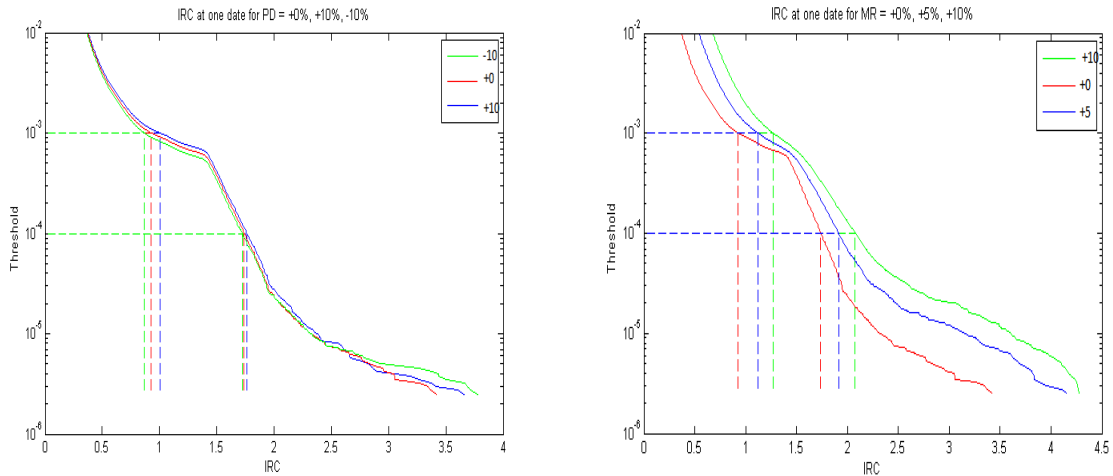


Figure 7.16: Tail distributions of the IRC.

Figure 7.16 gives the tails of the IRC distributions after the IRC simulation and the two thresholds we are interested in: 99.9% and 99.99%. A first conclusion is that higher a migration rate leads to a smoother tail. This was to be expected because defaults lead to larger variations of the portfolio than migrations. A second conclusion is that an increase in the PDs does not seem to lead to a large increase of the IRC. The reason why the PDs does not seem to have a large impact is because the shocks that are applied are very small. Table 7.7 shows that an increase by 10% of the PDs does not shock the good ratings by more than 0.1bp. As it is shown in Figure 7.16, the tail is not smoothed but has a "step" at 1.5. This is due to concentration: A

large exposure on one issuer at certain rating gives this type of behaviour. In that case, this is caused by a large exposure to one issuer at rating R6 (see Table 7.8).

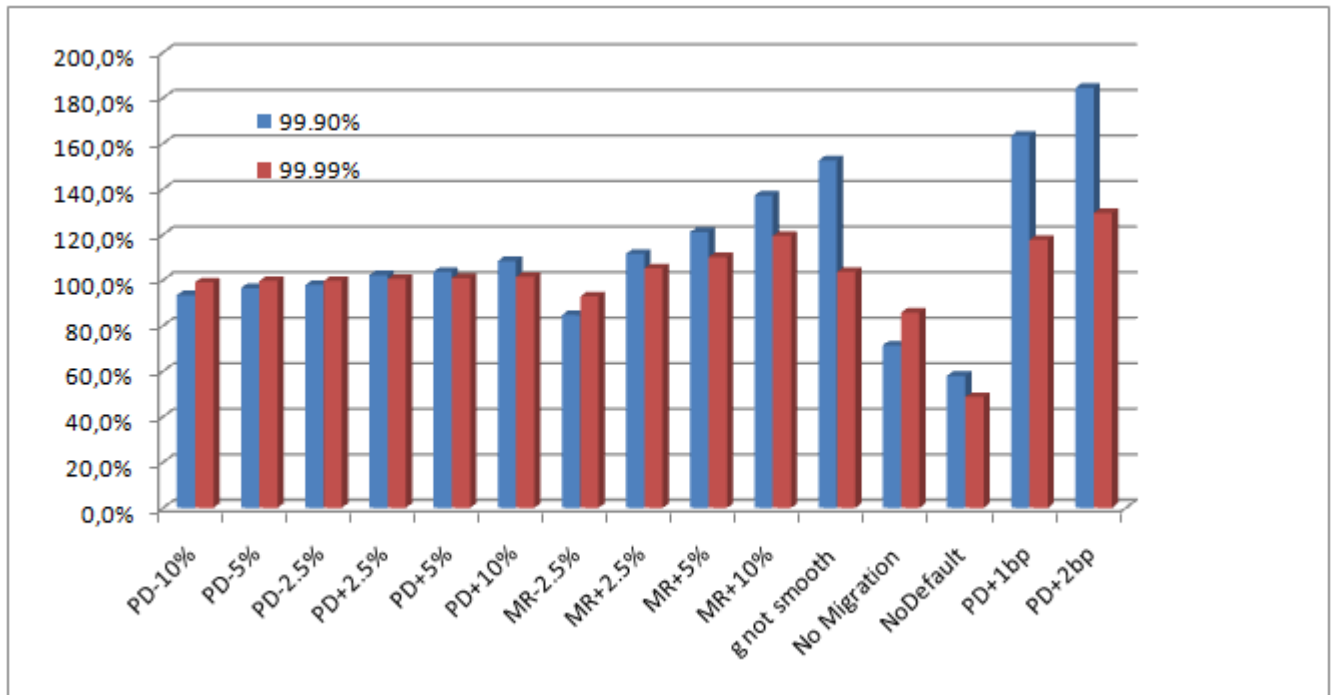


Figure 7.17: Sensitivity to the parameters.

Figure 7.17 gives sensitivity results for different changes in the migration matrix. A first conclusion is that the smoothing of the γ_i , $i = 1, \dots, n$ is not conservative for this portfolio. The reason is that there is a large exposure at one issuer rated R6 (see Table 7.8, this is an exposure of 40% of the largest exposure at the rating R6). The unsmoothed γ_6 is very low and, therefore, it increases the 3-month PD from 0.5bp for the smooth version to 17bp. This creates this increase of IRC at the 99.9% level. However, this is very portfolio dependent. If this issuer migrates to R7 then, the increase is less than 10% between the smooth and the unsmoothed γ_7 . The main reason why the PD sensitivity relative tests seem to have few impact is the variation in absolute value. The migration rates increase by approximately 600bp (see Table 7.7) whereas some PDs increase by only less than 0.1bp for investment ratings. Therefore, absolute shocks on the PDs are also performed and the increase of the IRC is larger while the increase of the PDs is only 1bp. In fact, 1bp is a large increase for investment ratings. The no migration and no default matrices show that the default risk is more important than the migration risk because it leads to a larger IRC.

Table 7.8: Exposures per rating.

Rating	Potential loss in % of the maximum	% of the total potential loss at this rating	Rating	Potential loss in % of the maximum	% of the total potential loss at this rating
R1	100.0%	53.6%	R11	0.2%	0.1%
R2	3.6%	3.4%	R12	1.0%	0.2%
R3	2.3%	1.3%	R13	1.7%	0.5%
R4	5.0%	1.2%	R14	0.0%	0.0%
R5	39.0%	7.0%	R15	15.7%	0.5%
R6	40.3%	7.2%	R16	0.0%	0.0%
R7	9.4%	6.7%	R17	0.1%	0.0%
R8	28.2%	10.6%	R18	0.0%	0.0%
R9	23.6%	4.1%	R19	0.0%	0.0%
R10	48.1%	3.6%	R20	0.0%	0.0%

7.6 Confidence Intervals

In this section, we derive confidence intervals for the migration matrix and the PDs. The objective is to show that the method gives stable results. A bootstrapping method is employed to measure confidence intervals. This method is applied to both the γ_i , $i = 1, \dots, n$, and the migration matrix by using several subsamples from the entire dataset. Subsamples are created by taking randomly 70% of the companies from the dataset. Many other procedures have been proposed in the literature. A realistic one consists in reproducing an artificial dataset based on the estimated migration matrix. A Monte Carlo method simulates rating processes and after the simulation the migration matrix is re-estimated. Many requirements have to be set on the dataset: Same number of companies, same arrival times and time-homogeneity. This method is completely explained in [11], however, it cannot be used for our purposes. There are two main reasons for this: time-homogeneity cannot be assumed for our dataset that covers 30 years (in [11], data covers only of 5 years to avoid this problem). Secondly, the method will lead to unreliable results due to the smoothing procedure and the fact that the PDs are not determined with the same methodology.

Two bootstrapping methods may be performed: one may be done by shifting the time window and the other takes subsamples from the companies. However, we saw in Figure 7.2 that time-homogeneity could not be assumed due to large variation in the PDs. This may affect the quality of the bootstrapping method that uses a time window, so only the bootstrapping method on companies is performed with 2000 simulations. The confidence interval level is 95%.

Partial results are given for the migration matrix in Table 7.9 because the complete migration matrix is too large to be displayed (they are representative of the confidence intervals of other migration probabilities) and complete results are given for the PDs in Table 7.10. The confidence intervals are pretty tight revealing that the procedure is relatively accurate and stable: For the probabilities of no migration, the confidence intervals are tight because there is a lot of data. For off-diagonal elements the confidence intervals are wider (for instance [1.31; 2.44

; $2.79 \cdot 10^{-7}$ from R4 to R20) but these probabilities do not heavily impact the IRC. For the PDs, the confidence intervals are also pretty tight. The upper confidence interval corresponds to an increase of the PDs of roughly 10% which was one of the scenarios from Section 7.5. We saw that for such a scenario the increase of the IRC is acceptable (less than 5%).

Table 7.9: 95% confidence intervals for the migration probabilities.

Lower CI	0.968951	0.000846	5.64E-06	3.47E-07	1.7E-07	1.7E-07
	0.000293	0.955014	0.000235	4.48E-06	7.35E-07	1.31E-07
	1.45E-06	0.000131	0.93646	0.000497	1.67E-05	4.25E-06
	1.12E-07	2.36E-06	0.000333	0.917689	0.000877	4.69E-05
	8.2E-08	1.21E-07	1.28E-05	0.000703	0.899034	0.001317
	8.51E-08	8.51E-08	8.51E-08	7E-05	0.001718	0.880761
Ratings	R1	R4	R8	R12	R16	R20
R1	0.97021	0.000972	8.00E-06	5.63E-07	3.14E-07	3.14E-07
R4	0.000369	0.95586	0.000273	5.73E-06	1.03E-06	2.44E-07
R8	2.30E-06	0.000166	0.93704	0.000538	1.92E-05	5.84E-06
R12	1.97E-07	3.37E-06	0.000387	0.9186	0.00094	5.70E-05
R16	1.46E-07	9.43E-07	1.60E-05	0.000806	0.90052	0.001464
R20	1.58E-07	1.58E-07	1.58E-07	8.71E-05	0.001933	0.8828
Upper CI	0.971385	0.001036	8.83E-06	6.22E-07	3.54E-07	3.54E-07
	0.000394	0.956622	0.000296	6.13E-06	1.1E-06	2.79E-07
	2.53E-06	0.000173	0.937631	0.000571	2.04E-05	6.97E-06
	2.17E-07	3.5E-06	0.000396	0.91954	0.001033	7.14E-05
	1.63E-07	9.81E-07	1.64E-05	0.00084	0.901998	0.001742
	1.82E-07	1.82E-07	1.82E-07	9.85E-05	0.002102	0.884858

Table 7.10: 95% confidence intervals for the PDs.

