

Imperial College London
Department of Mathematics

Correlated Continuous Time Markov Chains and Derivatives Pricing

Antonio Dalessandro

A Thesis presented for the degree of
Master of Philosophy of the
Imperial College London

Correlated Continuous Time Markov Chains and Derivatives Pricing

Antonio Dalessandro

A Thesis presented for the degree of
Master of Philosophy of the
Imperial College London

©Antonio Dalessandro 2009. All rights reserved.

Contents

Declaration	v
Acknowledgements	vii
Abstract	1
1 Introduction	5
1.1 Motivation and Dissertation objectives and contributions	7
1.2 Publications	8
2 Correlated Continuous Time Markov Chains	9
2.1 Setup and Assumptions	10
2.2 Continuous Time Markov Chain approximation of Diffusions	11
2.3 Weak Convergence of the approximate generator	15
2.3.1 Semigroup approach (I)	15
2.3.2 Semigroup approach (II)	16
2.3.3 Spectral representation approach	19
2.3.4 Convergence of the derivative price	22
2.4 Multidimensional CTMC and Tensor Product Analysis	23
2.4.1 Computation of approximated diffusive semigroups	29
2.4.2 Subspace approximations to the matrix exponential	32
2.5 Numerical example: stochastic volatility	35
3 Credit Migration Risk and Point-in-Time Credit Dynamics	43
3.1 Literature reviews	43
3.1.1 Model Uniqueness and its Applications	45
3.1.2 Rating Migration Matrices and Data source Description	47
3.2 Model building blocks and Distance to Default Process	48
3.2.1 DTD as a process with jumps	51
3.3 Point-in-time Model Dynamics	56
3.4 Calibration to Historical Rating Migration Matrices	60
3.4.1 Results - Fitted model parameters	63
Notes	70
3.4.2 Market Data	70
4 Dynamic Pricing of Index and Bespoke CDO	73
4.1 Literature reviews	73
4.1.1 Model Uniqueness and its Applications	75
4.2 Notation and Model Specification	75

4.2.1	Conditional Model Dynamics	79
4.3	Model Identification under the physical and risk neutral measure	80
4.3.1	Price sensitivities	85
4.4	Model performance	87
Notes		92
4.4.1	Credit default swap index tranches	92
4.4.2	Index Basis	93
4.4.3	Loss distribution and Index Tranches pricing	95
4.4.4	Change of measure for the VG process	99
5	Conclusion	103
A	Appendices	105
A.1	Exponential jump diffusion process	105
A.2	Variance Gamma (VG) process	108
A.3	Differential Evolution	111
	Bibliography	118

Declaration

I the undersigned hereby declare that the work presented in this thesis is my own. When material has been sourced from other authors, these have been duly acknowledged. This thesis has not previously been presented for this or any other MPhil examinations.

Antonio Dalessandro

Acknowledgements

I am very grateful to Prof. Mark Davis for the time we spent together discussing ideas and the way it is possible sometimes to make them work and happen. He has been advising me through this period at the Math Department of Imperial, being always supportive, helpful and transmitting me a great joy for research and hard work. A remarkable thank to Dr. Stefano Luzzatto and Prof. Manfred Gilli that gave me financial help at the beginning of my PhD studies, encouragements and stimulating discussions. I thank Prof. Darryl Holm, a true friend with whom I spent long hours discussing ideas, enjoying together the beauty of mathematics. I would like to thank Prof. Claudio Albanese for all the advice and teaching. I thank also, Dr. Alex Mijatović, Dr. Oliver Butterly, Paul Jones, Dr. Adele Peel, Dr. Ilir Roko, Dr. Kaisa Darlstadt, Dr. Achilleas Lazarides, Dr. Andreas Eckner, Prof. Enrico De Giorgi, Prof. Damiano Brigo, Dr. Christofer Oduneye and Matthew Wilhelm for their friendship and support given during these years.

I want to say thank to my beloved parents. I will be forever in debt with you and always love you.

I dedicate this work to my brother, Dr. Luca, that I admire for his impressive knowledge, determination and diligence towards study. Whenever I felt stuck in one spot in my calculations and writing, or I was frustrated that something was not going the way I thought it should, I looked at some of Luca's work. I promise it always inspired me, giving me strength, courage and helped me in trusting my vision and ideas. We share the same gift and, my dear Luca, I owe you much more than words can ever express.

Abstract

In this thesis, properties and results on the continuous-time Markov chain approximation for multivariate diffusions and Lévy processes are presented and applied to the problem of derivative pricing. Numerical methods are formulated in order to approximate a specific stochastic process by a corresponding continuous-time Markov chain. We produce desired convergence results for such approximation schemes. A great emphasis is given to the approximation of correlated multidimensional processes and to the construction of the associated correlated continuous-time Markov chain. In fact, the correlation among continuous-time Markov chains is the theme of this dissertation. We show how to build a multi-dimensional continuous-time Markov chain that closely follow the dynamics of a multivariate diffusion. We apply this result to compute the price of European options where the underlying process dynamics is given by the Heston's stochastic volatility model. Furthermore, we introduce a method to create high-dimensional correlated continuous-time Markov chains that approximate specified stochastic processes. We apply this methodology to model the point-in-time credit rating dynamics and also to price standardized Index collateralized debt obligations and more generally credit derivatives portfolios.

Notation

Probability and Markov processes

$\mathbb{E}$	Expectation
$P(A)$	Probability of event A
$\phi(s)$	Characteristic function computed at s
$p_t(x, y)$	Density of the random variable at time t , starting at x computed at y
p_X	Transition density of the random variable X
$F_t(x, y)$	Cumulative distribution function at time t , starting at x computed at y
$p(X_t = z x)$ or $\mathbb{E}_x[1_{\{z\}}]$	Density of the random variable X computed at t in z . conditional on the random variable X_t being x at time $t = 0$
$\sim$	Distributed as
$\mu(x)$	Drift function
$\sigma(x)$	Volatility function
W_t	Standard Wiener process
Δt	time step in discretization
s, t, u	three time instants, $0 \leq s \leq t \leq u$.
$0 = t_0, t_i, \dots, t_i, \dots$	A sequence of equally spaced time instants, $t_i - t_{i-1} = \Delta t$.
$\delta(x)$	Dirac delta function centered in x
Ω	Sigma algebra sets
$X^{(n)}$	Markov Chain with n states
H	Discrete state space
E, S, S_X	State space of X

Functional Analysis and Operators

$\mathcal{A}, \mathcal{L}$	Infinitesimal generator
$\mathcal{A}^{(n)}, \mathcal{A}_h$ or A, L	Approximated Infinitesimal generator with n states, equivalent to a matrix
$D(\mathcal{A}), D(A)$	Domain of the operator $\mathcal{A}$
$\overline{D}(\mathcal{A})$	Core of the operator $\mathcal{A}$
T_t or $T(t)$	Operator

Function spaces

B, X, Y	Banach space
X'	dual space of X
$C(A)$	continuous functions on a metric space A
$C_b(A)$	bounded continuous functions
$C_0(A)$	continuous functions with compact support
$B(A)$	bounded measurable functions
$M(A)$	real valued maps on A
$C_\infty(\mathbb{R}^n)$	continuous functions on $\mathbb{R}^n$ vanishing at infinity
$C^k(\Omega)$	k times continuously differentiable functions on Ω
$C^\infty(\Omega)$	arbitrarily often differentiable functions on Ω
$C_0^\infty(\Omega)$	testfunctions on Ω
$\mathcal{S}(\mathbb{R}^n)$	Schwartz space of rapidly decreasing testfunctions

Chapter 1

Introduction

This thesis presents properties and results on the continuous-time Markov chain approximation for multivariate diffusions and Lévy processes and their application to derivative pricing. We mainly investigate the problem of approximating the dynamics of correlated processes, that is the dominant theme throughout this dissertation, by means of a discretization scheme based on continuous-time Markov chains.

We start by providing a general characterization of Markov processes from the standpoint of the generators of their corresponding semigroups. This supports the formulation of the generator parametric local approximation. We then prove the weak convergence of the approximated generator and study its convergence rate, similarly to Albanese and Mijatovic [2006b]. While in the existing literature there are several examples of stochastic processes approximation schemes based on Markov chains, for single dimensional diffusive or Lévy processes, see, among others, Mijatović and Pistorious [2009], Albanese and Mijatovic [2006b], or uncorrelated processes, see Albanese et al. [2009], it is not straightforward to extend this specific discretization scheme to multiple-dimensions.

Lo [2009] proposes a discretization scheme for diffusion processes in $\mathbb{R}^d$ based on a continuous-time Markov chain approximation, but he wasn't able to provide in his analysis a precise structure of this large generator matrix. Furthermore in Lo [2009] the correlation is not modeled via a scheme extended to the cross space but by projecting the information of the correlated diffusions on the orthogonal dimensions. This can be obtained via the Cholesky decomposition of the covariance matrix. In this way Lo [2009], by restricting the generator of the Markov chain to be a tridiagonal matrix, is able to use the same discretization scheme as the one in single dimension, see Albanese and Mijatovic [2006b]. Lo [2009] presents a condition for the convergence of diffusion approximation to general multidimensional diffusions, that is the theorem 4.1 pag. 354 of Ethier and Kurtz [1985]. On the contrary, in this thesis, we solve the problem of finding a scheme that approximate the dynamics of correlated diffusions through corresponding parametric multidimensional correlated Markov chains. We also solve the problem of exponentiating large multidimensional generator matrices through the fast exponentiation of matrices based on the Krylov subspace approximation. In this dissertation we also investigate a more general way to induce correlation among stochastic process approximated with parametric continuous-time Markov chains. We do so by defining a local joint distribution between two transition kernels belonging to two distinct Markov chains. We apply our framework to the problem of modeling point-in-time rating migration matrices, and this represents a new approach with respect to the existing market practice Fligewski et al. [2006]. We

investigate also the problem of calibrating our parametric rating migration model to a term structure of market observed rating migration matrices. At this purpose, we define a spectral distance between two matrices that is very important and useful in the context of fitting a parametric generator matrix given an observed Markov chain. Finally we apply the method of correlated continuous-time Markov chain to model the dynamics of a credit derivative portfolio. The use of Markov chain in the context of CDO pricing is not new. Bielecki et al. [2009] is a comprehensive survey of various methods and results regarding application of the theory of continuous time Markov chains (CTMC) to the valuation of credit derivatives. A CDO pricing model based on Markov chains is presented in Digraziano and Rogers [2005].

The thesis is organized as follows. Chapter 1 presents some theoretical background on basic properties of stochastic processes and Markov processes. In this chapter we also discuss basic facts on Lévy processes. In Chapter 2 we present a new method for the approximation of multivariate correlated diffusions by their corresponding continuous-time Markov chain. The material in this chapter is based on Dalessandro [2010a]. We examine the construction of the discrete approximation and its convergence properties and provide explicit calculation for correlated Markov chains using tensor algebra. Examples of interest in statistics and financial derivative pricing are included along with results on the fast exponentiation of matrices.

Chapter 3 presents a credit migration model that aims to consistently capture the point-in-time dynamics of the credit worthiness of debt issuers and their obligations, and a calibration routine that permits the model to effectively fit historical ratings data. The conditional rating dynamics is modeled through two joint correlated continuous-time Markov chain that approximate the a market factor and the credit quality process respectively. The material in this chapter is based on Dalessandro [2010b].

The approach proposed in this Chapter is to view the rating migration matrices as the operator semigroup associated to an approximated parametric infinitesimal generators. Our credit model accounts not only for default risk dynamics but also for the entire transition among states of the rating migration matrix. This modeling feature is fundamental for an efficient risk management of credit derivatives and credit risk portfolios conditionally on a state of the economy or specific macro factors. We fit our model to the historical average rating migration matrices published by Moody's Investors Service, focusing on the banking sector over the period 1920-2005. Our results show that the model can identify the through-the-cycle transition across rating scales and that the point-in-time migration probabilities are only generated by stressed economic conditions and can only be justified by the influence of macro factors on the through-the-cycle unconditional probability values. The great level of modeling details and the accuracy of the produced results is an improvement over those of other available models.