	Lower PDs	PD	Upper PDs
R1	6.79E-06	7.94E-06	8.26E-06
R2	1.04E-05	1.21E-05	1.26E-05
R3	1.6E-05	1.84E-05	1.91E-05
R4	2.47E-05	2.81E-05	2.9E-05
R5	3.79E-05	4.28E-05	4.41E-05
R6	5.84E-05	6.53E-05	6.71E-05
R7	8.98E-05	9.96E-05	0.000102
R8	0.000138	0.000152	0.000155
R9	0.000224	0.000244	0.000249
R10	0.000374	0.000405	0.000413
R11	0.000617	0.000661	0.000674
R12	0.001015	0.001079	0.001101
R13	0.001665	0.001763	0.001798
R14	0.002732	0.002879	0.002938
R15	0.004479	0.004702	0.004804
R16	0.007328	0.00768	0.007852
R17	0.012008	0.012542	0.012853
R18	0.019677	0.020483	0.021057
R19	0.032241	0.033452	0.034521
R20	0.052613	0.054632	0.056702

7.7 Advanced Models to Reflect non-Markovian Migrations

The objectives of this section are to investigate the possible evolutions of the migration matrix and to show the limits of the current model. As it may be seen in Table 7.3, the current Markov chain does not capture that the migration matrix is influenced by the past ratings and that the rating process is not time-homogeneous. The assumption that the rating process is Markovian is impossible to confirm on historical ratings. This section investigates a simple method to have a more realistic model. The only requirement for this part is that we wish to keep the concept of migration matrix. A migration matrix is convenient to use from a business perspective and there is no doubt that the concept of migration matrix will still be used for the next 20 years in risk management. The ratings are merged for this analysis: the rating of S&P are used without the +/-: $S = \{AAA/AA, A, BBB, BB, B, CCC/CC/C, D\}$. The objective is still to obtain a 3-month migration matrix. The idea is to include some information about the past ratings to compute the migration matrix. Let $R(t)$ be a rating process at time t , the migration probability from i to j , $p_{i,j}$ is currently given by:

$$p_{i,j} = P(R(t + \Delta t) = j | R(t) = i),$$

where Δt is the time horizon of the matrix (here 3 months).

We wish to include information about the past ratings. Let \mathcal{F}_t be the information about the

past ratings up to time t of a given issuer (This is the same concept as a filtration⁸, which is information about the past). The following probability would be more accurate for the migration probability:

$$P(R(t + \Delta t) = j | R(t) = i, \mathcal{F}_t), \quad (7.36)$$

Therefore, the method that is presented in this section includes some information about the past ratings. One possibility is to take in consideration the previous rating of the issuer. The method that is presented in this section is mainly based on [5]. The advantage of this method is that it fits in the Markov framework used at many banks.

In the previous section, the state space only contained information about the current rating: $S = \{AAA/AA, A, BBB, BB, B, CCC/CC/C, D\}$. In this section, the ratings are divided in three categories depending if the rating of issuer a certain time ago, δt , was better, worse or equal to the current rating. Therefore, Equation (7.36) may be rewritten with $\mathcal{F}_t = \{R(t - \delta t) < R(t), R(t - \delta t) > R(t), R(t - \delta t) = R(t)\}$. This leads to three different probabilities:

$$p_{i,j}^+ = P(R(t + \Delta t) = j | R(t) = i, R(t - \delta t) > R(t)), \quad (7.37)$$

$$p_{i,j}^- = P(R(t + \Delta t) = j | R(t) = i, R(t - \delta t) = R(t)), \quad (7.38)$$

$$p_{i,j}^- = P(R(t + \Delta t) = j | R(t) = i, R(t - \delta t) < R(t)). \quad (7.39)$$

The first equation is the migration probability from i to j knowing that an upgrade occurred. This method would capture some non-Markovity (information on the past) of the rating process but also some time-inhomogeneity (matrix changing over time).

Some questions still have to be answered: Is it possible to write the method in form of one time-homogeneous Markov chain (with a migration matrix)? What δt should be equal to? Should it be equal to Δt ?

The three questions are related and we try to answer them in the following.

The answer to the first question is that it may be easily represented as a Markov chain if $\Delta t = \delta t$. For instance, if the issuer migrates up and is now rated Ri . Then, the migration probabilities are $p_{i,\cdot}^+$. The main problem is that if, after an upgrade, the issuer is now downgraded to Rk , the migration matrix has to be different because the new migration probabilities are $p_{k,\cdot}^-$. We can define the following state space for the Markov chain $S = \{R1^+, R2^+, \dots, Rn^+, R1^-, R2^-, \dots, Rn^-, R1^-, R2^-, \dots, Rn^-\}$. The migration matrix may be written as follows:

⁸No formal definition of filtration is introduced because it is not needed to understand the analysis.

$$M = \left(\begin{array}{cccc|cccc|cccc} 0 & 0 & \cdots & 0 & p_{1,1}^+ & 0 & \cdots & 0 & 0 & p_{1,2}^+ & \cdots & p_{1,n}^+ \\ p_{2,1}^+ & 0 & \cdots & 0 & 0 & p_{2,2}^+ & 0 & \cdots & 0 & 0 & p_{2,3}^+ & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n,1}^+ & \cdots & p_{n,n-1}^+ & 0 & 0 & \cdots & 0 & p_{n,n}^+ & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & p_{1,1}^- & 0 & \cdots & 0 & 0 & p_{1,2}^- & \cdots & p_{1,n}^- \\ p_{2,1}^- & 0 & \cdots & 0 & 0 & p_{2,2}^- & 0 & \cdots & 0 & 0 & p_{2,3}^- & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n,1}^- & \cdots & p_{n,n-1}^- & 0 & 0 & \cdots & 0 & p_{n,n}^- & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & p_{1,1}^- & 0 & \cdots & 0 & 0 & p_{1,2}^- & \cdots & p_{1,n}^- \\ p_{2,1}^- & 0 & \cdots & 0 & 0 & p_{2,2}^- & 0 & \cdots & 0 & 0 & p_{2,3}^- & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n,1}^- & \cdots & p_{n,n-1}^- & 0 & 0 & \cdots & 0 & p_{n,n}^- & 0 & 0 & \cdots & 0 \end{array} \right).$$

This defines a Markov chain. Therefore, this model satisfies the requirement as we still have a migration matrix. It may be noted that many elements are 0 because it is impossible to be rated R^+ after a downgrade, for instance.

The same method may be used in case there exists an integer k such that $\delta t = k\Delta t$. However, computing the migration matrix for previously upgraded or downgraded leads to a larger matrix. We can optimize δt such that the distance between the two migration matrices (the one after an update and after a downgrade) is the greatest as possible. This is performed by computing the norm of the difference of the migration matrices (PDs included). AJ method is employed. Figure 7.18 presents the norm of the difference for various norms. To maximize the distance between the two matrices, the best δt is 60 days for $\Delta t = 3m$. However, the norm remains high for $\delta t = 3m$, because the overall migration matrix and the implementation are simpler if $\delta t = \Delta t$ we fix $\delta t = 3m = 90d$. Another way to determine δt would have been to compute the IRC for different δt and then to take the one impacting the results the most. However, this method would be portfolio dependent.

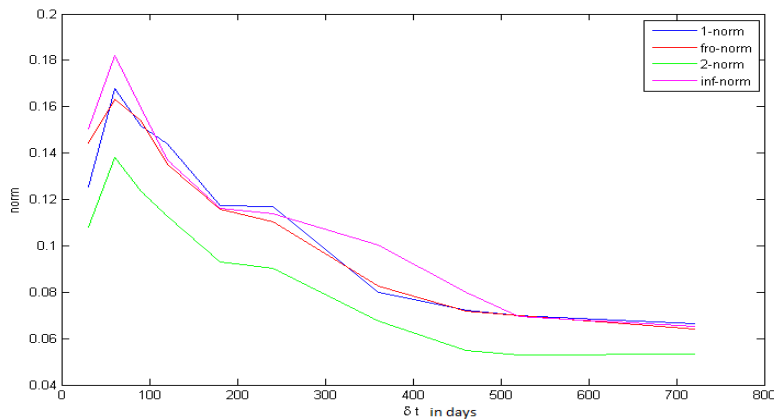


Figure 7.18: Distance between the migration matrices.

This method gives the following migration matrices:

Table 7.11: 3-month migration matrices knowing the previous rating.

	AAA/AA	A	BBB	BB	B	CCC/CC/C	D
upgrade	AAA/AA	98.39%	1.26%	0.34%	0.00%	0.00%	0.00%
	A	0.09%	99.42%	0.32%	0.17%	0.00%	0.00%
	BBB	0.00%	0.31%	98.92%	0.29%	0.20%	0.28%
	BB	0.00%	0.00%	0.45%	99.14%	0.35%	0.06%
	B	0.00%	0.02%	0.08%	0.12%	98.71%	1.07%
	CCC/CC/C	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
	AAA/AA	A	BBB	BB	B	CCC/CC/C	D
stable	AAA/AA	97.98%	1.89%	0.10%	0.02%	0.01%	0.00%
	A	0.55%	97.74%	1.56%	0.10%	0.04%	0.00%
	BBB	0.06%	1.17%	97.20%	1.36%	0.18%	0.02%
	BB	0.02%	0.11%	1.59%	95.68%	2.35%	0.17%
	B	0.02%	0.06%	0.10%	1.55%	95.71%	1.88%
	CCC/CC/C	0.00%	0.10%	0.17%	0.26%	4.43%	86.46%
	AAA/AA	A	BBB	BB	B	CCC/CC/C	D
downgrade	AAA/AA	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	A	0.01%	95.56%	4.25%	0.15%	0.01%	0.00%
	BBB	0.12%	0.27%	96.19%	2.06%	0.67%	0.31%
	BB	0.08%	0.01%	0.50%	92.58%	4.73%	0.92%
	B	0.00%	0.06%	0.00%	0.49%	88.42%	5.86%
	CCC/CC/C	0.00%	0.05%	0.02%	0.26%	1.82%	68.66%
	AAA/AA	A	BBB	BB	B	CCC/CC/C	D
all	AAA/AA	97.92%	1.95%	0.10%	0.02%	0.01%	0.00%
	A	0.51%	97.82%	1.54%	0.09%	0.03%	0.00%
	BBB	0.05%	1.04%	97.44%	1.28%	0.15%	0.02%
	BB	0.02%	0.08%	1.45%	95.87%	2.34%	0.16%
	B	0.01%	0.05%	0.09%	1.45%	95.75%	1.95%
	CCC/CC/C	0.00%	0.08%	0.15%	0.22%	4.06%	83.74%

The difference of the PDs between the matrices is quite large. For instance, a company rated BB has the following PDs: 0.06%, 0.07% and 1.18% for respectively an upgrade, a stable rating and a downgrade. As mentioned previously, implementation is not easy task because of the IRC software employed and because banks do not always have the history of the rating of every issuers. We have only a two options to try to analyse the potential impact on the IRC. The first one is to measure the IRC independently for the three matrices on the same portfolio. The second option is to assume that the issuers stay in the sub-matrix (upgrade, stable, downgrade) where they started the simulation. It means that if an issuer was upgraded 2 months before the start of the simulation, then, it remains in the upgraded matrix for the complete simulation. The second option is more realistic because it is consistent with the constant risk assumption which states that the risk should be the same at the beginning of every liquidity periods. The

matrix with the constant risk assumption is the following:

$$M = \left(\begin{array}{cccc|cccc|cccc} p_{1,1}^+ & p_{1,2}^+ & \cdots & p_{1,n}^+ & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ p_{2,1}^+ & p_{2,2}^+ & \cdots & p_{2,n}^+ & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n,1}^+ & p_{n,2}^+ & \cdots & p_{n,n}^+ & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & p_{1,1}^- & p_{1,2}^- & \cdots & p_{1,n}^- & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & p_{2,1}^- & p_{2,2}^- & \cdots & p_{2,n}^- & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & p_{n,1}^- & p_{n,2}^- & \cdots & p_{n,n}^- & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & p_{1,1}^+ & p_{1,2}^+ & \cdots & p_{1,n}^+ \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & p_{2,1}^+ & p_{2,2}^+ & \cdots & p_{2,n}^+ \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & p_{n,1}^+ & p_{n,2}^+ & \cdots & p_{n,n}^+ \end{array} \right).$$

The impact on IRC is measured by creating artificial portfolios with some issuers that have just been downgraded or upgraded or that are stable. This method may lead to large differences in the results.