We continue the investigation on correlated continuous-time Markov chain in Chapter 4. The material in this chapter is based on Dalessandro [2010c]. In this Chapter we present a framework to model correlated default events that can be used to price and hedge standard and exotic credit baskets whose values depend on the realized losses of a default portfolio. The model consists of parametric continuous time Markov chain and aims to accurately capture the point-in-time dynamics of the credit worthiness of debt issuers, the single name dynamics, and the dynamics of the correlation among the names underlying the credit portfolio. Key features of the model are the implied credit rating and the local correlation as a function of the implied rating and time. We illustrate through an example

how to price and hedge with this model standard CDO index tranches. This method is computationally fast and parsimonious because uses just one matrix. The framework is simple and at the same time very rich in information and this can easily accommodate for the evaluation of MBSs, CMOs and CDO of ABS tranches.

1.1 Motivation and Dissertation objectives and contributions

The theory and proposed financial modeling approaches which is presented in the following pages conceivably comes from the blend of methods belonging to functional analysis, probability theory, optimization and linear algebra and constitutes a new way to approach the analysis and pricing of financial derivatives, specifically when dealing with derivatives where the underlying stochastic processes are correlated. My contributions can be summarized as it follows:

- General characterization of Markov processes from the standpoint of the generators of their corresponding semigroups. This supports the formulation of the generator local approximation. I prove the weak convergence of the approximated generator and I study its convergence rate.
- I propose a new discretization scheme for correlated diffusions based on continuous-time Markov chains and tensor algebra.
- I solve the problem of exponentiating large multidimensional generator matrices through the fast exponentiation of matrices based on the Krylov subspace approximation.
- I propose a new approach to model correlated continuous-time Markov chains and apply this framework to the problem of modeling point-in-time rating migration matrices.
- I define a spectral distance between two matrices that is very important and useful in the context of fitting a parametric generator matrix given an observed Markov chain. This is the case of fitting a term structure of rating migration matrices with a parametric continuous-time Markov chain that approximate over a discrete set of states the dynamics of a specific stochastic process for the credit quality.
- A new dynamic model for portfolio credit derivative pricing based on correlated continuous-time Markov chains.

Furthermore I develop all the computational routines to make this theory applicable to real problems. All the chapters are self contained and I will point to the relevant literature and references where necessary.

1.2 Publications

Published Papers

1. BRIGO, D., DALESSANDRO, A., NEUGEBAUER, M., AND TRIKI, F. (2009). A stochastic processes toolkit for risk management. Geometric Brownian Motion, jumps, GARCH and Variance Gamma models. *Journal of Risk Management in Financial Institutions*, Vol. 2, 4 365-393.
2. BRIGO, D., DALESSANDRO, A., NEUGEBAUER, M., AND TRIKI, F. (2009). A stochastic processes toolkit for risk management. Mean reverting processes and jumps. *Journal of Risk Management in Financial Institutions*, Vol. 3, 1.

Submitted for Publication

3. DALESSANDRO, A. (2010). Correlated Continuous Time Markov Chains in Multi-dimensional Vector Spaces with Application to Derivatives Pricing, 1-44.
4. DALESSANDRO, A. (2010). Credit Migration Risk and Point-in-Time Credit Dynamics: A New Perspective for Credit Risk Management, 1-48.
5. DALESSANDRO, A. (2010). All your Hedges in one Matrix, 1-32.

Conference Proceedings

- C1. A. DALESSANDRO. CMS 20-22 April 2007, Geneva. Quantitative Finance Conference. Calibration of Rating Migration Matrices using Stochastic Operators.
- C2. A. DALESSANDRO. CMS 17 May 2006, Amsterdam. Quantitative Finance Conference. A Hidden Markov Model for Equity return Levels.
- C3. A. DALESSANDRO. Summer School of Financial derivatives, 23 June 2006, London. Optimization Problem in calibrating Markov Operators to Rating Migration Matrices.

Chapter 2

Correlated Continuous Time Markov Chains

This Chapter presents a new method to construct an approximating continuous time Markov chain that closely follows the dynamics of the given multivariate correlated Markov model. The main motivation for investigating this topic is the apparent absence in the literature of a reference that reports a clear and explicit way how to build correlated continuous time Markov chains that approximate a multivariate diffusion. Therefore the need for an effective representation method that can be exploited in financial modeling and statistical applications.

Our investigation and findings mainly rely on the following ideas. The first idea is that certain multidimensional parabolic boundary problems can be solved by solving a few one dimensional problems - it is the basis of the classical method of separation of variables of mathematical physics. In the case of partial difference equations, this idea leads to tensor product analysis of the matrices involved as reported in Lynch et al. [1963]. The second idea is that infinitesimal generators of diffusions can be approximated as continuous time Markov chains that have a matrix representation, see Albanese and Mijatovic [2006a], and use such generator matrices to define the dynamics in every dimension of a multidimensional stochastic problem. Furthermore, the third idea is that tensor algebra can effectively describe the multivariate boundary value problem when dimensions are orthogonal and also when spaces are coupled with the presence of cross derivative terms - correlated diffusions.

On the basis of our ideas we formulate a methodology and accomplish the following: (i) exact representation of the differential operator for a diffusion as a matrix - continuous time Markov chain (CTMC) and exact associated solution, (ii) explicit exact solutions of problems consisting of separable partial difference equations and boundary conditions are obtained using tensor algebra, (iii) exact representation of correlated variables using tensor algebra, (iv) a stable algorithm is devised with which these exact solutions can be easily and quickly computed.

An important result in this proposed model is that the fast computation of the solution is guaranteed by an effective dimension reduction of the formulated large-scale numerical problem.

The methodology proposed in this Chapter can be seen as an alternative to Alternating Direction Implicit (ADI) schemes for the numerical solution of initial-boundary value problems for convection-diffusion equations with cross derivative terms, see Mitchell and Griffiths

[1980].

The Chapter is organized as follows. Section 1 sets out the model and assumptions. In section 2, we describe how to approximate a diffusion by its corresponding continuous time Markov chain. The important concept of weak convergence of the approximated diffusion is discussed in section 3. Section 4 reports the main result of the Chapter, that is the CTMC approximation of correlated diffusions and the fast exponentiation of large matrices through the Krylov subspace approximation. In section 5 we apply our method to concrete cases and provide examples and numerical results.

2.1 Setup and Assumptions

The aim of this Chapter is to describe a method to construct an approximating continuous time Markov chain that closely follows the dynamic of multi-dimensional correlated diffusion processes. In this section we introduce multivariate diffusive stochastic differential equation (SDE) and give background on diffusions and introduce all the necessary assumptions for the uniqueness and existence of the SDE solution to exist. In order to fix ideas, let us consider a d -dimensional Markov process $X = \{\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\}$, and assume that X has continuous path. Consider the following SDE for a multivariate time-homogeneous diffusion

$$dX_t = \mu(X_t)dt + \Sigma(X_t)dW_t \quad (2.1)$$

where X_t and $\mu(X_t)$ are $d \times 1$ vectors, Σ is a $d \times d$ matrix, which need not be symmetric and W_t is an $d \times 1$ vector of independent Brownian motions. Arbitrary correlation between the shocks to the different equations can be modeled through the inclusions of off-diagonal terms in the Σ matrix. In time-inhomogeneous diffusions, the coefficients are allowed to depend on time directly, as in $\mu(t, X_t)$ and $\Sigma(t, X_t)$, beyond their dependence on time via the state vector. The time-inhomogeneous case can be reduced to the time-homogeneous case by treating time as an additional state variable and so it suffices to return to the model specified in (2.1).

We denote by $\mathcal{S}_X$, a subset of $\mathbb{R}^d$, the domain of the diffusion X , assumed, for simplicity, to be of the following form.

Assumption 2.1 $\mathcal{S}_X$ is a product of d intervals with limits $\underline{x}_i$ and $\bar{x}_i$, where possibly $\underline{x}_i = -\infty$ and/or $\bar{x}_i = \infty$, in which case, the intervals are open at infinite limits.

We consider the variance-covariance matrix of the process

$$\Sigma(\mathbf{x}) \equiv \sigma(x)\sigma^T(x) \quad (2.2)$$

and assume that this matrix satisfy the following regularity condition:

Assumption 2.2 The matrix $\Sigma(\mathbf{x})$ is positive definite for all x in the interior of $\mathcal{S}_X$.

Further assumptions are required to ensure the existence and uniqueness of a solution to (2.1) and to make the computation of the solution possible. We will assume the following.

Assumption 2.3 $\mu(x)$ and $\sigma(x)$ are infinitely differentiable in x on $\mathcal{S}_X$.

Assumption (2.3) ensures the uniqueness of the solution to (2.1). Indeed, Assumption (2.3) implies in particular, that the coefficients of the stochastic differential equation are locally Lipschitz under their assumed (once) differentiability, which can be seen by applying the mean value theorem. This ensures that a solution, if it exists, will be unique (cf. Karatzas and Shreve [1991], Th. 5.2.5). The next assumption restricts the growth behavior of the coefficients near the boundaries of the domain.

Assumption 2.4 *The drift and diffusion functions satisfy linear growth conditions, that is, there exists a constant K such that for all $x \in \mathcal{S}_X$ and i, j ,*

$$|\mu_i(x)| \leq K(1 + \|x\|) \quad \text{and} \quad |\sigma_{ij}(x)| \leq K(1 + \|x\|). \quad (2.3)$$

Their derivatives exhibit at most polynomial growth.

The role of Assumption (3) is to ensure the existence of a solution to the stochastic differential equation (1) by preventing explosions of the process in finite expected time. While it can be relaxed in specific examples, it is not possible to do so in full generality. In dimension one, however, finer results are available (cf. Karatzas and Shreve [1991], Th. 5.5.15 Engelbert-Schmidt criterion), allowing linear growth to be imposed only when the drift coefficient pulls the process toward an infinity boundary (cf. Ait-Sahalia [2002], Proposition 1). In all dimensions, the linear growth condition in Assumption (2.4) is only an issue near the boundaries of $\mathcal{S}_X$. In the special case where $\mathcal{S}_X$ is compact, the growth condition (boundedness, in fact) follows from the continuity of the functions. The additional assumption that the derivatives of the drift and diffusion functions grow at most polynomially simplifies matters in light of the exponential tails of the transition density p_X . Finally, the diffusion process X is fully defined by the specification of the functions μ and σ its behavior at the boundaries of $\mathcal{S}_X$. In many examples, the specification of μ and σ predetermines the boundary behavior of the process, but this will not be the case for models that represent limiting situations. We have, further, that the following limit, defines the infinitesimal generator $\mathcal{A}$ of the Markov process, see pag. 8 Ethier and Kurtz [1985],

$$\lim_{t \downarrow 0} \frac{1}{t} \left[\mathbb{E}_x f(X_t) - f(x) \right] = (\mathcal{A}f)(x) \quad \forall x \in \mathbb{R}^d \quad (2.4)$$

and holds for every f in the space $C^2(\mathbb{R}^d)$ of real-valued twice continuously differentiable functions on $\mathbb{R}^d$; the operator $\mathcal{A}f$ in (2.4) is given by

$$(\mathcal{A}f)(x) := \sum_i^d \mu_i(x) \frac{\partial f(x)}{\partial x_i} + \frac{1}{2} \sum_i^d \sum_k^d \sigma_{ik}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_k} \quad (2.5)$$

for suitable Borel-measurable functions $\mu_i, \sigma_{ik} : \mathbb{R}^d \rightarrow \mathbb{R}$, $i, k = 1, \dots, d$, satisfying all the above assumptions.

2.2 Continuous Time Markov Chain approximation of Diffusions

In this section we describe how to construct an approximating continuous time Markov chain that closely follows the dynamic of a single dimension given Markov model. We

recollect in this section some important notions and results from Markov processes and semigroup theory already given in the previous Chapter and review some key concepts of the standard Markovian setup that we will use through the Chapter. Our main references are Ethier and Kurtz [1985], Rogers and Williams [2000] and Engel and Nagel [1999]. Let the process $X = \{X_t\}_{t \geq 0}$ be a one dimensional diffusion with domain $\mathbf{E} \subset \mathcal{S}_X$ and satisfy the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad \text{for } t \geq 0, \quad (2.6)$$

where W_t is a 1-dimensional Brownian motion and $\mu(x) : \mathbf{E} \rightarrow \mathbb{R}$, $\sigma(x) : \mathbf{E} \rightarrow \mathbb{R}_+$ satisfy the assumptions made in the previous section

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C|x - y|, \quad x, y \in \mathbf{E}$$

i.e. that $\mu(\cdot)$ and $\sigma(\cdot)$ are Lipschitz continuous and of sublinear growth, with C a positive constant.

X a Markov process defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. X takes values in $\mathbf{E}$ and satisfy the Markov property:

$$\mathbb{E}[f(X_{t+s})|\mathcal{F}_t] = P_s f(X_t)$$

for all $s, t \geq 0$ and $f : \mathbf{E} \rightarrow B$ is a measurable and nonnegative function, B a Banach space (For definition of the Banach space $B := l^p(\mathbf{E})$, see Kreyszig [1978]) and where $\mathbb{E}$ denotes the expectation under the probability measure $\mathbb{P}$ and $P_s f$ is given by

$$P_s f(x) := \mathbb{E}_x[f(X_s)] := \mathbb{E}[f(X_s)|X_0 = x].$$

The family $\{P_t\}_{t \geq 0}$ forms a strongly continuous semigroup on B that is continuous in the strong operator topology. A strongly continuous semigroup on B is a map

$$P_t : \mathbf{E} \rightarrow B$$

such that the following axioms hold

$$\begin{aligned} P_0 &= I \\ P_{t+s} &= P_t P_s, \quad \text{for all } s, t \geq 0 \\ \lim_{t \downarrow 0} \|P_t f - f\| &= 0 \quad \text{for all } f \in B \end{aligned}$$

The first two axioms are algebraic, and state that $\{P_t\}_{t \geq 0}$ is a representation of the semigroup $(\mathbf{E}, B)$; the last is topological, and states that the map P_t is continuous in the strong operator topology.

Remark 2.1 *It is immediate to extend the notions and formalism just introduced to the price of derivative securities. In fact denoting by r a constant rate of discounting, and by d the dividend yield, for any non-negative Borel function f the expected discounted cash flow at time t with maturity T , that determines by standard no arbitrage argument the price of the derivative contract is given by an expression very similar to the ones above:*

$$\mathbb{E}_{X_t}[e^{-(r-d)(T-t)} f(X_T)] \quad (2.7)$$

We make the following assumption:

Assumption 2.5 *X is a Feller process on $\mathbf{E}$, that is, for any $f \in C_0(\mathbf{E})$, the family $\{P_t f\}_{t \geq 0}$ satisfies:*

$$\begin{aligned} P_t f &\in C_0(\mathbf{E}) \quad \text{for any } t > 0 \\ \lim_{t \downarrow 0} P_t f(x) &= f(x) \quad \text{for any } x \in \mathbf{E} \end{aligned}$$

The semigroup $P_t f(x)$ is a contraction semigroup as it is straightforward to see that

$$|P_t f(x)| = |\mathbb{E}_x[f(X_t)]| \leq \sup_{y \in \mathbf{E}} |f(y)| \leq \|f(x)\|$$

The Feller property is a standard condition, which guarantees that a version of the process X with cadlag paths exists, satisfying the strong Markov property.

We can furthermore state, as in Ethier and Kurtz [1985], that the infinitesimal generator $\mathcal{A}$ of a strongly continuous contraction semigroup $\{P_t f\}_{t \geq 0}$ is defined by the map

$$\mathcal{A}f := \lim_{t \rightarrow 0} \frac{(P_t f - f)(x)}{t} \quad (2.8)$$

for all f in $\mathcal{D}$, with $\mathcal{D}$ the domain of $\mathcal{A}$.

The generator $\mathcal{A} : l^p(\mathbf{E}) \rightarrow l^p(\mathbf{E})$ and in particular $\mathcal{D} \in C_0(\mathbf{E})$ is the set of all $f \in C_0(\mathbf{E})$ for which the right end side of (2.8) converges in the strong sense, therefore with respect to the norm $\|f\| = \sup_{x \in \mathbf{E}} |f(x)|$ of the Banach space $B, (C_0(\mathbf{E}), \|\cdot\|)$. $\mathcal{D}$ is dense in $C_0(\mathbf{E})$ and for any $f \in \mathcal{D}$. $\mathcal{A}$ is linear in its domain, and it is closed, although not necessarily bounded.

Definition 2.1 *We define $\overline{\mathcal{D}}$ as the core of $\mathcal{A}$, such that $\overline{\mathcal{D}} \in C_0(\mathbf{E})$ is the dense set $\{(\lambda - \mathcal{A})f : f \in C_0(\mathbf{E})\}$ in $C_0(\mathbf{E})$ for some $\lambda > 0$.*

The relation between a given linear operator $\mathcal{A}$ and the corresponding semigroup can be expressed with the abstract Cauchy problem

$$\begin{cases} \frac{du(t)}{dt} = \mathcal{A}u(t), & t \in [0, T] \\ u(0) = v, & \text{initial probability distribution} \end{cases} \quad (2.9)$$

The infinitesimal generator in eq.(2.8) associated to eq.(2.6) is

$$\mathcal{A}f(x) = \mu(x) \frac{\partial f(x)}{\partial x} + \frac{1}{2} \sigma^2(x) \frac{\partial^2 f(x)}{\partial x^2} \quad (2.10)$$

that acts on $C_0^2(\mathbf{E})$ as a differential operator for all $f \in \mathcal{D}$, being $\mathcal{D}$ the operator domain as previously defined.

In order to approximate eq. (2.10), we construct a state space $\mathbf{H} \subset \mathbf{E}$ with $n \in \mathbb{N}$ elements and define the sets

$$\mathbf{H} := \{\underline{x} := -\frac{2^{2n}}{2^n} = x_1, -\frac{2^{2n}}{2^n} + \frac{1}{2^n} := x_2, \dots, \frac{1}{2^n}, \frac{2}{2^n}, \dots, x_m = \frac{2^{2n}}{2^n} := \bar{x}\} \quad \text{and} \quad \mathbf{H}^o := \mathbf{H} \setminus \partial \mathbf{H}$$

with $h = \frac{1}{2^n}$ a positive constant that represents the discretization unit, and where the boundary $\partial\mathbf{H}$ consist of the smallest (i.e. $\underline{x}$) and largest (i.e. $\bar{x}$) elements in $\mathbf{H}$, possibly a countably infinite set, and the interior $\mathbf{H}^\circ$ is the complement of the boundary. It is possible to construct the continuous time Markov chain $X^{(n)} := \{X_t^n\}_{t \geq 0}$ as the discrete approximation of X on $\mathbf{H}$ by building a matrix $\mathcal{A}^{(n)} = \{a(x_i, x_j)\}$, $i, j = 1, \dots, 2^{2n+1} + 1 = m$, that is the discretized counterpart of $\mathcal{A}$ in eq.(2.10) and each entry can be calculated by solving the following system:

$$\begin{cases} a(x_1, x_2) = a(x_m, x_{m-1}) = 0 \\ a(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{\mu(x_i)}{h} + \frac{\sigma^2(x_i)}{h^2} \right) \\ a(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma^2(x_i)}{h^2} - \frac{\mu(x_i)}{h} \right) \\ a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})) \end{cases}$$

for all $i = 2, \dots, 2^{2n+1}$, with $\frac{\sigma^2(x_i)}{h} \leq \mu(x_i) \leq \frac{\sigma(x_i)^2}{h}$. However the discrete state space $\mathbf{H}$ does not need to be uniform and alternative discretization routines are presented in Tavella and Randall [2000].

Remark 2.2 *Note that for $x \in \partial\mathbf{H}$, in line with the assumption and arguments of Section 3.1.2, and for computational aspects, we impose an absorbing boundary condition. However it is important to choose the boundary states far enough that the laws of the processes $X^{(n)}$ and X are close to each other during the finite time interval of interest. However this this will not be our setting for the study of weak convergence.*

The resulting matrix $\mathcal{A}^{(n)}$ is a tridiagonal matrix in $\mathbb{R}^{m \times m}$, $m = 2^{2n+1} + 1$, with always positive extra-diagonal elements. The previous system calculates the entries of the generator by specifying the first and second instantaneous moments of the process $X^{(n)}$ that have to coincide with those of X on the set $\mathbf{H}^\circ$. This is equivalent to satisfy the following conditions

$$\mathbb{E}_{X_t}[(X_{t+\Delta t} - X_t)^z] = \mathbb{E}_{X_t^n}[(X_{t+\Delta t}^n - X_t^n)^z] + o(\Delta t), \quad z \in \{1, 2\} \text{ and } X^{(n)} \in \mathbf{H}^\circ \quad (2.11)$$

Furthermore one could in principle produce more accurate results by matching higher instantaneous moments of the two processes.

The numerical problem we face is of the same type as in eq.(2.9) and its analytic solution is $p(t) = e^{tA}v$ and represents the transient probability of the Markov chain with n states. More generally if $\mathcal{L}$ is a bounded operator on the Banach space B , then

$$P_t = e^{\mathcal{L}t} := \sum_{k=0}^{\infty} \frac{\mathcal{L}^k}{k!} t^k \quad (2.12)$$

is a strongly continuous semigroup. Conversely any uniform continuous semigroup is necessarily of this form for some bounded linear operator $\mathcal{L}$. In particular if B is a finite dimensional Banach space, then any strongly continuous semigroup is necessarily of this form for some linear operator $\mathcal{L}$.

Therefore the core of our numerical implementation is the computation of a matrix exponential. This is not a trivial task, especially when n is a very large number. We deal in detail about this problem in section 4.

2.3 Weak Convergence of the approximate generator

In this section we review the existence of a weak limit of continuous Markov chains as $h \downarrow 0$, result that was established by Sofa [1967] and Kurtz by using operator semigroup methods, and is clearly reported in Ethier and Kurtz [1985]. Convergence of operators is also been extensively studied in functional analysis Kreyszig [1978], while in probability Stroock and Varadhan proposed a result by reducing convergence to the martingale problem. Further studies are present in the numerical analysis literature where it has been used the spectral representation of the Markov generator to derive desired convergence results Suzuki [1978]. In what follows, we analyze the convergence from the semigroup viewpoint and also using the spectral representation of the operator.

Three approaches to investigate and prove the weak convergence of the discretized process to a diffusion are proposed, and they all share the main principle used to prove convergence, namely that the analysis done on the operator spectrum properties can be mapped and transferred into related semigroup properties.

Assumption 2.6 *Without loss of generality we assume, for the analysis of weak convergence, $\mathbf{H}$ to be the countably infinite set $h\mathbb{Z}$. Therefore the Markov generator of the discrete process is given by the operator $\mathcal{A}^{(n)} : l^p(h\mathbb{Z}) \rightarrow l^p(h\mathbb{Z})$.*

In what follows, we adopt the following conventions and assumptions, see pag. 28 Ethier and Kurtz [1985]. $B_n = l^p(\mathbf{H})$, $n = 1, 2, \dots$, in addition to B , is a Banach space (with norm also denoted $\|\cdot\|$) and $\pi_n : B \rightarrow B_n$ is a bounded linear transformation. We assume that $\sup_n \|\pi_n\| < \infty$. We write $f_n \rightarrow f$ if $f_n \in B_n$ for each $n \geq 1$, $f \in B$, and $\lim_{n \rightarrow \infty} \|f_n - \pi_n f\| = 0$. $P_t^{(n)}$, $n = 1, 2, \dots$, and P_t are strongly continuous contraction semigroups (see def.(??)) on B_n and B with generators $\mathcal{A}^{(n)}$ and $\mathcal{A}$ respectively, and let $\overline{\mathcal{D}}$ be a core for $\mathcal{A}$.