The first portfolio P1 tested has one issuer per rating and they all have the same potential loss in case of a default (normalized to 1). The second portfolio P2 is the following: The stable ratings have 90% of the exposure and 5% for the upgraded ratings and 5% for the downgraded ratings. Finally, P3 is the same as P2 but a large issuer has just been downgraded from BBB to BB. Results are provided in Table 7.12.

Table 7.12: Impact on IRC of three migration matrices.

P / α	99.9%	99.99%
P1	+18%	+12%
P2	+2%	+2%
P3	+37%	+43%

Results are rather heterogeneous among the portfolios. As expected, in all cases the IRC is increasing, however, the increases are very different depending on the portfolio.

This matrix may be extended to the case $\delta t = k \cdot \Delta t$, $k \in \mathbb{N}$, by increasing the dimension of M . In that case, it is also possible to take into account the date of the last migration more precisely. This may be done through the same procedure as in the previous section but the matrix should be of size $3 \cdot k \cdot n + 1$. Calibration is as complex as the previous section. This model may also include the time since the last rating changes. For instance there may be a matrix for the downgrade that happened 3 month of less ago, the one that occurred 3-6 months ago and so on ... This would lead to a more complex migration matrix. Also, calibration may become quite complex and some more steps may have to be done, such as a smoothing between these matrices.

For now, the banking industry does not capture the risk of successive downgrades at all. The

regulators prefer to insist on implementing several migration matrices for various types of issuers (financial institutions, sovereigns, corporates...).

7.8 Conclusion and Further Research

In this chapter, a method to compute the migration matrix and the PDs was defined consistently with the requirements given. The outcome was a rather stable matrix (confirmed by tight confidence intervals) and reliable PDs.

Further research should probably reconsider the requirements of the matrix and the PDs. The consistency with the credit risk Economic Capital lead to two steps in the methodology: PDs and migration matrix. The PDs were not fully consistent with the historical PDs, and these were very conservatives compared to historical PDs.

As shown in this chapter, another point of concern was the modelling of the rating process by a Markov chain directly on the rating space. We showed that this process was historically not time-homogeneous nor Markov. As seen in Section 7.7, the impact on the migration matrix and on the IRC was quite substantial, more research may be done in this area. We presented a simple but reliable alternative that employed a migration matrix on a different space of states in order to keep the Markov chain model but captured some non-Markovity of the rating process.

Chapter 8

Conclusion and Outlook

The Economic Capital is a major aspect of risk management and investment decisions. This thesis focused on the two main components of the Economic Capital for the trading book, which are the market price (fluctuation) risk and the IRC that capture the credit risk.

In Chapter 3, we saw that for the market fluctuation risk, historical simulated overlapping PLs are used because these were directly available and they provided reliable and realistic shocks. This option did not require a large computing power and was easily understandable for daily users of the Economic Capital.

In Chapter 4, a general overview of the model was provided. The idea was to have a model that covers 99.99% of all the possible yearly scenarios. Subsequently, the market fluctuation risk and the IRC were defined and explained. The first one covered the market risk and the second one covered the credit risk of the trading book.

The market fluctuation risk was modelled through a sampling method where 25 randomly chosen 10-day PLs were summed up to obtain a possible yearly scenario. Then, by repeating this summation many times, a complete 1-year distribution was empirically composed and the $\text{VaR}_{99.99\%}$ was taken. This capitalized for the market risk that is present in the trading book. However, the final model employed an approximation: $SF \cdot \rho(X)$ where SF is a scaling factor, ρ a risk measure and X a random variable representing the PLs. In Chapter 5, two possible risk measures for ρ were analysed, which were VaR and ES. This chapter provided a complete statistical analysis of the estimators of the risk measures. The outcome was that for non-autocorrelated time series, the estimators were biased but the bias remained quite low when employing the complete dataset (1000 PLs). The MSE was also rather low for both risk measures. For autocorrelated (so for the historical PLs), the behaviour of the estimators was more problematic especially for ES. Employing 1000 PLs, the bias may be as severe as -4% with a MSE of 20%. VaR lead to similar results, although the bias was lower (in absolute value), the MSE was about the same. The conclusion was that the use of overlapping PLs diminished the quality of the data and increased significantly the bias and the MSE for classical processes (AR process and T5). The threshold α should be chosen such that $n \cdot (1 - \alpha)$ was an integer, because a rounding or a linear interpolation lead to a higher error or a higher complexity of the estimators, respectively. A general rule was that the maximum amount of data that is available

should be used and the threshold should not be too extreme. Although, the use overlapping PLs diminished the quality, they should be used because they also brought information.

Chapter 6 employed the estimators to compute the scaling factor and the market fluctuation risk. The criteria that were analysed were the correlation and the error with the $\text{VaR}_{99.99\%}$ from the sampling methodology, but also the stability of the scaling factor and the volatility of the risk measure. The results were very much in favour of ES. It showed a higher correlation and a lower error with the sampled VaR than a scaling based on VaR. Subsequently, optimal threshold levels were derived to minimize the error and maximize the correlation with the sampling VaR. We showed that for ES thresholds of 95% or 96% were appropriate and for VaR the choice is more difficult, due to a lack of a clear optimum, but thresholds ranging from 97.5% to 99% were suitable. A change to ES lead to a model reflecting the tail risk better by its very definition and was therefore recommended.

The last topic was IRC and more precisely the migration matrix and the PDs. Several general methods were tested and the most advanced was chosen (AJ methods) because it took into account the specificities of the rating process the best. To avoid too much reliance on limited historical data, a smoothing procedure was performed consistently with other methods employed at the bank. The outcome was a smooth, reliable, and robust matrix consistent with the requirements. Sensitivity tests were also performed and they showed the impact of both the smoothing procedure and some artificial shocks (e.g. PD +10%). The unsmoothed PDs test lead to a high reliance on the data and to a decrease of the IRC in case of a downgrade of a large issuer because the unsmoothed PDs were not all increasing as the rating was worse. A downgrade of an issuer is supposed to lead to higher IRC. In a last part of this thesis, we showed the limits of a model only based on one Markov migration matrix. A Markov model associated to the rating space was not able to capture non-markovity, therefore, another model was proposed. This takes into account the previous migration (upgrade, downgrade, no migration). By mean of experiments, we show that this may have a consequences on the IRC figures and the migration matrix.

Further research may still be performed about Economic Capital modelling.

A first step could be to reconsider some hypotheses. The constant risk assumption, which states that the risk should be the same at the beginning of each liquidity horizon, is probably very conservative. A constant position assumption may be more realistic although more complex to implement. This change may be tested for both the market fluctuation risk and the IRC.

Many other risk measures may also be investigated. A family of them is the spectral risk measure¹ defined as:

$$\rho(X) = - \int_0^1 \varphi(x) \cdot F_X^{-1}(x) dx,$$

where X is random variables (like a PL distribution) with cumulative distribution function F_X . φ is a function to be determined. Typically, this function gives a certain weight to each PL (in case X is a discrete random variable). Some requirements of this function may ensure coherence of the spectral risk measure (see [2]). This type of risk measures relates to the economic concept of utility function (see [4]) and, therefore, may be meaningful in terms of interpretation. The main problem with this risk measure is to determine the function φ .

¹ES is a spectral risk measure.

A new regulation for the Regulatory Capital has been proposed and will be made final next year. Although the regulation is not final, it will most likely impact the way banks model the market risk. Therefore, this will be the opportunity for banks to re-assess their Economic Capital model and to investigate a possible alignment or a divergence with the Regulatory Capital. A first point is that this new regulation will, very likely, require a change to ES. This review will also probably lead to new and different liquidity horizons for each type of products. A trend in risk management is to separate different products and asset classes. There are two reasons for this, the first one is that during the crisis many different assets were in the same framework and the historical correlations, which seemed to be stable, changed drastically and lead to a large underestimation of the risk. Separating the asset classes and products allows risk managers to set more conservative correlations. The second reason is that some products need specific models because they carry certain types of risks. This will lead so some interesting possibilities to develop new methods to aggregate the different types of risks although the Basel Committee may standardize many models.

Bibliography

- [1] O. Aalen and Johansen. An empirical transition matrix for non homogeneous markov chains based on censored observation. *Scandinavian journal of statistics*, 1978.
- [2] C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion.
- [3] P.J. Acklam. An algorithm for computing the inverse normal cumulative distribution function. <http://home.online.no/pjacklam/notes/invnorm/>.
- [4] K.J. Arrow. The theory of risk aversion.
- [5] R. Bahar and K. Nagpal. Dynamics of rating transition. *Algo Research Quarterly*, Vol.4:71–92, 2001.
- [6] Barron. Dictionary. 2012.
- [7] S.A. Broda and M.S. Poalella. Expected shortfall for distributions in finance. *Statistical Tools for Finance and Insurance*, pages 57–99, 2012.
- [8] M.v. Buren, G.v. Roekel, K. Moorman, and N. Bouman. Ec for market risk, 2011.
- [9] S.X. Chen. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, Vol.6:87–107, 2005.
- [10] S.Xi. Chen and C.Y. Tang. Nonparametric inference of value at risk for dependent financial return. *Journal of financial econometrics*.
- [11] J. Christensen, E. Hansen, and D. Lando. Confidence sets for continuous-time rating transition probabilities. *Recent Research on Credit Ratings*, Vol.28(11):25752602, 2004.
- [12] P.F. Christoffersen, F.X. Debold, and T. Schuermaan. Horizon problem and extreme events in financial risk management. *Economic Policy Review*, Vol. 4(3), 1998.
- [13] J. Danielsson, B.N. Jorgensen, G. Samorodnisky, M. Sarma, and C.G. de Vries. Subadditivity re-examined : the case for value-at-risk. *Working Paper*, 2005.
- [14] K. Dowd. *Measuring Market Risk*. Wiley Finance, 2005.
- [15] D. Dwyer, Z. Li, S. Qu, H. Russell, and J. Zhang. Cds-implied edf credit measures and fair-value spreads. *Moody's Analytics*, 2010.
- [16] A. Elizade and R. Repullo. Economic and regulatory capital in banking: What is the difference? *International Journal of Central Banking*, Vol.3(3), 2007.

-
- [17] W.G. Faris. Lectures on statistics. 2003.
 - [18] GE.B.Vrugt. Estimating implied default probabilities and recovery values. *The Journal of Fixed Income*, Vol. 21(2), 2010.
 - [19] U. Grzybowska, M. Karwansky, and A. Orlowski. Examples of migration matrices models and their performance in credit risk analysis. *Acta Phisica Polonica A*, Vol. 121(2B):40–45, 2012.
 - [20] G.A. Holton. History of value-at-risk: 1922-1998. *working paper*, 2002.
 - [21] J. Hull, I. Nelken, and A. White. Merton’s model, credit risk and volatility skews. *Journal of Credit Risk*, Vol.1(1):3–28, 2005.
 - [22] K. Inui and M. Kijma. On the significance of epected shortfall as a coherent risk measure. Vol.29(4):853–864, 2004.
 - [23] R.B. Israel, J.S Rosenthal, and J.Z. Wei. Finding generators for markov chains via empirical transition matrices, with appliction to credit ratings. *Mathematical Finance*, Vol. 11(2):245–265, 2001.
 - [24] R. Jarrow, D. Lando, and S. Turnbull. A markov model for the term structure of credit spreads. *Review of Financial Studies*, Vol. 10(2), 1997.
 - [25] E. Jokivuolle and S. Peura. Rating targeting and dynamic economic capital. *The Journal of Risk*, Vol.12(4):87–117, 2010.
 - [26] M.v Jole and M. Mesters. Mapping rabobank risk ratings to s&p ratings. *Rabobank Group, GRM*, 2008.
 - [27] J.H.T. Kim and M.R.Hardy. Quantifying and correcting the bias in estimated risk measures. pages 365–386, 2007.
 - [28] Moody’s KMV. Modeling default risk. 2003.
 - [29] A. Kreinin and M. Sidelnikova. Regularization algorithms for transition matrices. *Algo Research Quarterly*, 2001.
 - [30] A. Kreinin and M. Sidelnikova. Regularization algorithms for transition matrices. 2001.
 - [31] U. Kuchler and M. Sorensen. Exponential families of stochastic processes. *New-York Springer*, 1997.
 - [32] J. Kwiatkowski and R. Rebonato. A coherent aggregation framework for stress testing and scenario analysis. *Applied Mathematical Finance*, Vol.17(2), 2011.
 - [33] D. Lando and T.M. Skodeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance.*, Vol.26(2-3):423–444, 2002.
 - [34] G. Lebenon. Bias, variance and mse of estimator. *working paper*, 2010.
 - [35] A J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concept Techniques tools*. Princeton series in finance, 2005.

-
- [36] M.B. Miller. *Mathematics and Statistics for Financial Risk Management*. Willey Finance, 2012.
 - [37] C. Moussu and A. Petit-Romec. Bank capital and risk-taking: Old and new perspectives from the crisis. *Bankers, Markets & Investors*, Vol.122, 2013.
 - [38] Basel Committee on Banking Supervision. The new basel capital accord. 2001.
 - [39] Basel Committee on Banking Supervision. Guideline for computing capital for incremental risk in the trading book. 2009.
 - [40] Basel Committee on Banking Supervision. Principles for sound stress testing practices and supervision. 2009.
 - [41] Basel Committee on Banking Supervision. Range of practices and issues in economic capital frameworks. 2009.
 - [42] Basel Committee on Banking Supervision. Messages from the academic literature on risk measurement for the trading book. 2011.
 - [43] Basel Committee on Banking Supervision. Fundamental review of the trading book, 2012.
 - [44] Basel Committee on Banking Supervision. Regulatory consistency assessment programme-analysis of risk-weighted assets for market risk. 2013.
 - [45] V. Proviziatou, S. Markose, and O. Menkens. Empirical scaling rules for value at risk. *working paper*, 2006.
 - [46] W.T. Shaw. Sampling student's t-distribution - use of the inverse cumulative distribution function. *Journal of Computational Finance, Online only*, 2004.
 - [47] M. Sherman. A short history of financial deregulation in the united states. *Center for Economic and Policy Research*, 2009.
 - [48] J. Skoglund, D. Erdman, and W. Chen. On the time scaling of value-at-risk. *The Journal of Risk Model Validation*, 2011.
 - [49] D. Tasche. Expected shortfall and beyond. *Journal of banking and finance*, Vol.26(7):1519–1533, 2002.
 - [50] Financial Times. Article <http://www.zerohedge.com/news/2013-04-12/jpmorgan-changes-var-calculation-fourth-time-past-year>.
 - [51] A. Vaart. *Time Series*. 2010.
 - [52] S& P Website. S&p. 2012.
 - [53] D. Xu and F. Nencioni. Introducing the j.p.morgan implied default probability model : A powerful tool for bond valuation. 2000.
 - [54] Y. Yamai and T. Yoshida. Comparative analyses of expected shortfall and value-at-risk under market stress. *Bank for international settlements*, 2002.
 - [55] K. Yoshihara. The bahadur representation of sample quantiles for sequences of strongly mixing random variable. *Statistics and probability letters*, Vol.24(4):299–304, 1995.

Appendix A

Regulatory Capital Model

We defined the Regulatory Capital in 2.2.2. By completeness, we give a more detailed explanation of the main terms that compose it. It is said to reflect the 99.9% worst case. This formula is very likely to change over the next years. The formula is not the same for every banks. For major international banks, it is given by the following:

$$RC_{Tradingbook} = SF \cdot (VaR_{99\%} + SVaR_{99\%}) + IRC + ADD, \quad (A.1)$$

where:

- The $VaR_{99\%}$ is taken over the historical PLs of last year. This 1-year window is rolling over time.
- $SVaR_{99\%}$ is the VaR taken over the PLs of the stress period June 2008 - June 2009. These PLs contain the credit crisis and Lehman bankruptcy. This is called the Stressed VaR.
- SF is a Scaling Factor expending the $VaR_{99\%}$ and the $SVaR_{99\%}$ to a 99.9% threshold and to a 1-year Capital horizon. Although the threshold 99.9% is not explicitly mentioned by the Basel committee it is the threshold of the IRC. It is determined by backtesting: A large number of exception during backtesting increases the scaling factor. Its value is within the range 3-4 for the Netherlands.
- IRC is Incremental Risk Charge as defined in Section 4.2 taken with a 99.9% threshold.
- As for the Economic Capital, there are other terms in that formula (combined in ADD) reflecting specific risks of the bank. They have a marginal impact on the Regulatory Capital for banks that do not enter very complex trades.

Appendix B

Background of Statistics

This appendix provides some statistical definitions and theorems required to understand the estimation of VaR and ES. The first one is the definition of the empirical Cumulative Distribution Function (CDF) of a sample.

B.1 Definitions and Classical Theorems

Definition B.1.1

Given an observed sample/time series $X = \{X_1, \dots, X_n\}$ the empirical cumulative distribution function is defined as follows:

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq t}. \quad (\text{B.1})$$

This is the trivial estimator that represents the empirical distribution as a step function. We have seen in Section 3.2.2 that the 10-day PLs are overlapping. Therefore, this behaviour needs to be taken in consideration in this analysis. The classical central limit theorem only considers the case where the observations are independent. Theorems taking in consideration the dependency have to be defined.

The next definition a classical property in time series theory.

Definition B.1.2 (Stationary time series)

A time series $\{X_t, \dots, X_{t+h}\}$ is strictly stationary if the distribution of (X_t, \dots, X_{t+h}) is independent of t , $\forall h \geq 0$.

It means that the sample can be autocorrelated but the distribution of the process does not change over time. The definition of a measure of dependency is given. This definition is widely used in many theorems and applications.

Definition B.1.3 (α -mixing coefficients, [51])

The α - mixing coefficients of a time series $(X_t)_{t \in \mathbb{R}}$ are defined as follows:

$$\alpha(0) = \frac{1}{2}, \quad \alpha(k) = 2 \cdot \sup_t \sup_{A \in \sigma(\dots, X_{t-1}, X_t), B \in \sigma(X_{t+k}, \dots)} |P(A)P(B) - P(A \cap B)| \quad \forall k > 0. \quad (\text{B.2})$$

A time series is said α -mixing if $\lim_{k \rightarrow +\infty} \alpha(k) = 0$.

In the definition, $\sigma(\dots, X_{t-1}, X_t)$ is the σ -field generated by the random vector (\dots, X_{t-1}, X_t) . In non-measure-theory terms, it is the collection of all the possible outcome of the random vector (\dots, X_{t-1}, X_t) . The supremum is taken over this collection.

The financial interpretation of an α -mixing time series is that the autocorrelation of the PLs decreases as the lag increases. However, the α -mixing condition is stronger.

The next theorem is a central limit theorem for α -mixing time series.

Theorem B.1.1 (Ibragimov's theorem, [51])

If $X = \{X_1, \dots, X_n\}$ is a strictly stationary time series with mean 0 such that:

$E|X_1|^r < \infty$ and,

$\sum_k \alpha(k)^{1-\frac{2}{r}} < \infty$ for some $r > 2$ then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightsquigarrow N(0, v). \quad (\text{B.3})$$

The complete proof is given in [51]. The theorem gives an analytical formula for the asymptotic variance and it is possible to prove that this variance exists and is finite.

The variance of a dependent time series can be computed: Let X_1, \dots, X_n be a dependent time series. It is assumed that the time series is strictly stationary. We denote \bar{X}_n the mean of the time series.

$$\text{var}(\sqrt{n}\bar{X}_n) = \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n \text{cov}(X_s, X_t).$$

We change the order of summation:

$$\text{var}(\sqrt{n}\bar{X}_n) = \frac{1}{n} \sum_{s=1}^n \sum_{h=-s+1}^{n-s} \text{cov}(X_s, X_{s+h}).$$

The covariance does not depend on s because the time series is strictly stationary.

$$\text{var}(\sqrt{n}\bar{X}_n) = \sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \text{cov}(X_s, X_{s+h}), \quad \forall s. \quad (\text{B.4})$$

We take the limit then we obtain the variance of Theorem B.1.1:

$$v = \sum_h \text{cov}(X_s, X_{s+h}), \quad \forall s. \quad (\text{B.5})$$

Property B.1.1 Under the conditions of Theorem B.1.1, the sum of Equation (B.5) is absolutely convergent.

Theorem B.1.2

Let X_1, \dots, X_n be a sample from i.i.d. random variables with cumulative distribution function F . Let \hat{F}_n be the empirical cumulative distribution function then, we have:

$$\hat{F}_n(t) \longrightarrow F(x) \text{ almost surely,} \quad (\text{B.6})$$

$$\sqrt{n}(\hat{F}_n(t) - F(t)) \rightarrow N(0, F(t)(1 - F(t))). \quad (\text{B.7})$$

Let (X_1, \dots, X_n) be a sample from identically distributed random variables with cumulative distribution function F , autocorrelated time series with empirical cumulative distribution \hat{F}_n if $\sum_k \alpha(k)^{3/4-r} < \infty$ for some $r > 2$ (α – mixing terms of $(\mathbb{I}_{X_1 \leq t}, \dots, \mathbb{I}_{X_n \leq t})$) then, the following holds:

$$\hat{F}_n(t) \longrightarrow F(x) \text{ almost surely,} \quad (\text{B.8})$$

and,

$$\sqrt{n}(\hat{F}_n(t) - F(t)) \rightarrow N(0, F(t)(1 - F(t)) + 2 \sum_{h=1}^{\infty} \text{cov}(\mathbb{I}_{X_1 \leq t}, \mathbb{I}_{X_{1+h} \leq t})). \quad (\text{B.9})$$

Proof:

The result (B.6) is a simple application of the strong law of large numbers and B.7 is an application of the central limit theorem to B.1. \square

Result (B.9) is an application of Theorem B.1.1 to the time series : $(\mathbb{I}_{X_1 \leq t}, \dots, \mathbb{I}_{X_n \leq t})$.

We notice the condition $E[\mathbb{I}_{X_1 \leq t}] < \infty$ is not necessary because:

$$|E[\mathbb{I}_{X_1 \leq t}]| = \int_{-\infty}^{+\infty} |\mathbb{I}_{X_1 \leq t}| dF(t) \leq \int_{-\infty}^{+\infty} dF(t) = 1.$$

Therefore, the conditions of the theorem are satisfied. \square

B.2 Distribution of the Order Statistics

This section provides a theorem that gives the distribution of the order statistics. However, under a general distribution (even classical: Normal, Student...) the complete distribution is difficult to compute because of the computational cost.

Theorem B.2.1 ([17])

Let X_1, \dots, X_n be independent random variables with a common continuous distribution F_X . Let $X_{1:n}, \dots, X_{n:n}$ be their order statistics. For each x , let $N_n(x)$ be the number i such that $X_i \leq x$. Then, $N_n(x)$ is a binomial random variable with parameters n and $F_X(x)$. Furthermore, the following holds:

$$P(X_{i:n} \leq x) = P(N_n(x) \geq i). \quad (\text{B.10})$$

Therefore, we can write:

$$P(X_{i:n} \leq x) = \sum_{k=i}^n C_n^k F(x)^k (1 - F(x))^{n-k}. \quad (\text{B.11})$$

B.3 Expected Shortfall and Value at Risk of Classical Distributions

B.3.1 Normal distribution

The density of a normal distribution with mean μ and variance σ^2 is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}. \quad (\text{B.12})$$

In this section, we assume that X is a normally distributed with mean μ and variance σ^2 . Let ϕ be the cumulative distribution function and f the probability density function. The VaR is given by the inverse cumulative distribution function:

$$\text{VaR}_\alpha(X) = \mu + \sigma \cdot \phi^{-1}(\alpha). \quad (\text{B.13})$$

In the above formula, the symmetry of the normal distribution is used. The ES is computed in Section 2.1.3 for a zeros mean random variable.