2.3.1 Semigroup approach (I)

We want that the continuous Markov chain $X^{(n)} := \{X_t^n\}_{t \geq 0}$ has a dynamics as close as possible to the corresponding approximated process X . At this purpose we can define an error

$$\epsilon_n(f) := \sup_{x \in \mathbf{H}} \|\mathcal{A}^{(n)} f(x) - \mathcal{A} f(x)\| \quad (2.13)$$

Using the semigroup approach to weak convergence we can state that, if $\epsilon_n(f)$ tends to zero as n tends to infinity for f in $\overline{\mathcal{D}}$, then the sequences of processes $X^{(n)}$, converges weakly to X , in the space $D_{\mathbf{E}}$. The space $D_{\mathbf{E}}$, is the space of right continuous functions $f : \mathbf{E} \rightarrow \mathbb{R}$ with left limits, see Ch. 5, pp. 115 of Ethier and Kurtz [1985] for more details. Formally we can state the following result,

Theorem 2.1 *Let X be a Feller process with state space $\mathbf{E}$ and infinitesimal generator $\mathcal{A}$ satisfying above properties and assumptions, and $X^{(n)}$ be a sequence of Markov chains with generator matrices $\mathcal{A}^{(n)}$. Then it holds that*

$$\lim_{n \rightarrow \infty} \epsilon_n(f) = 0 \quad (2.14)$$

for every function f in the core of $\mathcal{A}$, is equivalent to

$$P_t^{(n)} \pi_n f \rightarrow P_t f, \quad \text{for all } f \in C_0(\mathbf{E}), \quad t \geq 0$$

Then the sequences of processes $X^{(n)}$, converges weakly to X , in the space $D_{\mathbf{E}}$.

Proof (Th. (2.1)) From Theorem 6.1 in Chapter 1 pag. 28 of Ethier and Kurtz [1985], we have that if $P_t^{(n)}$, $n = 1, 2, \dots$, and P_t are strongly continuous contraction semigroups on B_n and B with generators $\mathcal{A}^{(n)}$ and $\mathcal{A}$ respectively, and let $\overline{\mathcal{D}}$ be a core for $\mathcal{A}$, then the following are equivalent:

(i) For each $f \in B$

$$P_t^{(n)} \pi_n f \rightarrow P_t f$$

uniformly on bounded intervals.

(ii) For each $f \in B$

$$P_t^{(n)} \pi_n f \rightarrow P_t f$$

for all $t \geq 0$.

(iii) For each $f \in \overline{\mathcal{D}}$, there exists $f_n \in D(\mathcal{A}^{(n)})$ ($D(\mathcal{A}^{(n)})$ is the domain of $\mathcal{A}^{(n)}$) for each $n \geq 1$ such that $f_n \rightarrow f$ and $\mathcal{A}^{(n)} f_n \rightarrow \mathcal{A} f$.

Furthermore, following Ethier and Kurtz [1985], Chapter 4 pag. 172 Theorem 2.11, we have that if P_t is a Feller Semigroup on $C_0(\mathbf{E})$ and that for each $t \geq 0$ and $f \in C_0(\mathbf{E})$

$$P_t^{(n)} \pi_n f \rightarrow P_t f.$$

If $X_0^{(n)}$ has limiting distribution ν , then there is a Markov Process X corresponding to P_t with initial distribution ν and sample paths in $D_{\mathbf{E}}$, and

$$X^{(n)} \rightarrow X \quad \text{in } D_{\mathbf{E}}.$$

where $D_{\mathbf{E}}$ is the Skorokhod space of càdlàg real-valued function endowed with the Skorokhod topology.

2.3.2 Semigroup approach (II)

This alternative proof of the weak convergence based on the semigroup approach is local and based on the same results of Ethier and Kurtz [1985], presented in the previous section. We remind the reader that the settings and the assumptions on the process X are as specified above, and its behavior on the boundary can, in certain cases, also be specified. In this proof we also consider the case where $\partial \mathbf{H}$ are finite and absorbing states of the chain.

The semigroups of the Markov chain $X^{(n)}$ that takes values on $\mathbf{H}$ are given by $P_t^{(n)}$ with

$$P_t^{(n)} \pi_n f = \mathbb{E}_x[f(X_t^{(n)})] \tag{2.15}$$

for all $f \in C_b(\mathbf{H})$. Here $C_b(\mathbf{H})$ denotes the set of bounded functions from $\mathbf{H}$ to $\mathbb{R}$. We have that

$$\begin{aligned} P_0^{(n)} f_n(x) &= \mathbb{E}_x[f(X_0^{(n)})] = f(x) \\ P_{s+t}^{(n)} &= P_s^{(n)} P_t^{(n)} \end{aligned}$$

The last equation follows directly from the Chapman-Kolmogorov equation.

$P_t^{(n)}$ is a contraction semigroup because for all f

$$|P_t^{(n)} f(x)| = |\mathbb{E}_x[f(X_t^{(n)})]| \leq \sup_{y \in \mathbf{H}} |f(y)| = \|f\|. \quad (2.16)$$

Let $C_n = \max_{x \in \mathbf{H}} |f(x_{i+1}) - f(x_i)|$, and $\mathcal{A}^{(n)} = \{a(x_i, x_j)\} \quad i, j = 1, \dots$. Then

$$\begin{aligned} |P_t^{(n)} f_n(x) - f_n(x)| &= |\mathbb{E}_x[f_n(X_t^n) - f(x)]| \\ &\leq |(a(x_i, x_{i+1})t + o(t))[f(x_{i+1}) - f(x_i)]| \\ &\quad + |(a(x_i, x_{i-1})t + o(t))[f(x_{i-1}) - f(x_i)]| + o(t) \\ &\leq \left(\max_{x \in \mathbf{H}} a(x_i, x_{i+1})t + o(t) \right) C_n + \left(\max_{x \in \mathbf{H}} a(x_i, x_{i-1})t + o(t) \right) C_n \\ &= C_n t \left(\max_{x \in \mathbf{H}} a(x_i, x_{i+1}) + \max_{x \in \mathbf{H}} a(x_i, x_{i-1}) + \frac{o(t)}{t} \right). \end{aligned}$$

We can then obtain the inequality

$$\sup_{x \in \mathbf{H}} |P_t^{(n)} f_n(x) - f_n(x)| \leq C_n t \left(\max_{x \in \mathbf{H}} a(x_i, x_{i+1}) + \max_{x \in \mathbf{H}} a(x_i, x_{i-1}) + \frac{o(t)}{t} \right) \xrightarrow{t \rightarrow 0} 0,$$

that shows that the semigroup $P_t^{(n)}$ is strongly continuous. If the semigroups $P_t^{(n)}$ and P_t are strongly continuous holds the result of the Theorem 6.1 in Chapter 1 pag. 28 of Ethier and Kurtz [1985]. Then the desired convergence result is given by the following theorem 2.11, Chapter 4 of Ethier and Kurtz [1985].

Theorem 2.2 *If P_t is a Feller Semigroup on $C_0(\mathbf{E})$ and that for each $t \geq 0$ and $f \in C_0(\mathbf{E})$*

$$P_t^{(n)} \pi_n f \rightarrow P_t f.$$

If $X_0^{(n)}$ has limiting distribution ν , then there is a Markov Process X corresponding to P_t with initial distribution ν and sample paths in $D_{\mathbf{E}}$, and

$$X^{(n)} \rightarrow X \quad \text{in } D_{\mathbf{E}}.$$

Furthermore we show in the following theorem that the generator $\mathcal{A}^{(n)}$ converges to the the generator $\mathcal{A}$.

Theorem 2.3 *For all $f \in \overline{\mathcal{D}}$, with $\overline{\mathcal{D}}$ core of the generator $\mathcal{A}$ as previously defined,*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbf{H}} |\mathcal{A}^{(n)} f(x) - \mathcal{A} f(x)| = 0$$

This mode of convergence is called strong convergence.

Proof (Th. (2.3)) Let $f \in \overline{\mathcal{D}}$ and $x \in \mathbf{H}$, and $h := \Delta x \in \mathbb{R}$ the discretization unit. We can locally write the generator of the semigroup $P_t^{(n)}$ as

$$\mathcal{A}^{(n)} f(x) = a^{(n)}(x, x+h) (f(x+h) - f(x)) + a^{(n)}(x, x-h) (f(x-h) - f(x))$$

By Taylor approximation

$$f_n(x) \approx f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)(\Delta x)^2 + o((\Delta x)^2)$$

we obtain for all $i = 1, 2, \dots, x_i \in \mathbf{H}$,

$$\begin{aligned} \mathcal{A}^{(n)}f_n(x_i) &= a(x_i, x_{i+1})\left(f(x_i) + f'(x_i)\Delta x + \frac{1}{2}f''(x_i)(\Delta x)^2 + o((\Delta x)^2) - f(x_i)\right) \\ &+ a(x_i, x_{i-1})\left(f(x_i) - f'(x_i)\Delta x + \frac{1}{2}f''(x_i)(\Delta x)^2 + o((\Delta x)^2) - f(x_i)\right) \\ &= f'(x_i)\Delta x\left(a(x_i, x_{i+1}) - a(x_i, x_{i-1})\right) + \frac{1}{2}f''(x_i)(\Delta x)^2\left(a(x_i, x_{i+1}) - a(x_i, x_{i-1})\right) \\ &+ \left(a(x_i, x_{i+1}) - a(x_i, x_{i-1})\right)o((\Delta x)^2) \end{aligned}$$

Due to the fact that $f \in \overline{\mathcal{D}}$, the error term $o((\Delta x)^2)$ is uniform in x . We have that

$$\begin{aligned} a(x_i, x_{i+1}) + a(x_i, x_{i-1}) &= \frac{\sigma^2(x_i)}{h^2} \\ a(x_i, x_{i+1}) - a(x_i, x_{i-1}) &= \frac{\mu(x_i)}{h} \end{aligned}$$

Please remind that in section 3.1.2 we made precise assumptions on the behavior at the boundary of the process. Without loss of generality we can assume that the boundary is at a point at infinite, not attainable in finite time and it is furthermore absorbing. It is possible to express $\mathcal{A}^{(n)}f_n(x)$ for all $x \in \mathbf{H}$, as

$$\begin{aligned} \mathcal{A}^{(n)}f_n(x) &= f'(x)\Delta x\left(\frac{\mu_i}{\Delta x}\right) + \frac{1}{2}f''(x)(\Delta x)^2\left(\frac{\sigma^2(x_i)}{(\Delta x)^2}\right) + \left(\frac{\mu(x_i)}{\Delta x}\right)o((\Delta x)^2) \\ &= \mathcal{A}f(x) + \left(\frac{\mu(x_i)}{\Delta x}\right)o((\Delta x)^2) \end{aligned}$$

We obtain

$$\sup_{x \in \mathbf{H}} |\mathcal{A}_n f(x) - \mathcal{A}f(x)| = \sup_{x \in \mathbf{H}} \left| \frac{\mu(x_i)}{\Delta x} o((\Delta x)^2) \right| = C_1 o((\Delta x)^2) \xrightarrow{n \rightarrow \infty} 0. \quad (2.17)$$

In case the elements of $\partial\mathbf{H}$, consisting of the smallest (i.e. $\underline{x}$) and largest (i.e. $\overline{x}$) elements in $\mathbf{H}$, are absorbing states of the Markov chain we can continue the above analysis with a further investigation of the weak convergence. Furthermore the behavior on the boundary of the functions $f \in \overline{\mathcal{D}}$ is $f'(\underline{x}) = f'(\overline{x}) = 0$. For $x_i = \underline{x}$ we have,

$$\begin{aligned} |\mathcal{A}^{(n)}f(\underline{x}) - \mathcal{A}f(\underline{x})| &\leq |\mathcal{A}^{(n)}f(\underline{x}) - \mathcal{A}^{(n)}f(\underline{x} + h)| \\ &+ |\mathcal{A}^{(n)}f(\underline{x} + h) - \mathcal{A}f(\underline{x} + h)| \\ &+ |\mathcal{A}f(\underline{x} + h) - \mathcal{A}f(\underline{x})| \end{aligned}$$

The second term tends to 0 as shown above, the third term by continuity of $\mathcal{A}f$ in $\underline{x}$. For the first term we have

$$\begin{aligned} |\mathcal{A}^{(n)}f(\underline{x}) - \mathcal{A}^{(n)}f(\underline{x} + h)| &= |a(\underline{x}, \underline{x} + h)(f(\underline{x} + h) - f(\underline{x})) \\ &- a(\underline{x} + h, \underline{x})(f(\underline{x} + 2h) - f(\underline{x} + h)) \\ &- a(\underline{x} + h, \underline{x} + 2h)(f(\underline{x} + h) - f(\underline{x}))| \\ &= O(n)o(h) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

since $a(\underline{x}, \underline{x} + h)$, $a(\underline{x} + h, \underline{x})$, $a(\underline{x} + h, \underline{x} + 2h)$ are on order n and $(f(\underline{x} + h) - f(\underline{x}))$, $(f(\underline{x} + 2h) - f(\underline{x} + h))$, $(f(\underline{x} + h) - f(\underline{x}))$ are of order $o(h)$ because $f'(\underline{x}) = 0$ for $f \in \overline{D}$. The result for $\bar{x}$ follows in the same way. ■

2.3.3 Spectral representation approach

In what follows we derive some weak converge properties of the continuous Markov chain $X^{(n)}$ by comparing the spectral representations of the probability density functions for the continuous infinitesimal generator and its approximated counterpart. We are presenting this approach because it can be easily generalized to a broader class of Feller processes and process with jumps.

In order to show some desirable convergence properties of the transition kernel we consider a discrete operator associate to a discrete version of Brownian motion¹ with drift taking values in $\mathbf{H} \subseteq h\mathbb{Z} \subset \mathbf{E} \subseteq \mathbb{R}$, where h is a small positive real number

$$\mathcal{A}_h f(x) = \mu \nabla_h(f)(x) + \frac{\sigma^2}{2} \Delta_h(f)(x) \quad (2.18)$$

with

$$\nabla_h(f)(x) = \frac{f(x+h) - f(x)}{h}$$

being the discrete gradient operator and

$$\Delta_h(f)(x) = \frac{f(x+h) - 2f(x) - f(x-h)}{h^2}$$

the discrete Laplace operator. Both operators are defined on the Hilbert space of sequences indexed by $h\mathbb{Z}$ and are applied to every sequence $f \in l^2(h\mathbb{Z})$ for all $x \in h\mathbb{Z}$. The parameters $\mu, \sigma \in \mathbb{R}^{++}$ and $\frac{\sigma^2}{h^2} > \frac{|\mu|}{h}$. We can obtain a spectral representation of the operator $\mathcal{A}_h$ by applying the unitary transformation introduce in the previous Chapter, eq.??, leading to the following diagonal operator,

$$q_h(s) = \mathcal{F}(\mathcal{A}_h(f))\mathcal{F}^{-1}(s, s) \quad (2.19)$$

The calculation $q_h(s)$ of is straightforward and it is just an application of the shift theorem.

$$\mathcal{F}_n(\mathcal{A}_h)(f)(s) = \sum_{x \in \mathbf{H}} F_{s,k} \left(\mu \nabla_h(f)(x) + \frac{\sigma^2}{2} \Delta_h(f)(x) \right)$$

$$\begin{aligned} \mathcal{F}_n(\mu \nabla_h)(f)(s) &= \sum_{x \in \mathbf{H}} F_{s,k} \mu \nabla_h(f)(x) \\ &= \mu \sum_{k=0}^{n-1} F_{s,k} \frac{f(kh+h) - f(kh-h)}{2h} \\ &= \frac{\mu}{2h} \left(e^{ihs} \sum_{k=0}^{n-1} F_{s,k} f(kh) - e^{-ihs} \sum_{k=0}^{n-1} F_{s,k} f(kh) \right) \\ &= \mu \frac{e^{ihs} - e^{-ihs}}{2h} \mathcal{F}(f)(s) = -i\mu \frac{\sin hs}{h} \mathcal{F}(f)(s) \end{aligned}$$

¹A Markov chain approximates the Brownian motion on the state space $\mathbf{H} \subseteq h\mathbb{Z} \subset \mathbb{R}$

Doing a similar calculation for $\mathcal{F}(\frac{\sigma^2}{2}\triangle_h)(f)(s)$ we obtain

$$q_h(s) = \left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs) - 1}{h^2} \right)(s) \quad (2.20)$$

We have

$$\begin{aligned} \frac{1}{h}P_{n,h}^t(x, y) &= \frac{1}{h}p_n(y \leq X_t \leq y+h | X_0 = x) \\ &= \frac{1}{hn} \sum_{s \in B_n} e^{q_n(s)t} e^{is(y-x)} \end{aligned}$$

Therefore at this stage we would like to derive some converge properties of we do this by comparing the spectral representations of the probability density functions for the continuous infinitesimal generator and its approximated counterpart, namely

$$\frac{1}{h}P_{n,h}^t(x, y) = \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{q_n(s)t} e^{is(y-x)} ds$$

and

$$p_t(x, y) = \frac{1}{2\pi} \int e^{(i\mu s - \frac{\sigma^2}{2}s^2)t} e^{is(y-x)} ds$$

In this way we are able to assess the order of convergence of the error

$$\epsilon_h := \left| p_t(x, y) - \frac{1}{h}P_{n,h}^t(x, y) \right| \quad (2.21)$$

In order to prove convergence and assess the rate of convergence in eq.(2.21), we exploit the relationship between the distribution function and its corresponding characteristic function and in particular we refer to the Continuity Theorem, cf. Lukacs [1958], Th. 3.6.1. which states

Theorem 2.4 *Let $\{F_n(x)\}$ be a sequence of distribution functions and denote by $\{f_n(s)\}$ the sequence of the corresponding characteristic functions. The sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if, and only if, the sequence $\{f_n(s)\}$ converges for every s to a function $f(s)$ which is continuous at $s = 0$. The limiting function is then the characteristic function of $F(x)$.*

Theorem 2.5 (cf. Lukacs [1958], Th. 3.2.3.) *Let $f(s)$ be an arbitrary characteristic function. For every real x the limit*

$$p(x) = \lim_{T \downarrow 0} \frac{1}{2T} \int_T^T e^{-isx} f(s) ds \quad (2.22)$$

exists and is equal to the saltus of the distribution function of $f(s)$ at the point x .

In our case we compare the discrete transition probability

$$\frac{1}{h}P_{n,h}^t(x, y) = \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{q_n(s)t} e^{is(y-x)} ds \quad (2.23)$$

where q_n is the discrete characteristic exponent of the process, and

$$p_t(x, y) := F'_t(x, y) = \frac{1}{2\pi} \int e^{\phi(s)t} e^{is(y-x)} ds = \frac{1}{2\pi} \int e^{(i\mu s - \frac{\sigma^2}{2}s^2)t} e^{is(y-x)} ds$$

where $F_t(x, y)$ is the distribution function.

We can therefore focus on the analysis of the passage to the limit $h \rightarrow 0$ of the following spectral representation, conditional on a time t ,

$$\lim_{h \downarrow 0} \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{\left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs)-1}{h^2}\right)(s)t} e^{is(y-x)} ds = \frac{1}{2\pi} \int e^{(i\mu s - \frac{\sigma^2}{2}s^2)t} e^{is(y-x)} ds \quad (2.24)$$

and the result of weak convergence by mean of the weak representation of the operator can be formulated as a corollary of Th.(2.4)

Corollary 2.5.1 *For all x we consider the sequence of distribution functions $\{P_{n,h}^t(x, y)\}$ and by $\{e^{q_n(s)t}\}$ the sequence of the corresponding characteristic functions. The sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if, and only if, the sequence $\{e^{q_n(s)t}\}$ converges for every s to a function $\phi(s)$ which is continuous at $s = 0$. The limiting function is then the characteristic function of $F(x)$.*

Proof Without loss of generality, this result is a particular case of the continuity theorem 2.4, and the convergence error of eq.(2.21) is measured in correspondence of saltus point of the distribution, see Th.(2.5). Lets consider the integral on the left end side of eq.(2.24) and it can be rewritten as

$$\begin{aligned} \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{\left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs)-1}{h^2}\right)(s)t} e^{is(y-x)} ds &= \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{-K} e^{\left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs)-1}{h^2}\right)(s)t} e^{is(y-x)} ds \\ &+ \frac{1}{2\pi} \int_{-K}^K e^{\left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs)-1}{h^2}\right)(s)t} e^{is(y-x)} ds \\ &+ \frac{1}{2\pi} \int_K^{\frac{\pi}{h}} e^{\left(-i\mu_h \frac{\sin hs}{h} + \sigma_h^2 \frac{\cos(hs)-1}{h^2}\right)(s)t} e^{is(y-x)} ds \end{aligned}$$

and it is possible to make the first and the third integral on the right end side arbitrary small by choosing a large number $K > 0$ and by selecting $h > 0$ sufficiently small. If we consider the second integral we can analyze the behaviour as $h \rightarrow 0$. We notice that the function

$$\begin{aligned} \lim_{h \rightarrow 0} \left(i\mu \frac{\sin hs}{h} \right) &= \lim_{h \rightarrow 0} i\mu \frac{s \sin hs}{s h} \\ &= \lim_{h \rightarrow 0} is\mu \frac{\sin hs}{hs} = is\mu \end{aligned}$$

and

$$\begin{aligned} \lim_{h \rightarrow 0} \left(\sigma^2 \frac{\cos hs - 1}{h^2} \right) &= \lim_{h \rightarrow 0} \sigma^2 \frac{s^2 \cos hs - 1}{s^2 h^2} \\ &= \lim_{h \rightarrow 0} \sigma^2 s^2 \frac{\cos hs - 1}{(hs)^2} = -\frac{1}{2} \sigma^2 s^2 \end{aligned}$$

We would like to examine in more details the order of convergence of the above functions as $h \rightarrow 0$. For the limit

$$\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1,$$

using $\sin h = h - \frac{h^3}{6} + \dots$, we get

$$\frac{\sin h}{h} - 1 = \frac{\sin h - h}{h} = -\frac{h^3}{6h} + \dots = -\frac{h^2}{6} + \dots$$

we find an order of $O(h^2)$. In the same way using $\cos h = 1 - \frac{h^2}{2} + \dots$, we can assess the order of convergence of the limit

$$\lim_{h \rightarrow 0} \frac{1 - \cos h}{h^2} = \frac{1}{2},$$

$$\frac{1 - \cos h}{h^2} = \frac{1 - \left(1 - \frac{h^2}{2} + \dots\right)}{h^2} = \frac{1}{2} + \dots$$

and convergence order of $O(1)$. Therefore the order of convergence is at most $O(h^2)$. ■

Along the lines of this proof there is a result, cf. Albanese and Mijatovic [2006a], for the problem of obtaining the transition probability kernel constructively as a limit of increasingly fine triangularizations schemes and assessing the rate of convergence with pointwise bounds of the kernel itself.

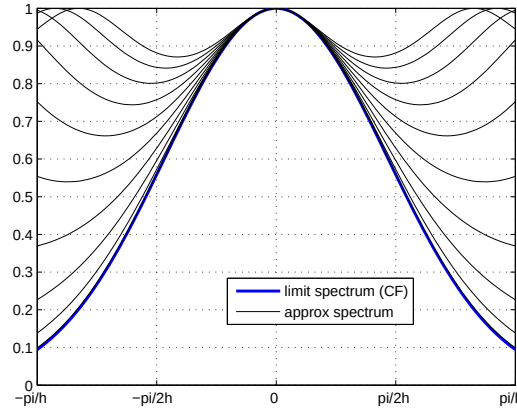


Figure 2.1: Convergence of the discrete spectrum of the Markov generator to the continuous one as $h \downarrow 0$.

2.3.4 Convergence of the derivative price

The weak convergence on the level of the process implies, for f continuous function, that the marginal distributions of $X^{(n)}$ will weakly converge to the distribution of the process X and therefore the values of European options converge, that is

$$\mathbb{E}_x[f(X_T^n)] \rightarrow \mathbb{E}_x[f(X_T)] \quad (2.25)$$

2.4 Multidimensional CTMC and Tensor Product Analysis

In this section we extend the approximation of the infinitesimal generator based on Markov chain introduced in section 3 to the case of multidimensional diffusions. This problem has been investigated by, among others, Lo [2009], that proposes a discretization scheme for diffusion processes in $\mathbb{R}^d$ based on a continuous-time Markov chain and points out that it is not straight-forward to extend the single dimension discretization scheme to multiple-dimensions. However in Lo [2009] the correlation is not modeled via a scheme extended to the cross space but by projecting the information of the correlated processes on the orthogonal dimensions. This can be obtained via the Cholesky decomposition of the covariance matrix. In this way Lo [2009], by restricting the generator of the Markov chain to be a tridiagonal matrix, is able to use the same discretization scheme as the one in one dimension Albanese and Mijatovic [2006b]. Lo [2009] presents a condition for the convergence of diffusion approximation to general multidimensional diffusions, that is the theorem 4.1 pag. 354 of Ethier and Kurtz [1985].