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \frac{\sigma}{\sqrt{2\pi}} e^{\frac{-(\text{VaR}_\alpha(X)-\mu)^2}{2\sigma^2}} + \mu. \quad (\text{B.14})$$

B.3.2 Student's t-distribution

The density of a Student's t-distribution with ν degrees of freedom is given by:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \forall x \in \mathbb{R}, \quad (\text{B.15})$$

where Γ is the gamma function given by:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad \forall z \in \mathbb{R}^+.$$

Assume X follows a Student t-distribution with ν degrees of freedom. For small $\nu \in \mathbb{N}$ the VaR may be computed directly. This is shown in [46]. However, no general closed formula can be given for $\mu > 6$. The analytical formulas are given for $\mu = 4, 5, 6$.

$$\begin{aligned} \mu & -\text{VaR}_\alpha, \\ 4 & \frac{1}{2} + \frac{x(x^2 + 6)}{2(x^2 + 4)^{3/2}}, \\ 5 & \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x}{\sqrt{5}}\right) + \frac{\sqrt{5} x(3x^2 + 25)}{3\pi(x^2 + 5)^2}, \\ 6 & \frac{1}{2} + \frac{x(2x^4 + 30x^2 + 135)}{4(x^2 + 6)^{5/2}}. \end{aligned} \quad (\text{B.16})$$

As for the normal random variables, the ES has a closed form expression but the result still needs a numerical integration. The results are explained in detail in [7].

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \frac{\nu^{-1/2}}{B(\nu/2, 1/2)} \left(1 + \frac{1}{\nu}\right)^{(\nu+1)/2} \left(\frac{\nu+1}{\nu-1}\right), \quad (\text{B.17})$$

where $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ with $a, b \in \mathbb{R}$.

This result may be adapted for Student's t-distribution (also called t-Location scale distribution) with non-zero location parameter and scale parameter different than one (see [7]).

Appendix C

Tables from Chapter 4

C.1 Tables for Uncorrelated Time Series

Table C.1: VaR^U Normal

VaR^U	Normal	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	-0.3%	1.1%	2.1%	-1.1%	-0.3%	0.6%	1.4%	-0.9%	-0.2%
	SD (%)	11.0%	10.5%	10.2%	9.4%	9.4%	9.4%	9.5%	9.2%	9.3%
	sMSE (%)	11.0%	10.5%	10.4%	9.5%	9.4%	9.4%	9.6%	9.2%	9.3%
	Theo std	11.2%	10.4%	10.0%	9.7%	9.6%	9.5%	9.4%	9.4%	9.4%
n=500	bias (%)	1.6%	-0.2%	1.0%	-0.1%	-1.0%	-0.1%	0.6%	-0.1%	0.5%
	SD (%)	8.1%	7.2%	7.1%	6.8%	6.6%	6.6%	6.6%	6.5%	6.6%
	sMSE (%)	8.3%	7.2%	7.2%	6.8%	6.7%	6.6%	6.6%	6.5%	6.6%
	Theo std	7.9%	7.4%	7.1%	6.9%	6.8%	6.7%	6.6%	6.6%	6.6%
n=1000	bias (%)	0.8%	0.5%	0.4%	0.4%	-0.5%	0.3%	0.3%	-0.4%	0.3%
	SD (%)	5.6%	5.2%	5.0%	4.9%	4.7%	4.7%	4.7%	4.7%	4.7%
	sMSE (%)	5.7%	5.3%	5.1%	4.9%	4.8%	4.7%	4.7%	4.7%	4.7%
	Theo std	5.6%	5.2%	5.0%	4.9%	4.8%	4.7%	4.7%	4.7%	4.7%

Table C.2: VaR^U t-student 4 degrees

VaR^U	t-4	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	1.0%	3.4%	4.7%	-1.0%	0.4%	1.6%	2.6%	-0.8%	0.2%
	SD (%)	21.8%	19.8%	18.3%	15.6%	15.1%	14.7%	14.6%	13.6%	13.6%
	sMSE (%)	21.9%	20.0%	18.9%	15.6%	15.1%	14.8%	14.8%	13.7%	13.6%
	Theo std	21.1%	18.3%	16.7%	15.7%	14.9%	14.4%	14.0%	13.7%	13.5%
n=500	bias (%)	4.6%	0.6%	2.5%	0.4%	-1.0%	0.3%	1.5%	0.2%	1.2%
	SD (%)	16.8%	13.3%	12.5%	11.2%	10.5%	10.3%	10.2%	9.8%	9.7%
	sMSE (%)	17.4%	13.3%	12.8%	11.3%	10.5%	10.3%	10.3%	9.8%	9.8%
	Theo std	14.9%	13.0%	11.8%	11.1%	10.6%	10.2%	9.9%	9.7%	9.5%
n=1000	bias (%)	2.2%	1.6%	1.3%	1.0%	-0.5%	0.8%	0.7%	-0.4%	0.6%
	SD (%)	11.1%	9.4%	8.6%	8.0%	7.5%	7.3%	7.1%	6.8%	6.8%
	sMSE (%)	11.3%	9.6%	8.7%	8.1%	7.5%	7.4%	7.2%	6.8%	6.8%
	Theo std	10.6%	9.2%	8.4%	7.8%	7.5%	7.2%	7.0%	6.9%	6.7%

Table C.3: VaR^U t-student 5 degrees

VaR^U	t-5	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	0.6%	2.8%	4.1%	-1.1%	0.2%	1.4%	2.3%	-0.7%	0.3%
	SD (%)	19.1%	17.3%	16.3%	13.9%	13.7%	13.5%	13.3%	12.6%	12.6%
	sMSE (%)	19.1%	17.5%	16.8%	14.0%	13.7%	13.5%	13.5%	12.6%	12.6%
	Theo std	18.7%	16.4%	15.1%	14.2%	13.6%	13.2%	12.9%	12.7%	12.5%
n=500	bias (%)	3.5%	0.3%	2.0%	0.2%	-1.1%	0.1%	1.2%	0.1%	1.1%
	SD (%)	14.4%	11.7%	11.1%	10.1%	9.5%	9.4%	9.3%	9.0%	9.0%
	sMSE (%)	14.8%	11.7%	11.3%	10.1%	9.6%	9.4%	9.4%	9.0%	9.1%
	Theo std	13.2%	11.6%	10.7%	10.1%	9.7%	9.3%	9.1%	9.0%	8.8%
n=1000	bias (%)	1.7%	1.2%	0.9%	0.8%	-0.6%	0.6%	0.5%	-0.5%	0.4%
	SD (%)	9.7%	8.4%	7.7%	7.3%	6.7%	6.6%	6.4%	6.3%	6.2%
	sMSE (%)	9.9%	8.5%	7.8%	7.3%	6.7%	6.6%	6.5%	6.3%	6.2%
	Theo std	9.4%	8.2%	7.6%	7.1%	6.8%	6.6%	6.5%	6.3%	6.2%

Table C.4: VaR^U t-student 6 degrees

VaR^U	t-6	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	0.3%	2.2%	3.5%	-1.3%	0.1%	1.1%	2.1%	-0.8%	0.1%
	SD (%)	17.7%	16.0%	15.0%	13.1%	13.0%	12.8%	12.6%	12.0%	12.0%
	sMSE (%)	17.7%	16.2%	15.4%	13.1%	13.0%	12.8%	12.8%	12.0%	12.0%
	Theo std	17.3%	15.2%	14.1%	13.4%	12.9%	12.5%	12.2%	12.0%	11.9%
n=500	bias (%)	3.0%	0.0%	1.6%	0.0%	-1.2%	0.0%	1.0%	0.1%	0.9%
	SD (%)	13.2%	10.8%	10.3%	9.5%	9.0%	8.9%	8.9%	8.6%	8.6%
	sMSE (%)	13.6%	10.8%	10.4%	9.5%	9.0%	8.9%	8.9%	8.6%	8.6%
	Theo std	12.2%	10.8%	10.0%	9.5%	9.1%	8.8%	8.6%	8.5%	8.4%
n=1000	bias (%)	1.5%	1.1%	0.9%	0.8%	-0.5%	0.6%	0.5%	-0.4%	0.5%
	SD (%)	9.0%	7.9%	7.2%	6.8%	6.5%	6.4%	6.3%	6.1%	6.0%
	sMSE (%)	9.1%	7.9%	7.2%	6.9%	6.5%	6.4%	6.3%	6.1%	6.1%
	Theo std	8.6%	7.6%	7.1%	6.7%	6.4%	6.2%	6.1%	6.0%	5.9%

Table C.5: VaR^L Normal

VaR^L	Normal	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	7.9%	7.5%	2.0%	2.9%	3.5%	4.1%	1.4%	2.0%	2.6%
	SD (%)	13.1%	11.9%	10.3%	10.1%	10.0%	9.9%	9.5%	9.5%	9.6%
	sMSE (%)	15.3%	14.1%	10.4%	10.5%	10.6%	10.7%	9.6%	9.7%	9.9%
	Theo std	11.2%	10.4%	10.0%	9.7%	9.6%	9.5%	9.4%	9.4%	9.4%
n=500	bias (%)	1.6%	2.7%	1.0%	1.9%	0.8%	1.5%	0.6%	1.3%	0.6%
	SD (%)	8.2%	7.8%	7.2%	7.0%	6.8%	6.8%	6.7%	6.7%	6.7%
	sMSE (%)	8.3%	8.2%	7.2%	7.3%	6.9%	7.0%	6.7%	6.8%	6.7%
	Theo std	7.9%	7.4%	7.1%	6.9%	6.8%	6.7%	6.6%	6.6%	6.6%
n=1000	bias (%)	0.7%	0.6%	0.4%	0.4%	0.3%	0.3%	0.3%	0.2%	0.2%
	SD (%)	5.7%	5.2%	5.0%	4.9%	4.8%	4.8%	4.7%	4.7%	4.7%
	sMSE (%)	5.7%	5.3%	5.1%	4.9%	4.8%	4.8%	4.7%	4.7%	4.7%
	Theo std	5.6%	5.2%	5.0%	4.9%	4.8%	4.7%	4.7%	4.7%	4.7%

Table C.6: VaR^L t-student 4 degrees

VaR^L	t-4	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	20.3%	16.4%	4.6%	5.6%	6.4%	7.0%	2.5%	3.4%	4.2%
	SD (%)	34.0%	25.7%	18.4%	17.4%	16.6%	16.1%	14.6%	14.4%	14.3%
	sMSE (%)	39.6%	30.5%	18.9%	18.3%	17.8%	17.5%	14.8%	14.8%	14.9%
	Theo std	21.1%	18.3%	16.7%	15.7%	14.9%	14.4%	14.0%	13.7%	13.5%
n=500	bias (%)	4.3%	5.7%	2.2%	3.5%	1.6%	2.6%	1.2%	2.2%	1.1%
	SD (%)	16.6%	14.6%	12.3%	11.8%	10.9%	10.6%	10.1%	10.0%	9.7%
	sMSE (%)	17.2%	15.7%	12.5%	12.3%	11.0%	10.9%	10.2%	10.3%	9.8%
	Theo std	14.9%	13.0%	11.8%	11.1%	10.6%	10.2%	9.9%	9.7%	9.5%
n=1000	bias (%)	2.0%	1.5%	1.2%	1.0%	0.9%	0.8%	0.7%	0.6%	0.6%
	SD (%)	11.0%	9.5%	8.5%	7.9%	7.4%	7.2%	7.0%	6.9%	6.7%
	sMSE (%)	11.2%	9.6%	8.6%	8.0%	7.5%	7.2%	7.0%	6.9%	6.8%
	Theo std	10.6%	9.2%	8.4%	7.8%	7.5%	7.2%	7.0%	6.9%	6.7%