We start our analysis by reviewing some tensor algebra that is based on matrix manipulation because it turns to be a key tool to represent multidimensional diffusions when approximated by CTMC's. We continue by illustrating how to build multidimensional approximate generators through practical examples, focusing first on the case of bivariate diffusions with no correlation, second on the correlated case, and then deriving a more general procedure for larger dimensions. Lo [2009] presents also an example, where he builds an approximate generator for two correlated diffusions. We will closely point out, referring to this specific example, the differences between the approach in Lo [2009] and the one we are proposing.

Definition 2.2 *The tensor (Kronecker) product of two rectangular matrices $A \in \mathbb{R}^{n_A \times m_A}$ and $B \in \mathbb{R}^{n_B \times m_B}$ is written as*

$$C = A \otimes B$$

and yields to the rectangular matrix $C \in \mathbb{R}^{n_A n_B \times m_A m_B}$ whose elements satisfy

$$c(i_C, j_C) = a(i_A, j_A)b(i_B, j_B)$$

with $i_C = (i_A - 1)n_B + i_B$ and $j_C = (j_A - 1)m_B + j_B$, for $(i_A, j_A) \in \{1, 2, \dots, n_A\} \times \{1, 2, \dots, m_A\}$ and $(i_B, j_B) \in \{1, 2, \dots, n_B\} \times \{1, 2, \dots, m_B\}$, where $\otimes$ is the Cartesian product operator. Note that in a 2-dimensional representation, the row indices of C are in $\{1, 2, \dots, n_A\} \times \{1, 2, \dots, n_B\}$, whereas its columns indices are in $\{1, 2, \dots, m_A\} \times \{1, 2, \dots, m_B\}$. Hence, the ordering of rows and columns of C with respect to this 2-dimensional representation is lexicographical, since

$$c(i_C, j_C) = c((i_A, i_B), (j_A, j_B)) = c((i_A - 1)n_B + i_B, (j_A - 1)m_B + j_B). \quad (2.26)$$

The tensor product can be written also as a matrix in block partition form:

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1m_A}B \\ a_{21}B & & a_{2m_A}B \\ \vdots & & \vdots \\ a_{n_A1}B & \dots & a_{n_A m_A}B \end{bmatrix} \quad (2.27)$$

A detailed account of properties of tensor products and vector spaces is given in Halmos [1947]. Some elementary properties are:

$$\begin{aligned} (A + B) \otimes C &= A \otimes B + B \otimes C, & (A \otimes B)^{-1} &= A^{-1} \otimes B^{-1} \\ (A \otimes B)(C \otimes D) &= AC \otimes BD, & A \otimes (B + C) &= A \otimes B + B \otimes C \end{aligned}$$

We now propose an example in order to show how this formalism can be very effective to characterize a boundary problem and express its solution.

Example (Uncorrelated diffusions) Let's consider a second order linear parabolic partial differential equation in two independent variables x and y ,

$$\frac{du}{dt} = \mathcal{L}u \quad (2.28)$$

in which the differential operator $\mathcal{L}$ has the form of

$$\mathcal{L} = \mathcal{L}_x + \mathcal{L}_y$$

where $\mathcal{L}_x$ ($\mathcal{L}_y$) is an operator with derivatives only on x (y). In line with eq.(2.5), $\mathcal{L}_x$ is of the form

$$\mathcal{L}_x = \mu(x) \frac{d}{dx} + \sigma(x) \frac{d^2}{dx^2}$$

We treat the problem consisting of eq.(2.28) in the rectangle

$$R = \{(x, y) | \underline{x} \leq x \leq \bar{x}, \underline{y} \leq y \leq \bar{y}\} \quad (2.29)$$

and any number of conditions specified on the boundary of R such as Dirichlet, mixed, or periodic boundary conditions.

We approximate $\mathcal{L}$ in eq.(2.28) with the discrete operator L over a set of discrete points in R , as described in section (2.2). However in this specific bidimensional example, we first approximate one dimension per time and then build the entire operator. Therefore we compute the matrices L_x and L_y that are known explicitly and for diffusions² are sparse and squared tridiagonal matrices. Assuming constant coefficients for the drift and the volatility within each operator, namely $\mu_x(x) = \mu_x$, $\mu_y(y) = \mu_y$, $\sigma_x(x) = \sigma_x$ and $\sigma_y(y) = \sigma_y$, the matrices L_x and L_y are fully specified,

$$L_x = \{\lambda_x(i, j)\}_{i,j=1,\dots,n} = L_x(S_x, \mu_x, \sigma_x) \quad L_y = \{\lambda_y(i, j)\}_{i,j=1,\dots,m} = L_y(S_y, \mu_y, \sigma_y)$$

where S_x and S_y are the grids respectively over the dimension x and y . Note that the spacing between points need not to be uniform, see Tavella and Randall [2000]. Then we build the matrix L as

$$L = I_{(y)} \otimes L_x + L_y \otimes I_{(x)} \quad (2.30)$$

where $I_{(\cdot)}$ denotes identity matrices of appropriate sizes. The result is an ordinary difference equation which can be written as

$$\frac{dp}{dt} = Lp \quad (2.31)$$

²See section 2.2.

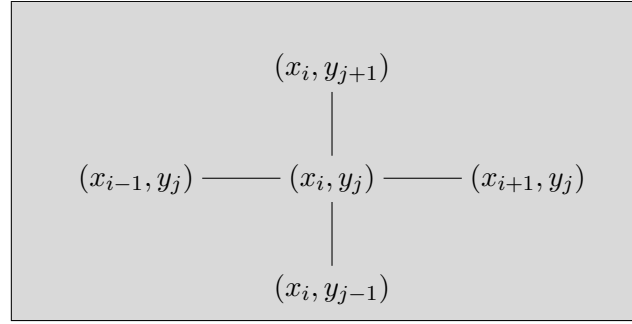


Figure 2.2: States used in the bidimensional discretization scheme of eq.(2.30). The dimensions x and y are orthogonal, and applies for all i and j .

where p is the approximation to the solution u at the discretization points. In particular, given an initial condition p_0 , the approximated solution reads

$$p_t = e^{Lt} p_0$$

The solution at time t can be readily computed calculating

$$p_t = \exp(I_{(y)} \otimes L_x + L_y \otimes I_{(x)}) p_0 \quad (2.32)$$

where p_0 is a vector that define the starting condition at time $t = 0$. Furthermore the dimensionality of the problem is as following: $L_x \in \mathbb{R}^{n \times n}$, $L_y \in \mathbb{R}^{m \times m}$, $I_{(x)} \in \mathbb{R}^{n \times n}$, $I_{(y)} \in \mathbb{R}^{m \times m}$, with $n, m \in \mathbb{N}$, $L, p_t \in \mathbb{R}^{mn \times mn}$.

Remark 2.3 *The amount of computational time necessary for evaluating the solution (2.33) only depends on the technique adopted to compute the exponential of a very large sparse matrix. At this regard, in the next subsection 3.4.1, we provide an extensive investigation of numerical methods to compute matrix exponential, and our privileged approach is based on Krylov subspace approximation.*

An extension of this example to multiple dimension is very straightforward. In fact supposing we have a third dimension z with approximated dynamics fully specified by the matrix operator L_z , we can rewrite a solution to the boundary problem

$$p_t = \exp\left(I_{(z)} \otimes (I_{(y)} \otimes L_x + L_y \otimes I_{(x)}) + L_z \otimes I_{(x,y)}\right) p_0 \quad (2.33)$$

We no show how to solve the same multidimensional problem but this time with correlated variables.

Example(Correlated diffusions) Let's consider a two dimensional parabolic boundary value problem in the same way as in the previous section but with correlated variables. The differential operator $\mathcal{L}$, that is the operator of two correlated difussions, now has the form

$$\begin{aligned} \mathcal{L} &= \mu_x(x) \frac{d}{dx} + \sigma_x(x) \frac{d^2}{dx^2} + \mu_y(y) \frac{d}{dy} + \sigma_y(y) \frac{d^2}{dy^2} + \rho(x, y) \sigma_x(x) \sigma_y(y) \frac{d^2}{dxdy} \\ &= \mathcal{L}_x + \mathcal{L}_y + \rho(x, y) \sigma_x(x) \sigma_y(y) \frac{d^2}{dxdy} \end{aligned} \quad (2.34)$$

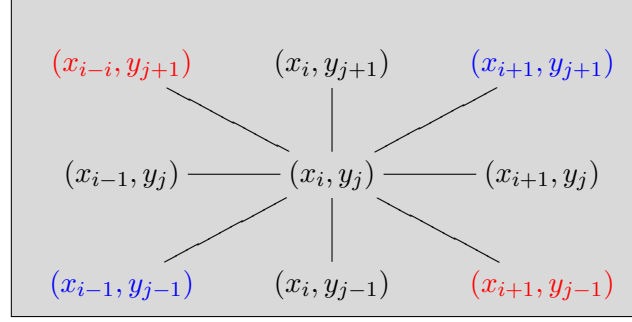


Figure 2.3: States used in the bidimensional discretization scheme of eq.(2.34) considering cross states. The scheme allows for positive (blue) and negative (red) correlation between dimensions x and y , and applies for all i and j .

and we want to solve this problem in the rectangle R of eq.(2.29), given a boundary condition. At this purpose we extend the analysis, illustrated through the previous example, and consider the contribution to the cross term $\rho(x, y)\sigma_x(x)\sigma_y(y)\frac{d^2}{dxdy}$ in the discretization scheme of eq.(2.30). This means that we take into account not only of the process moving along the states of each orthogonal dimension but also along the cross states, where the process can jump during the infinitesimal time dt . We want to approximate $\mathcal{L}$ in eq.(2.34), with the discrete operator L_{xy} over a set of discrete point in R .

To this end, we introduce an auxiliary variable Z , (useful for the bidimensional generator construction) that specifies the dynamics of the multivariate process (X, Y) in its cross-space, and it is fully defined by the covariance of (X, Y) . The intuition behind the approach we propose can be expressed through the Bayes' rule as,

$$P(X, Y) = \int P(X|Z)P(Y|Z)dz \quad (2.35)$$

Therefore, for this bidimensional case, we need three building blocks for L_{xy} :

1. $L_x^{(c)}$, a conditional generator matrix for X , conditional on the process Z .
2. $L_y^{(c)}$, a conditional generator matrix for Y , conditional on the process Z .
3. $L_{xy}^{(c)(+/-)}$, an auxiliary generator for the process Z .

We calculate the entries for the tridiagonal matrices $L_x^{(c)}$ and $L_y^{(c)}$, as shown in section 2.2, conditionally on the information in the cross space, that is simply reflected in the variance,

$$L_x^{(c)} = \{\lambda_x^{(c)}(i, j)\}_{i,j=1,\dots,n} = L_x^{(c)}(S_x, \mu_x, \sqrt{\sigma_x^2 - \rho\sigma_x\sigma_y}) \in \mathbb{R}^{n \times n} \quad (2.36)$$

$$L_y^{(c)} = \{\lambda_y^{(c)}(i, j)\}_{i,j=1,\dots,m} = L_y^{(c)}(S_y, \mu_y, \sqrt{\sigma_y^2 - \rho\sigma_x\sigma_y}) \in \mathbb{R}^{m \times m} \quad (2.37)$$

where $S_x(S_y)$ corresponds to the discretization of the rectangle R in the $x(y)$ direction. The matrices $L_x^{(c)}$ and $L_y^{(c)}$ are used for the construction of the conditional bidimensional generator $L_{xy}^{(c)}$. The generator $L_{xy}^{(c)}$ approximates the operator L_{xy} in the two orthogonal

dimensions x and y conditionally on the information in the cross space, namely the instantaneous covariance of the process $\rho\sigma_x\sigma_y$ and is a penta-diagonal matrix with entries given by

$$L_{xy}^{(c)} = \{\lambda_{xy}^{(c)}(i, j)\}_{i,j=1,\dots,nm} = (I_{(y)} \otimes L_x^{(c)} + L_y^{(c)} \otimes I_{(x)}), \quad (2.38)$$

Let's now define the steps to build the matrices $L_{xy}^{(c)(+)} \in \mathbb{R}^{nm \times nm}$ and $L_{xy}^{(c)(-)} \in \mathbb{R}^{nm \times nm}$ of eq.(2.41). Such matrices entries simply impose the instantaneous covariance of the process $\rho\sigma_x\sigma_y$ over the cross states of the bidimensional space (X, Y) . Let I_m and I_p be matrices containing only a nonzero diagonal, namely

$$I_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \quad I_m = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

and they are important to populate the matrix entries that correspond to the cross states of the xy-space. In the case of positive correlation we define the cross-space matrix

$$\begin{aligned} L_{xy}^{(c)(+)}(i, j) &= \{\lambda_{xy}^{(c)(+)}(i, j)\}_{i,j=1,\dots,nm} \\ &= \left(\left(I_m^{(y)} \sigma_y \rho \right) \otimes \left(I_m^{(x)} \frac{\sigma_x}{2h^2} \right) + \left(I_p^{(y)} \sigma_y \rho \right) \otimes \left(I_p^{(x)} \frac{\sigma_x}{2h^2} \right) \right)(i, j) \end{aligned} \quad (2.39)$$

where, h is the state space discretization unit and $h^2 = h_x h_y$ in case the discretization unit along the x-dimension differs from the one along the y-dimension, and $I_m^{(y)}$ denotes a matrix of type I_m and size consistent with the approximated matrix operator for the dimension y , in this example is a matrix $I_m^{(y)} \in \mathbb{R}^{m \times m}$. In the case of negative correlation we define in the same fashion as above the cross-space matrix

$$\begin{aligned} L_{xy}^{(c)(-)}(i, j) &= \{\lambda_{xy}^{(c)(-)}(i, j)\}_{i,j=1,\dots,nm} \\ &= \left(\left(I_m^{(y)} \sigma_y \otimes I_p^{(x)} \frac{\sigma_x}{2h^2} \right) + \left(I_p^{(y)} \sigma_y \rho \otimes I_m^{(x)} \frac{\sigma_x}{2h^2} \right) \right)(i, j) \end{aligned} \quad (2.40)$$

The entries of the matrices in eq. (2.39) and eq. (2.40) are computed by matching the instantaneous covariance of the process, using eq.(2.36). The infinitesimal generator for the problem in eq.(2.34) can be therefore expressed as the matrix

$$L_{xy} = \begin{cases} L_{xy}^{(c)} + L_{xy}^{(c)(+)}, & \text{if } \rho > 0 \\ L_{xy}^{(c)} + L_{xy}^{(c)(-)}, & \text{if } \rho < 0 \end{cases} \quad (2.41)$$

where $L_{xy}^{(c)}$ is the term acting on the orthogonal x-y space, while the terms $L_{xy}^{(c)(+)}$ and $L_{xy}^{(c)(-)}$ act on the cross x-y space.

For clarity, we illustrate how the intensities approximating the process dynamics, as defined above, are located in the bidimensional discretized state space R . The following matrix is a squared representation of a row $L_{xy}(i, \cdot) \in \mathbb{R}^{n \times m}$ of the L_{xy} generator. The matrix reports all the non-zero intensities across a L_{xy} row as explained through the equations in this section. The central/diagonal term $-\sum \lambda$ is equal and opposite in sign

to the sum all the extra-diagonal elements on the row, and this condition guarantees probability conservation.

$$L_{xy}(i, \cdot) = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & & & \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{xy}^{(c)(-)} & \lambda_y^{(\text{pos})} - \overline{\lambda_{xy}^{(c)}} & \lambda_{xy}^{(c)(+)} & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & \lambda_x^{(\text{neg})} - \overline{\lambda_{xy}^{(c)}} & -\sum \lambda & \lambda_x^{(\text{pos})} - \overline{\lambda_{xy}^{(c)}} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{xy}^{(c)(+)} & \lambda_y^{(\text{neg})} - \overline{\lambda_{xy}^{(c)}} & \lambda_{xy}^{(c)(-)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \ddots & \vdots & \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where $\lambda_{xy}^{(c)} = \lambda_k^{(\xi)} - \overline{\lambda_{xy}^{(c)}}$, with $\xi = \text{pos, neg}$, $k = x, y$, $\lambda_y^{(\text{pos})}$ and $\lambda_y^{(\text{neg})}$ represent the unconditional forward and the backward intensities respectively along the y-dimension and $\lambda_x^{(\text{pos})}$ and $\lambda_x^{(\text{neg})}$ represent the unconditional forward and the backward intensities respectively along the x-dimension. The entries $\lambda_{xy}^{(c)(+)}$ correspond to eq.(2.39), while $\lambda_{xy}^{(c)(-)}$ correspond to eq.(2.40), and they occupy the cross space non-simultaneously, as reported in eq.(2.41).

Remark 2.4 *The only condition to impose to this approach is the integrability condition of the auxiliary variable Z , to ensure that the resulting integrated variance is positive, and that all the extradiagonal intensities entries of the approximate generator matrix are positive. However such condition is satisfied in this case by conditional independence and reflected into the construction of the generator matrices L_x^c and L_y^c .*

Remark 2.5 (Approach comparison) *In Lo [2009] the author shows in an example, illustrative of his proposed methodology to approximate the generator of multidimensional diffusions based on Markov chains, how to build the generator for a bivariate diffusion. To this end he calculates a set of intensities that describes the instantaneous jumps of the (X, Y) process along the orthogonal directions x and y , and they are:*

$$\begin{aligned} \lambda_x^{(\text{pos})} &= \frac{\sigma_x^2}{2h^2} + \frac{\mu_x}{2h} & \lambda_x^{(\text{neg})} &= \frac{\sigma_x^2}{2h^2} - \frac{\mu_x}{2h} \\ \lambda_y^{(\text{pos})} &= \frac{\sigma_y^2(1-\rho^2)}{2h^2} + \frac{\mu_y + \frac{\sigma_y\rho}{\sigma_x}\mu_x}{2h} & \lambda_y^{(\text{neg})} &= \frac{\sigma_y^2(1-\rho^2)}{2h^2} - \frac{\mu_y + \frac{\sigma_y\rho}{\sigma_x}\mu_x}{2h} \end{aligned}$$

Note that the intensities $\lambda_y^{(\cdot)}$ are simply calculated from the instantaneous moments of the conditional distribution of Y given X , in the bivariate normal case. In fact

$$Y|X = x \sim N\left(\mu_y + \frac{\sigma_y\rho}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right)$$

Expressing the intuition behind this approach through the Bayes' rule we have

$$P(X, Y) = P(Y|X)P(X) \quad (2.42)$$

This method guarantees that the joint transition kernel calculated through the bidimensional infinitesimal generator will exhibit the specified correlation between the marginals. However it is not clear in Lo [2009] whether $P(Y) = \int P(X, Y)dy$.

We can generalize our example to cases with higher dimensionality. The operator of this problem is the d-dimensional infinitesimal generator in eq.(2.5). At this end, the first step to build the generator matrices $L_k^{(c)}$ that approximate the dynamics of the multivariate process along each orthogonal dimension conditional on the instantaneous covariances, namely

$$L_k^{(c)} = \{\lambda_k^{(c)}(i, j)\}_{i,j=1,\dots,n_k} = L_k^{(c)} \left(S_k, \mu_k, \sqrt{\sigma_k^2 - \sum_{z,z \neq k}^d \rho \sigma_k \sigma_z} \right) \in \mathbb{R}^{n_k \times n_k} \quad k = 1, \dots, d. \quad (2.43)$$

The second step is to build two auxiliary generator matrices that define the intensities in the cross-space, corresponding to the instantaneous covariance, for every pair of dimensions

$$L_{kz}^{(c)(+)}(i, j) \quad L_{kz}^{(c)(-)}(i, j) \quad \text{for all } z, \text{ given } k \quad (2.44)$$

where the indices (k, z) span the whole d-dimensional space. For notation convenience we denote the matrices in eq.(2.44) as $L_{kz}^{(c)(\pm)}$.

We can then rewrite a solution to the multidimensional boundary problem using the following recursion

$$\begin{aligned} A^{(2)} &= I_{(2)} \otimes L_1^{(c)} + L_2^{(c)} \otimes I_{(1)} + L_{1z}^{(c)(\pm)}, \quad z = 2 \\ A^{(3)} &= I_{(3)} \otimes A^{(2)} + L_3^{(c)} \otimes I_{(1,2)} + L_{3z}^{(c)(\pm)} \quad z = 1, 2 \\ A^{(4)} &= I_{(4)} \otimes A^{(3)} + L_4^{(c)} \otimes I_{(1,2,3)} + L_{4z}^{(c)(\pm)} \quad z = 1, 2, 3 \\ &\vdots \\ A^{(d)} &= I_{(d)} \otimes A^{(d-1)} + L_d^{(c)} \otimes I_{(1,\dots,d)} + L_{d-1z}^{(c)(\pm)} \quad z = 1, \dots, d-1 \end{aligned}$$

$$p_t = e^{tA^{(d)}} p_0 \quad (2.45)$$

2.4.1 Computation of approximated diffusive semigroups

Once we numerically compute all the operator's entries $\mathcal{A}^{(n)} = \{a(x_i, x_j)\}$ and obtain the approximated infinitesimal generator $\mathcal{A}^{(n)} := A \in \mathbb{R}^{n \times n}$ the numerical problem we face is of the type

$$\begin{cases} \frac{dp(t)}{dt} = Ap(t), & t \in [0, T] \\ p(0) = v, & \text{initial probability distribution} \end{cases}$$

Its analytic solution is $p(t) = e^{tA}v$ and represent the transient probability of the Markov chain with n states. A direct way to define the matrix exponential e^{tA} where t is a real number is undoubtedly through the power series expansion,

$$e^{tA} = I + \sum_{k=1}^{\infty} \frac{(tA)^k}{k!}, \quad (2.46)$$

where I is the identity matrix and whose convergence is guaranteed for any squared matrix A . The algorithm in this case consists in building a finite series of the type

$$P_m = \sum_{k=1}^m \frac{(tA)^k}{k!},$$

and then find m such that,

$$\|P_m - P_{m+1}\| \leq \epsilon$$

where ϵ is some desired error tolerance. However, such an algorithm is known to be inefficient even in the scalar case and it is problematic to decide where to truncate the Taylor series because directly summing up the series terms seems not to work very well Moler and Loan [1978], Liou [1966]. Another method is to diagonalize the matrix A and to retrieve the associated discrete probability density kernel at any time t as

$$\begin{aligned} e^{At} &= e^{V\Lambda V^{-1}t} \\ &= 1 + V\Lambda V^{-1}t + \frac{1}{2!}V\Lambda^2 V^{-1}t^2 + \frac{1}{3!}V\Lambda^3 V^{-1}t^3 + \dots \\ &= V e^{t\Lambda} U, \end{aligned}$$

where V is the right eigenvectors, column wise and Λ contains the corresponding eigenvectors, represented as a diagonal matrix with elements in the order as the column vector in V , and U is the inverse of V . The exponent $e^{t\Lambda}$ is very easy to compute and there are satisfactory built-in routines in many computer languages that calculate the exponential of a scalar,

$$e^{t\Lambda} = \text{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t})$$

The difficulty with this approach is not find the eigenvalues of the matrix A in itself, but occurs when A does not have a complete set of linearly independent eigenvectors. In this case, A is defective and there is no invertible matrix of V , therefore, the algorithm will break down. In real computing world, the difficulties occur even when A is nearly defective. Define the condition number as $\text{cond}(V) = \|V\| \|V^{-1}\|$. While A is nearly defective, then $\text{cond}(V)$ will be very large. The errors of computing e^A , including the roundoff errors from the eigenvalues computation, may be magnified by $\text{cond}(V)$. As a result, the computed exponential of a matrix will most likely be inaccurate when $\text{cond}(V)$ is very large. An example had been demonstrated in this paper Moler and Loan [1978]. In order to overcome the inaccuracy of terminating the Taylor series for e^{tA} , we can use the Padé approximant of the function e^{tA} . Padé approximant is the ‘best’ approximation of a function by a rational function of given order. Developed by Henri Padé, a Padé approximant often gives better approximation of the function than truncating its Taylor series, and it may still work where the Taylor series does not converge. For these reasons Padé approximants are used extensively in computer calculations. A Padé approximation approximates a function in only one variable, an approximation of a function in two variables is called a Chisholm approximation, and in multiple variables is called a Canterbury approximation.