Table C.7: VaR^L t-student 5 degrees

VaR^L	t-5	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	16.9%	14.3%	4.2%	5.1%	5.7%	6.5%	2.5%	3.4%	4.1%
	SD (%)	27.0%	22.0%	16.7%	15.8%	15.2%	14.7%	13.4%	13.4%	13.2%
	sMSE (%)	31.8%	26.2%	17.2%	16.6%	16.2%	16.1%	13.6%	13.8%	13.8%
	Theo std	18.7%	16.4%	15.1%	14.2%	13.6%	13.2%	12.9%	12.7%	12.5%
n=500	bias (%)	3.6%	4.9%	1.9%	3.2%	1.4%	2.4%	1.1%	2.0%	0.9%
	SD (%)	14.5%	12.8%	11.1%	10.6%	9.9%	9.6%	9.2%	9.1%	8.9%
	sMSE (%)	14.9%	13.7%	11.2%	11.1%	10.0%	9.9%	9.3%	9.3%	8.9%
	Theo std	13.2%	11.6%	10.7%	10.1%	9.7%	9.3%	9.1%	9.0%	8.8%
n=1000	bias (%)	1.8%	1.3%	1.0%	0.8%	0.7%	0.6%	0.5%	0.5%	0.4%
	SD (%)	9.8%	8.5%	7.7%	7.2%	6.9%	6.7%	6.5%	6.4%	6.3%
	sMSE (%)	9.9%	8.6%	7.8%	7.3%	7.0%	6.7%	6.5%	6.4%	6.3%
	Theo std	9.4%	8.2%	7.6%	7.1%	6.8%	6.6%	6.5%	6.3%	6.2%

Table C.8: VaR^L t-student 6 degrees

VaR^L	t-5	99.0%	98.5%	98.0%	97.5%	97.0%	96.5%	96.0%	95.5%	95.0%
n=250	bias (%)	16.9%	14.3%	4.2%	5.1%	5.7%	6.5%	2.5%	3.4%	4.1%
	SD (%)	27.0%	22.0%	16.7%	15.8%	15.2%	14.7%	13.4%	13.4%	13.2%
	sMSE (%)	31.8%	26.2%	17.2%	16.6%	16.2%	16.1%	13.6%	13.8%	13.8%
	Theo std	18.7%	16.4%	15.1%	14.2%	13.6%	13.2%	12.9%	12.7%	12.5%
n=500	bias (%)	3.6%	4.9%	1.9%	3.2%	1.4%	2.4%	1.1%	2.0%	0.9%
	SD (%)	14.5%	12.8%	11.1%	10.6%	9.9%	9.6%	9.2%	9.1%	8.9%
	sMSE (%)	14.9%	13.7%	11.2%	11.1%	10.0%	9.9%	9.3%	9.3%	8.9%
	Theo std	13.2%	11.6%	10.7%	10.1%	9.7%	9.3%	9.1%	9.0%	8.8%
n=1000	bias (%)	1.8%	1.3%	1.0%	0.8%	0.7%	0.6%	0.5%	0.5%	0.4%
	SD (%)	9.8%	8.5%	7.7%	7.2%	6.9%	6.7%	6.5%	6.4%	6.3%
	sMSE (%)	9.9%	8.6%	7.8%	7.3%	7.0%	6.7%	6.5%	6.4%	6.3%
	Theo std	9.4%	8.2%	7.6%	7.1%	6.8%	6.6%	6.5%	6.3%	6.2%

Table C.9: ES^U Normal

ES^U	Normal	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	-4.8%	-1.6%	-2.4%	-1.0%	-1.7%	-0.8%	-1.4%	-0.6%	-1.2%
	SD (%)	10.7%	9.8%	8.9%	8.6%	8.3%	8.2%	8.0%	8.0%	7.9%
	sMSE (%)	11.8%	10.0%	9.2%	8.7%	8.4%	8.2%	8.1%	8.0%	8.0%
	Theo std	11.8%	10.1%	9.3%	8.8%	8.6%	8.4%	8.3%	8.2%	8.2%
n=500	bias (%)	-1.5%	-0.9%	-0.7%	-0.6%	-0.5%	-0.5%	-0.4%	-0.4%	-0.3%
	SD (%)	8.2%	7.0%	6.5%	6.2%	6.0%	5.8%	5.7%	5.6%	5.6%
	sMSE (%)	8.3%	7.1%	6.5%	6.2%	6.0%	5.8%	5.7%	5.7%	5.6%
	Theo std	8.4%	7.1%	6.6%	6.3%	6.1%	5.9%	5.9%	5.8%	5.8%
n=1000	bias (%)	-0.6%	-0.3%	-0.2%	-0.2%	-0.1%	-0.1%	-0.1%	-0.1%	-0.1%
	SD (%)	5.9%	5.0%	4.6%	4.4%	4.2%	4.1%	4.0%	4.0%	4.0%
	sMSE (%)	5.9%	5.0%	4.6%	4.4%	4.2%	4.1%	4.0%	4.0%	4.0%
	Theo std	5.9%	5.0%	4.6%	4.4%	4.3%	4.2%	4.1%	4.1%	4.1%

Table C.10: ES^U t-student 4 degrees

ES^U	t-4	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	-8.8%	-2.4%	-3.8%	-1.4%	-2.5%	-1.0%	-2.0%	-0.8%	-1.7%
	SD (%)	27.5%	23.9%	19.7%	18.6%	16.8%	16.2%	15.2%	14.9%	14.3%
	sMSE (%)	28.8%	24.0%	20.0%	18.6%	16.9%	16.3%	15.3%	15.0%	14.4%
	Theo std	30.6%	23.3%	20.1%	18.2%	16.9%	16.0%	15.3%	14.8%	14.4%
n=500	bias (%)	-1.9%	-1.0%	-0.7%	-0.5%	-0.4%	-0.3%	-0.3%	-0.3%	-0.2%
	SD (%)	21.8%	16.6%	14.3%	12.9%	12.0%	11.3%	10.8%	10.4%	10.1%
	sMSE (%)	21.9%	16.6%	14.3%	12.9%	12.0%	11.3%	10.8%	10.4%	10.1%
	Theo std	22.3%	16.9%	14.5%	13.1%	12.1%	11.4%	10.9%	10.5%	10.1%
n=1000	bias (%)	-1.0%	-0.5%	-0.3%	-0.2%	-0.1%	-0.1%	-0.1%	-0.1%	0.0%
	SD (%)	15.8%	11.9%	10.2%	9.2%	8.5%	8.0%	7.6%	7.4%	7.1%
	sMSE (%)	15.8%	11.9%	10.2%	9.2%	8.5%	8.0%	7.6%	7.4%	7.1%
	Theo std	15.9%	12.0%	10.3%	9.2%	8.6%	8.1%	7.7%	7.4%	7.2%

Table C.11: ES^U t-student 5 degrees

ES^U	t-5	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	-7.4%	-1.8%	-3.2%	-1.0%	-2.2%	-0.8%	-1.7%	-0.6%	-1.4%
	SD (%)	22.4%	19.5%	16.3%	15.4%	14.1%	13.7%	12.9%	12.7%	12.2%
	sMSE (%)	23.6%	19.6%	16.6%	15.5%	14.2%	13.7%	13.0%	12.7%	12.3%
	Theo std	25.4%	19.6%	17.1%	15.6%	14.6%	13.9%	13.3%	12.9%	12.6%
n=500	bias (%)	-2.1%	-1.2%	-0.8%	-0.7%	-0.6%	-0.5%	-0.4%	-0.4%	-0.4%
	SD (%)	17.8%	13.7%	11.9%	10.9%	10.2%	9.6%	9.3%	9.0%	8.7%
	sMSE (%)	17.9%	13.8%	12.0%	10.9%	10.2%	9.6%	9.3%	9.0%	8.7%
	Theo std	18.0%	13.9%	12.1%	11.0%	10.3%	9.8%	9.4%	9.1%	8.9%
n=1000	bias (%)	-1.2%	-0.7%	-0.5%	-0.4%	-0.3%	-0.3%	-0.2%	-0.2%	-0.2%
	SD (%)	12.7%	9.7%	8.5%	7.7%	7.2%	6.8%	6.6%	6.3%	6.2%
	sMSE (%)	12.7%	9.8%	8.5%	7.7%	7.2%	6.8%	6.6%	6.3%	6.2%
	Theo std	12.7%	9.8%	8.5%	7.8%	7.3%	6.9%	6.7%	6.5%	6.3%

Table C.12: ES^U t-student 6 degrees

ES^U	t-6	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	-7.0%	-1.8%	-3.0%	-1.0%	-2.0%	-0.7%	-1.6%	-0.6%	-1.3%
	SD (%)	19.5%	17.2%	14.6%	13.9%	12.7%	12.4%	11.8%	11.6%	11.2%
	sMSE (%)	20.8%	17.3%	14.9%	13.9%	12.9%	12.4%	11.9%	11.6%	11.3%
	Theo std	22.4%	17.5%	15.3%	14.1%	13.2%	12.6%	12.2%	11.9%	11.6%
n=500	bias (%)	-1.8%	-1.0%	-0.7%	-0.6%	-0.5%	-0.4%	-0.4%	-0.4%	-0.3%
	SD (%)	16.0%	12.5%	10.9%	9.9%	9.3%	8.8%	8.5%	8.2%	8.0%
	sMSE (%)	16.1%	12.5%	10.9%	9.9%	9.3%	8.8%	8.5%	8.2%	8.0%
	Theo std	15.9%	12.4%	10.8%	10.0%	9.4%	8.9%	8.6%	8.4%	8.2%
n=1000	bias (%)	-0.9%	-0.5%	-0.3%	-0.3%	-0.2%	-0.2%	-0.2%	-0.2%	-0.1%
	SD (%)	11.2%	8.7%	7.6%	7.0%	6.5%	6.2%	6.0%	5.8%	5.6%
	sMSE (%)	11.3%	8.7%	7.6%	7.0%	6.5%	6.2%	6.0%	5.8%	5.6%
	Theo std	11.2%	8.8%	7.7%	7.0%	6.6%	6.3%	6.1%	5.9%	5.8%

Table C.13: ES^V Normal

ES^V	Normal	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	-0.2%	-1.8%	-0.1%	-1.1%	-0.1%	-0.9%	-0.1%	-0.8%	-0.1%
	SD (%)	12.5%	9.9%	9.3%	8.7%	8.5%	8.2%	8.2%	8.0%	8.0%
	sMSE (%)	12.5%	10.0%	9.3%	8.8%	8.5%	8.3%	8.2%	8.0%	8.0%
	Theo std	11.8%	10.1%	9.3%	8.8%	8.6%	8.4%	8.3%	8.2%	8.2%
n=500	bias (%)	-1.4%	-0.8%	-0.6%	-0.5%	0.4%	-0.4%	0.4%	-0.3%	-0.3%
	SD (%)	8.2%	7.0%	6.5%	6.1%	6.0%	5.8%	5.8%	5.7%	5.6%
	sMSE (%)	8.3%	7.0%	6.5%	6.2%	6.1%	5.8%	5.8%	5.7%	5.6%
	Theo std	8.4%	7.1%	6.6%	6.3%	6.1%	5.9%	5.9%	5.8%	5.8%
n=1000	bias (%)	-0.7%	-0.4%	-0.3%	-0.2%	0.3%	-0.2%	0.2%	-0.2%	-0.1%
	SD (%)	5.9%	5.0%	4.6%	4.4%	4.2%	4.1%	4.1%	4.0%	4.0%
	sMSE (%)	5.9%	5.0%	4.6%	4.4%	4.2%	4.1%	4.1%	4.0%	4.0%
	Theo std	5.9%	5.0%	4.6%	4.4%	4.3%	4.2%	4.1%	4.1%	4.1%

Table C.14: ES^L t-student 4 degrees

ES^L	t-4	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	1.4%	-2.4%	0.4%	-1.4%	0.2%	-1.1%	0.1%	-1.0%	0.0%
	SD (%)	37.2%	24.3%	21.9%	18.8%	17.9%	16.4%	16.0%	15.1%	14.8%
	sMSE (%)	37.3%	24.4%	21.9%	18.8%	17.9%	16.5%	16.0%	15.1%	14.8%
	Theo std	30.6%	23.3%	20.1%	18.2%	16.9%	16.0%	15.3%	14.8%	14.4%
n=500	bias (%)	-2.1%	-1.2%	-0.8%	-0.7%	0.9%	-0.5%	0.7%	-0.4%	-0.4%
	SD (%)	21.9%	16.7%	14.3%	12.9%	12.3%	11.3%	11.0%	10.4%	10.1%
	sMSE (%)	22.0%	16.7%	14.4%	12.9%	12.4%	11.3%	11.1%	10.4%	10.1%
	Theo std	22.3%	16.9%	14.5%	13.1%	12.1%	11.4%	10.9%	10.5%	10.1%
n=1000	bias (%)	-0.9%	-0.5%	-0.4%	-0.3%	0.5%	-0.3%	0.3%	-0.2%	-0.2%
	SD (%)	15.8%	11.9%	10.2%	9.2%	8.7%	8.1%	7.8%	7.4%	7.2%
	sMSE (%)	15.8%	11.9%	10.2%	9.2%	8.7%	8.1%	7.8%	7.4%	7.2%
	Theo std	15.9%	12.0%	10.3%	9.2%	8.6%	8.1%	7.7%	7.4%	7.2%

Table C.15: ES^L t-student 5 degrees

ES^L	t-5	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	0.8%	-2.2%	0.5%	-1.2%	0.4%	-0.8%	0.3%	-0.6%	0.3%
	SD (%)	28.5%	19.4%	17.7%	15.4%	14.8%	13.7%	13.4%	12.8%	12.6%
	sMSE (%)	28.5%	19.5%	17.7%	15.5%	14.8%	13.8%	13.4%	12.8%	12.6%
	Theo std	25.4%	19.6%	17.1%	15.6%	14.6%	13.9%	13.3%	12.9%	12.6%
n=500	bias (%)	-2.1%	-1.2%	-0.9%	-0.7%	0.7%	-0.5%	0.6%	-0.5%	-0.4%
	SD (%)	18.1%	14.0%	12.1%	11.0%	10.6%	9.8%	9.5%	9.1%	8.8%
	sMSE (%)	18.2%	14.0%	12.2%	11.1%	10.6%	9.8%	9.6%	9.1%	8.8%
	Theo std	18.0%	13.9%	12.1%	11.0%	10.3%	9.8%	9.4%	9.1%	8.9%
n=1000	bias (%)	-0.9%	-0.5%	-0.3%	-0.3%	0.4%	-0.2%	0.3%	-0.2%	-0.2%
	SD (%)	13.1%	10.0%	8.7%	7.9%	7.4%	6.9%	6.7%	6.4%	6.3%
	sMSE (%)	13.2%	10.1%	8.7%	7.9%	7.4%	7.0%	6.7%	6.4%	6.3%
	Theo std	12.7%	9.8%	8.5%	7.8%	7.3%	6.9%	6.7%	6.5%	6.3%

Table C.16: ES^L t-student 6 degrees

ES^L	t-6	99.0%	98.0%	97.0%	96.0%	95.0%	94.0%	93.0%	92.0%	91.0%
n=250	bias (%)	0.8%	-1.9%	0.6%	-0.9%	0.5%	-0.6%	0.5%	-0.4%	0.5%
	SD (%)	25.2%	17.2%	15.7%	13.8%	13.2%	12.3%	12.1%	11.5%	11.4%
	sMSE (%)	25.2%	17.3%	15.7%	13.8%	13.2%	12.3%	12.1%	11.5%	11.4%
	Theo std	22.4%	17.5%	15.3%	14.1%	13.2%	12.6%	12.2%	11.9%	11.6%
n=500	bias (%)	-2.0%	-1.2%	-0.9%	-0.7%	0.7%	-0.5%	0.5%	-0.4%	-0.4%
	SD (%)	15.6%	12.3%	10.7%	9.8%	9.5%	8.8%	8.6%	8.2%	8.0%
	sMSE (%)	15.8%	12.3%	10.8%	9.9%	9.5%	8.8%	8.6%	8.2%	8.0%
	Theo std	15.9%	12.4%	10.8%	10.0%	9.4%	8.9%	8.6%	8.4%	8.2%
n=1000	bias (%)	-0.8%	-0.4%	-0.3%	-0.3%	0.4%	-0.2%	0.3%	-0.2%	-0.2%
	SD (%)	11.4%	8.8%	7.6%	7.0%	6.6%	6.2%	6.0%	5.8%	5.7%
	sMSE (%)	11.4%	8.8%	7.6%	7.0%	6.6%	6.2%	6.0%	5.8%	5.7%
	Theo std	11.2%	8.8%	7.7%	7.0%	6.6%	6.3%	6.1%	5.9%	5.8%

C.2 Tables for Correlated Time Series

Table C.17: T5 Var^L

	T5 Var^L	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	7.0%	0.1%	2.9%	0.5%	2.1%	-0.1%	1.3%	-0.3%	1.0%
	SD (%)	37.5%	30.5%	31.1%	29.5%	30.0%	29.6%	30.4%	30.3%	31.2%
	sMSE (%)	38.2%	30.5%	31.2%	29.5%	30.1%	29.6%	30.4%	30.2%	31.1%
n=500	bias (%)	2.5%	1.6%	0.9%	0.5%	2.4%	0.4%	1.9%	0.7%	0.6%
	SD (%)	29.1%	24.3%	22.4%	22.1%	22.4%	22.2%	22.7%	22.9%	23.3%
	sMSE (%)	29.2%	24.4%	22.4%	22.0%	22.5%	22.2%	22.7%	22.9%	23.3%
n=1000	bias (%)	0.6%	0.3%	-0.1%	-0.1%	0.9%	-0.5%	0.2%	-0.4%	-0.8%
	SD (%)	19.0%	16.9%	15.6%	15.3%	15.7%	15.4%	15.7%	16.0%	16.2%
	sMSE (%)	19.0%	16.8%	15.6%	15.3%	15.7%	15.4%	15.7%	16.0%	16.2%

Table C.18: AR Var^L

	AR Var^L	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	0.4%	-1.7%	0.4%	-0.8%	0.6%	-0.7%	0.5%	-0.5%	0.7%
	SD (%)	18.6%	18.1%	18.7%	19.0%	19.6%	20.2%	21.0%	21.5%	22.2%
	sMSE (%)	18.6%	18.1%	18.6%	19.0%	19.6%	20.2%	21.0%	21.5%	22.2%
n=500	bias (%)	-0.3%	0.0%	0.5%	0.5%	1.8%	0.4%	1.5%	0.6%	0.6%
	SD (%)	13.6%	13.1%	13.3%	13.5%	14.1%	14.1%	14.6%	15.0%	15.3%
	sMSE (%)	13.6%	13.0%	13.3%	13.5%	14.2%	14.1%	14.7%	15.0%	15.3%
n=1000	bias (%)	-0.1%	0.1%	0.2%	0.2%	0.8%	0.3%	0.9%	0.4%	0.4%
	SD (%)	10.7%	10.1%	10.0%	10.1%	10.2%	10.3%	10.6%	10.9%	11.1%
	sMSE (%)	10.6%	10.0%	10.0%	10.1%	10.3%	10.3%	10.6%	10.9%	11.1%

Table C.19: AR VaR^U

	AR VaR^U	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-4.8%	-1.7%	-3.2%	-1.9%	-2.8%	-1.7%	-2.6%	-1.7%	-2.8%
	SD (%)	18.3%	19.0%	18.6%	18.9%	19.2%	19.6%	20.0%	20.6%	21.0%
	sMSE (%)	18.9%	19.0%	18.8%	19.0%	19.4%	19.7%	20.1%	20.6%	21.2%
n=500	bias (%)	-1.0%	-0.6%	-0.6%	-0.7%	-0.6%	-0.6%	-0.5%	-0.6%	-0.6%
	SD (%)	13.2%	12.7%	13.0%	13.1%	13.4%	13.7%	14.0%	14.3%	14.6%
	sMSE (%)	13.3%	12.7%	13.0%	13.1%	13.4%	13.7%	14.0%	14.3%	14.6%
n=1000	bias (%)	-0.8%	-0.8%	-0.8%	-0.8%	-0.5%	-0.6%	-0.7%	-0.7%	-0.7%
	SD (%)	9.9%	9.4%	9.2%	9.4%	9.6%	9.7%	9.8%	10.1%	10.4%
	sMSE (%)	9.9%	9.4%	9.3%	9.4%	9.6%	9.7%	9.8%	10.1%	10.4%

Table C.20: AR ES^L

	AR ES^L	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-8.0%	-7.4%	-5.4%	-5.7%	-4.6%	-4.9%	-4.1%	-4.4%	-3.7%
	SD (%)	16.7%	16.6%	16.9%	16.9%	17.1%	17.2%	17.4%	17.5%	17.7%
	sMSE (%)	18.5%	18.2%	17.7%	17.8%	17.7%	17.8%	17.9%	18.0%	18.1%
n=500	bias (%)	-4.9%	-3.5%	-3.0%	-2.7%	-1.7%	-2.3%	-1.6%	-2.2%	-2.1%
	SD (%)	13.9%	13.6%	13.4%	13.4%	13.4%	13.3%	13.4%	13.4%	13.5%
	sMSE (%)	14.7%	14.1%	13.8%	13.6%	13.5%	13.5%	13.5%	13.6%	13.6%
n=1000	bias (%)	-2.7%	-1.9%	-1.5%	-1.3%	-0.8%	-1.1%	-0.7%	-0.9%	-0.9%
	SD (%)	10.2%	9.6%	9.4%	9.3%	9.2%	9.2%	9.2%	9.2%	9.2%
	sMSE (%)	10.5%	9.8%	9.5%	9.4%	9.3%	9.2%	9.2%	9.2%	9.2%

Table C.21: AR ES^U

	AR ES^U	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-10.3%	-6.6%	-6.5%	-5.1%	-5.4%	-4.5%	-4.9%	-4.2%	-4.6%
	SD (%)	17.4%	17.9%	17.4%	17.3%	17.0%	17.0%	16.9%	17.0%	16.9%
	sMSE (%)	20.2%	19.0%	18.5%	18.1%	17.8%	17.6%	17.6%	17.5%	17.5%
n=500	bias (%)	-5.1%	-3.8%	-3.3%	-3.0%	-2.7%	-2.6%	-2.5%	-2.3%	-2.2%
	SD (%)	13.6%	13.2%	12.9%	12.8%	12.8%	12.8%	12.8%	12.9%	13.0%
	sMSE (%)	14.5%	13.7%	13.4%	13.2%	13.1%	13.1%	13.1%	13.1%	13.1%
n=1000	bias (%)	-2.1%	-1.4%	-1.1%	-1.0%	-0.8%	-0.7%	-0.6%	-0.5%	-0.5%
	SD (%)	9.7%	9.0%	8.9%	8.8%	8.8%	8.8%	8.8%	8.9%	8.9%
	sMSE (%)	9.9%	9.1%	8.9%	8.8%	8.8%	8.8%	8.8%	8.9%	8.9%

Table C.22: T5 ES^L

	T5 ES^L	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-13.9%	-12.1%	-8.3%	-8.6%	-6.6%	-7.2%	-5.8%	-6.4%	-5.4%
	SD (%)	40.3%	35.5%	35.3%	33.3%	33.0%	31.8%	31.8%	31.0%	31.1%
	sMSE (%)	42.6%	37.5%	36.3%	34.3%	33.7%	32.6%	32.3%	31.7%	31.5%
n=500	bias (%)	-8.8%	-6.1%	-4.9%	-4.1%	-2.4%	-3.3%	-2.0%	-2.8%	-2.7%
	SD (%)	30.8%	27.4%	25.6%	24.4%	24.1%	23.1%	22.9%	22.3%	22.1%
	sMSE (%)	32.0%	28.1%	26.0%	24.8%	24.2%	23.3%	23.0%	22.5%	22.2%
n=1000	bias (%)	-5.4%	-3.9%	-3.2%	-2.7%	-1.8%	-2.2%	-1.5%	-1.9%	-1.8%
	SD (%)	25.9%	21.2%	19.1%	17.8%	17.2%	16.5%	16.2%	15.8%	15.6%
	sMSE (%)	26.4%	21.5%	19.3%	18.0%	17.3%	16.6%	16.3%	15.9%	15.7%

Table C.23: T5 ES^U

	T5 ES^U	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-20.3%	-13.1%	-12.4%	-9.6%	-9.7%	-8.0%	-8.3%	-7.1%	-7.6%
	SD (%)	28.3%	29.1%	27.9%	28.2%	27.6%	27.9%	27.5%	27.8%	27.5%
	sMSE (%)	34.8%	31.9%	30.5%	29.8%	29.2%	29.0%	28.7%	28.6%	28.5%
n=500	bias (%)	-9.1%	-6.1%	-4.8%	-4.0%	-3.5%	-3.1%	-2.8%	-2.6%	-2.4%
	SD (%)	29.7%	26.7%	24.9%	23.7%	23.0%	22.5%	22.1%	21.9%	21.7%
	sMSE (%)	31.1%	27.4%	25.3%	24.1%	23.3%	22.7%	22.3%	22.0%	21.8%
n=1000	bias (%)	-3.7%	-2.9%	-2.5%	-2.2%	-2.0%	-1.8%	-1.7%	-1.6%	-1.5%
	SD (%)	24.7%	20.7%	18.8%	17.7%	17.0%	16.6%	16.3%	16.0%	15.9%
	sMSE (%)	25.0%	20.9%	19.0%	17.8%	17.1%	16.7%	16.3%	16.1%	15.9%

Table C.24: T5 VaR^U

	T5 VaR^U	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	bias (%)	-3.8%	0.4%	-2.8%	-1.2%	-3.4%	-2.0%	-4.2%	-2.8%	-4.1%
	SD (%)	41.2%	37.4%	31.9%	30.8%	29.5%	29.5%	29.4%	30.2%	30.7%
	sMSE (%)	41.4%	37.3%	32.0%	30.7%	29.7%	29.6%	29.6%	30.3%	30.9%
n=500	bias (%)	3.0%	2.5%	2.2%	1.8%	1.8%	1.4%	1.0%	1.0%	0.8%
	SD (%)	28.9%	24.1%	22.5%	21.4%	21.1%	20.7%	20.8%	21.0%	21.5%
	sMSE (%)	29.0%	24.2%	22.6%	21.5%	21.2%	20.8%	20.8%	21.0%	21.5%
n=1000	bias (%)	1.3%	0.8%	0.7%	0.7%	0.8%	0.6%	0.7%	0.7%	0.4%
	SD (%)	19.2%	16.0%	15.3%	14.9%	15.1%	15.0%	15.1%	15.1%	15.4%
	sMSE (%)	19.2%	16.0%	15.3%	14.9%	15.1%	15.0%	15.1%	15.1%	15.4%

Table C.25: Theoretical σ AR

	99%	98%	97%	96%	95%	94%	93%	92%	91%
n= 250	13.0%	12.2%	12.0%	11.8%	11.8%	11.7%	11.7%	11.7%	11.8%
n = 500	9.7%	9.1%	8.7%	8.6%	8.5%	8.4%	8.4%	8.4%	8.4%
n=1000	7.2%	6.6%	6.4%	6.2%	6.0%	6.0%	6.0%	6.0%	6.0%

Table C.26: theoretical σ T5

	99%	98%	97%	96%	95%	94%	93%	92%	91%
n=250	42.0%	32.0%	28.0%	26.0%	25.0%	23.0%	22.0%	22.0%	21.9%
n = 500	27.3%	22.0%	19.5%	18.1%	17.1%	16.5%	16.1%	15.7%	15.4%
n=1000	19.0%	15.4%	13.7%	12.7%	12.0%	11.5%	11.2%	10.9%	10.7%

Appendix D

Migration Matrices

Table D.1: 3-month Migration matrix

	1	2	3	4	5	6	7	8	9	10
1	9.7E-01	2.4E-02	4.5E-03	9.7E-04	2.4E-04	6.7E-05	2.2E-05	8.0E-06	3.4E-06	1.6E-06
2	1.3E-02	9.7E-01	1.7E-02	3.4E-03	7.5E-04	1.9E-04	5.6E-05	1.9E-05	7.1E-06	3.1E-06
3	2.0E-03	1.4E-02	9.6E-01	1.9E-02	3.7E-03	8.6E-04	2.3E-04	6.8E-05	2.3E-05	9.2E-06
4	3.7E-04	2.3E-03	1.6E-02	9.6E-01	2.0E-02	4.2E-03	1.0E-03	2.7E-04	8.5E-05	3.0E-05
5	8.1E-05	4.5E-04	2.7E-03	1.7E-02	9.5E-01	2.2E-02	4.7E-03	1.2E-03	3.3E-04	1.1E-04
6	2.1E-05	1.0E-04	5.5E-04	3.1E-03	1.9E-02	9.5E-01	2.4E-02	5.3E-03	1.3E-03	3.9E-04
7	6.4E-06	2.8E-05	1.3E-04	6.6E-04	3.6E-03	2.0E-02	9.4E-01	2.5E-02	5.8E-03	1.5E-03
8	2.3E-06	8.9E-06	3.7E-05	1.7E-04	7.9E-04	4.0E-03	2.2E-02	9.4E-01	2.7E-02	6.4E-03
9	9.7E-07	3.4E-06	1.2E-05	4.9E-05	2.1E-04	9.4E-04	4.5E-03	2.4E-02	9.3E-01	2.8E-02
10	4.8E-07	1.5E-06	4.9E-06	1.7E-05	6.4E-05	2.6E-04	1.1E-03	5.1E-03	2.5E-02	9.3E-01
11	2.8E-07	7.7E-07	2.2E-06	7.0E-06	2.3E-05	8.3E-05	3.2E-04	1.3E-03	5.7E-03	2.6E-02
12	2.0E-07	4.7E-07	1.2E-06	3.4E-06	1.0E-05	3.2E-05	1.1E-04	3.9E-04	1.5E-03	6.2E-03
13	1.6E-07	3.4E-07	7.8E-07	1.9E-06	5.0E-06	1.4E-05	4.2E-05	1.4E-04	4.7E-04	1.7E-03
14	1.5E-07	2.9E-07	5.9E-07	1.3E-06	3.0E-06	7.5E-06	2.0E-05	5.7E-05	1.7E-04	5.7E-04
15	1.5E-07	1.5E-07	5.3E-07	1.0E-06	2.1E-06	4.6E-06	1.1E-05	2.8E-05	7.5E-05	2.2E-04
16	1.5E-07	1.5E-07	1.5E-07	9.4E-07	1.7E-06	3.4E-06	7.1E-06	1.6E-05	3.8E-05	9.9E-05
17	1.4E-07	1.4E-07	1.4E-07	1.4E-07	1.4E-07	2.9E-06	5.4E-06	1.1E-05	2.3E-05	5.2E-05
18	1.3E-07	1.3E-07	1.3E-07	1.3E-07	1.3E-07	1.3E-07	4.8E-06	8.5E-06	1.6E-05	3.3E-05
19	1.3E-07	1.3E-07	1.3E-07	1.3E-07	1.3E-07	1.3E-07	1.3E-07	8.2E-06	1.4E-05	2.5E-05
20	1.6E-07	1.6E-07	1.6E-07	1.6E-07	1.6E-07	1.6E-07	1.6E-07	1.6E-07	1.6E-07	2.8E-05
D	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00

	11	12	13	14	15	16	17	18	19	20	D
1	8.9E-07	5.6E-07	4.1E-07	3.3E-07	3.1E-07	3.1E-07	3.1E-07	3.1E-07	3.1E-07	3.1E-07	7.9E-06
2	1.5E-06	8.7E-07	5.7E-07	4.2E-07	3.6E-07	3.5E-07	2.2E-07	2.2E-07	2.2E-07	2.2E-07	1.2E-05
3	4.2E-06	2.1E-06	1.3E-06	8.5E-07	6.5E-07	5.7E-07	2.3E-07	2.3E-07	2.3E-07	2.3E-07	1.8E-05
4	1.2E-05	5.7E-06	3.0E-06	1.9E-06	1.3E-06	1.0E-06	9.3E-07	2.4E-07	2.4E-07	2.4E-07	2.8E-05
5	3.9E-05	1.6E-05	7.9E-06	4.3E-06	2.7E-06	2.0E-06	1.6E-06	1.5E-06	2.6E-07	2.6E-07	4.3E-05
6	1.3E-04	4.9E-05	2.1E-05	1.1E-05	6.1E-06	3.9E-06	2.9E-06	2.5E-06	2.4E-06	2.7E-07	6.5E-05
7	4.6E-04	1.6E-04	6.2E-05	2.8E-05	1.4E-05	8.4E-06	5.7E-06	4.4E-06	3.8E-06	2.8E-07	1.0E-04
8	1.7E-03	5.4E-04	1.9E-04	7.8E-05	3.6E-05	1.9E-05	1.2E-05	8.1E-06	6.4E-06	5.8E-06	1.5E-04
9	6.9E-03	1.9E-03	6.2E-04	2.3E-04	9.6E-05	4.6E-05	2.5E-05	1.6E-05	1.1E-05	9.4E-06	2.4E-04
10	2.9E-02	7.5E-03	2.2E-03	7.2E-04	2.7E-04	1.2E-04	5.9E-05	3.3E-05	2.2E-05	1.6E-05	4.0E-04
11	9.2E-01	3.0E-02	8.0E-03	2.4E-03	8.3E-04	3.2E-04	1.5E-04	7.5E-05	4.4E-05	2.9E-05	6.6E-04
12	2.8E-02	9.2E-01	3.1E-02	8.5E-03	2.6E-03	9.4E-04	3.8E-04	1.8E-04	9.4E-05	5.7E-05	1.1E-03
13	6.9E-03	2.9E-02	9.1E-01	3.2E-02	9.0E-03	2.9E-03	1.1E-03	4.5E-04	2.1E-04	1.2E-04	1.8E-03
14	2.0E-03	7.5E-03	3.0E-02	9.1E-01	3.3E-02	9.5E-03	3.1E-03	1.2E-03	5.2E-04	2.6E-04	2.9E-03
15	6.8E-04	2.3E-03	8.1E-03	3.1E-02	9.1E-01	3.3E-02	9.9E-03	3.4E-03	1.3E-03	6.0E-04	4.7E-03
16	2.7E-04	8.1E-04	2.5E-03	8.6E-03	3.1E-02	9.0E-01	3.3E-02	1.0E-02	3.6E-03	1.5E-03	7.7E-03
17	1.3E-04	3.4E-04	9.4E-04	2.8E-03	9.1E-03	3.1E-02	9.0E-01	3.2E-02	1.0E-02	3.8E-03	1.3E-02
18	7.1E-05	1.7E-04	4.1E-04	1.1E-03	3.1E-03	9.6E-03	3.1E-02	8.9E-01	3.2E-02	1.0E-02	2.0E-02
19	4.8E-05	9.8E-05	2.2E-04	5.1E-04	1.3E-03	3.5E-03	1.0E-02	3.2E-02	8.9E-01	3.2E-02	3.3E-02
20	4.8E-05	8.7E-05	1.7E-04	3.6E-04	8.0E-04	1.9E-03	5.0E-03	1.4E-02	4.0E-02	8.8E-01	5.5E-02
D	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.0E+00