Definition 2.3 (*Padé Approximant*) *A Padé rational approximation to $f(x)$ is the quotient of two polynomials $N_p(x)$ and $D_q(x)$ of degrees p and q respectively. This is so-called (p, q) -degree type Padé approximation and we denote the quotient as*

$$R_{p/q}(x) = \frac{N_p(x)}{D_q(x)} \quad (2.47)$$

where x is a scalar. In the case of a matrix A , the quotient is defined as

$$R_{p/q}(A) = (D_q(A))^{-1} N_p(A) \quad (2.48)$$

where the polynomial in eq.(2.48) are

$$\begin{aligned} N_p(A) &= b_0 I + b_1 A + b_2 A^2 + \dots + b_p A^p \\ D_q(A) &= I + c_1 A + c_2 A^2 + \dots + c_q A^q \end{aligned}$$

Padé approximants are derived by expanding a function as a ratio of two power series and determining both the numerator and denominator coefficients. Padé approximations are usually superior to Taylor expansions when functions contain poles, because the use of rational functions allows them to be well-represented. In the case $q = 0$, the approximation (2.48) will reduce to the Maclaurin series for $f(A)$. There are $p + 1$ unknown coefficients in $N_p(A)$, and q unknown coefficients in $D_q(A)$, hence the rational function $R_{p/q}(A)$ has $p + q + 1$ unknown coefficients. Assume that $f(A)$ is analytic and has the Maclaurin expansion

$$f(A) = a_0 I + a_1 A + a_2 A^2 + \dots + a_k A^k + \dots$$

The coefficient in eq.(2.48) are found by setting

$$f(A) - (D_q(A))^{-1} N_p(A) = 0$$

In this way we are performing the approximation

$$\sum_{i=0}^p b_i A^i - \left(I + \sum_{i=0}^q c_i A^i \right) \left(\sum_{i=0}^{\infty} a_i A^i \right) = 0 \quad (2.49)$$

where we set to zero all the coefficient of the power of A^i for $i = 0, 1, \dots, p$, and the result is a system of $p + q + 1$ equations:

$$\begin{aligned} a_0 - b_0 &= 0 \\ a_0 c_1 + a_1 - b_1 &= 0 \\ a_0 c_2 + a_1 c_1 + a_2 - b_2 &= 0 \\ \vdots \quad \vdots \quad \vdots & \\ c_q a_{p-q} + c_{q-1} a_{p-q+1} + \dots + a_p - b_p &= 0 \end{aligned} \quad (2.50)$$

$$\begin{aligned} c_q a_{p-q+1} + c_{q-1} a_{p-q+2} + \dots + c_1 a_p + a_{p+1} &= 0 \\ c_q a_{p-q+2} + c_{q-1} a_{p-q+3} + \dots + c_1 a_{p+1} + a_{p+2} &= 0 \\ \vdots \quad \vdots \quad \vdots & \\ c_q a_p + c_{q-1} a_{p+1} + \dots + c_1 a_{p+q+1} + a_{p+q} &= 0 \end{aligned} \quad (2.51)$$

The q-equations in eq.(2.51) involve only the unknowns $(c_1, c_2, \dots, c_q)$ and have to be solved first. Then the eq.(2.50) are used successively to find $(b_1, b_2, \dots, b_p)$. Solving eq.(2.50) and eq.(2.51) together gives us,

$$N_p(A) = \sum_{i=0}^p \frac{(p+q-i)! p!}{(p+q)! i! (p-i)!} A^i \quad (2.52)$$

$$D_q(A) = \sum_{i=0}^q \frac{(p+q-i)! q!}{(p+q)! i! (q-i)!} A^i \quad (2.53)$$

It has been investigated that the diagonal approximant with $p = q$, are preferred over the off diagonal approximants, $p \neq q$ for stability and economy of computation Moler and Loan [1978]. For $p = q$ we have $b_0 = 1$, $b_i = b_{i-1} \frac{p+1-i}{(2p+1-i)i}$, and $c_i = (-1)^i n_i$. Furthermore eq.(2.48) can be written, Sidje [1994] as the following irreducible form

$$R_{p/p}(A) = \begin{cases} 1 + 2 \left(\sum_{i=0}^{p/2} b_{2i} A^{2i} - A \sum_{i=0}^{p/2-1} b_{2i+1} A^{2i} \right)^{-1} \left(A \sum_{i=0}^{p/2-1} b_{2i+1} A^{2i} \right), & \text{if } p \text{ is even} \\ -1 - 2 \left(\sum_{i=0}^{(p-1)/2} b_{2i+1} A^{2i} - \sum_{i=0}^{(p-1)/2} b_{2i} A^{2i} \right)^{-1} \left(\sum_{i=0}^{(p-1)/2} b_{2i} A^{2i} \right), & \text{if } p \text{ is odd} \end{cases}$$

Remark 2.6 *It is important to remind that the Padé rational approximation is only accurate near the origin, so that the approximation e^{tA} is not valid, when $\|tA\|$ is too large.*

This problem can be solved introducing the so-called ‘squaring and scaling’ technology Moler and Loan [1978], Sidje [1994]. We use the exponential property that

$$e^{tA} = \left(e^{2^{-s}tA} \right)^{2^s} \approx \left(R_{p/p}(2^{-s}tA) \right)^{2^s} \quad (2.54)$$

where s is chosen such that $\|2^{-s}tA\| \leq \frac{1}{2}$. The idea is to choose s to be a natural number for which $M = e^{2^{-s}tA}$ can be reliably and efficiently computed, and then compute the result $e^{tA} = M^{2^s}$ by repeated squaring. Some authors refer to this method as ‘fast exponentiation’ (p.19 Albanese [2007]), however it is important to take into account of the roundoff error. In fact in eq.(2.54) we have two parameters, p and s to choose. Because the result is evaluated by repeated squaring the exponential of scaled matrix, the drawback of this algorithm may come from the fact that if $s \gg 1$, then the roundoff errors may be large. In Moler and Loan [1978] is shown that if $\|tA\| \leq 2^{s-1}$, then

$$\left(R_{p/p}(2^{-s}tA) \right)^{2^s} = e^{tA+E}$$

where

$$\frac{\|E\|}{\|A\|} \leq 8 \left(\frac{\|A\|}{2^s} \right)^{2p} \left(\frac{(p!)^2}{(2p)!(2p+1)!} \right) \approx \begin{cases} 0.34 \times 10^{-15}, & p=6 \\ 0.11 \times 10^{-18}, & p=7 \\ 0.27 \times 10^{-22}, & p=8 \end{cases}$$

Therefore, a value of $p = 6$ is generally satisfactory for computer computing while using double precision.

Remark 2.7 *The Padé approximation combined with scaling and squaring will be used to compute e^{tA} . Furthermore this method of exponentiation will be the building block of more computationally involved algorithms needed for approximate the matrix exponentials in problems of very large scale.*

2.4.2 Subspace approximations to the matrix exponential

We would like to tackle the problem in eq.(2.9) and (2.9) with a Krylov subspace approximation to the matrix exponential operation $e^A v$. The basic principle of the Krylov subspace techniques is to approximately project the exponential of a very large matrix onto a small Krylov subspace. The only matrix exponential operation performed is

therefore with a much smaller matrix Saad [1990]. The idea is to approximate the matrix exponential $e^A v$ as

$$e^A v \approx w_{m-1}(A)v \quad (2.55)$$

where A is a matrix of dimension n , v an arbitrary nonzero vector, and w_{m-1} is a polynomial of degree $m-1$. Consider a large sparse n -by- n matrix A , a n -vector v and a scalar $t \in \mathbb{R}$. The underlying principle is to approximate, given an integer m , the Taylor truncated expansion at order $(m-1)$

$$p(t) = e^{tA} v = v + \frac{(tA)}{1!} v + \frac{(tA)^2}{2!} v + \dots \quad (2.56)$$

thereby yielding a polynomial approximation of degree $(m-1)$

$$c_0 v + c_1 (tA)v + c_2 (tA)^2 v + \dots + c_{m-1} (tA)^{m-1} v \quad (2.57)$$

with coefficients $\{c_i = \frac{1}{i!}\}$. However these are not necessarily the best coefficients and one can set out to look for a better linear combination. All polynomial approximations of degree at most $(m-1)$ (including therefore the truncated Taylor polynomial as well as the optimal polynomial approximation) are elements of the Krylov subspace of dimension m , defined as

$$\mathcal{K}_m(tA, v) = \text{span}\{v, (tA)v, \dots, (tA)^{m-1}v\} \quad (2.58)$$

Hence it is more general to state the problem as that of finding an element of $\mathcal{K}_m(tA, v)$ that approximates $p(t)$. Elements of the Krylov subspace are better manipulated via their representation onto an orthonormal basis.

Remark 2.8 *The dimension of the Krylov subspace m , is small compared to n . Usually $m \leq 30$ while n can exceed many thousands.*

Iterative methods for finding one (or a few) eigenvalues of large sparse matrices or solving large systems of linear equations avoid matrix-matrix operations, but rather multiply vectors by the matrix and work with the resulting vectors. Starting with a vector, v , one computes tAv , then one multiplies that vector by tA to find $(tA)^2 v$ and so on. Because the vectors tend very quickly to become almost linearly dependent, methods relying on Krylov subspace frequently involve some orthogonalization scheme, such as Lanczos iteration for Hermitian matrices or Arnoldi iteration for more general matrices.

```

Input: Matrix  $A$  and vector  $v$ 
Output: Matrix  $H_m$ , upper Hessenberg Matrix
 $v_1 := \frac{v}{\|v\|_2};$ 
for  $j = 1 : m$  do
     $w := Av_j;$ 
    for  $i = 1 : j$  do
        Calculate  $h_{ij} = v_i' w;$ 
        Calculate  $w = w - h_{ij} v_i;$ 
    end
     $h_{j+1,j} = \|w\|_2;$ 
    if  $h_{j+1,j} \leq tol \|A\|$  then
        Break;
    end
     $v_{j+1} = \frac{w}{h_{j+1,j}};$ 
end

```

Algorithm 1: Arnoldi algorithm.

The Arnoldi algorithm Golub and Loan [1996], generates an orthogonal basis of the Krylov subspace using $v_1 := \frac{v}{\|v\|_2}$ as an initial vector. The inner for loop in Algorithm 1 is nothing but a modified Gram-Schmidt process. The above algorithm produces an orthonormal basis $V_m = [v_1, v_2, \dots, v_m]$ of the Krylov subspace K_m . If we denote by $H_m \in \mathbb{R}^{m \times m}$ the upper Hessenberg matrix consisting of the coefficients h_{ij} computed by the algorithm, we have the relation

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m' \quad (2.59)$$

from which we get $H_m = V_m' A V_m$. Therefore H_m represents the projection of the linear transformation A onto the subspace K_m , with respect to the basis V_m . Based on this, the following approximation was introduced in Gallopoulos and Saad [1992]

$$e^A v \approx \beta V_m e^{H_m} e_1 \quad (2.60)$$

Since $V_m'(\tau A)V_m = \tau H_m$ and the Krylov subspaces associated with A and τA are identical we can also write

$$\tilde{p}(t) = e^A v \approx \beta V_m e^{\tau H_m} e_1 \quad (2.61)$$

for an arbitrary scalar τ , where e_1 is the first unit basis vector, $\beta = \|v\|_2$.

Remark 2.9 It has been observed in many applications, Gallopoulos and Saad [1992], Nour-Omid [1989], Friesner et al. [1989] that even with relatively small values of m a remarkably accurate approximation can be obtained from eq.(2.61).

Remark 2.10 Once the matrix H_m determined using Algorithm 1, the explicit computation of $e^{\tau H_m}$ is handled with the irreducible rational Padé method combined with scaling and squaring as described in the previous section.

The idea of the Arnoldi iteration as an eigenvalue algorithm, is to compute the eigenvalues of the orthogonal projection of A onto the Krylov subspace. This projection is represented by H_m . The eigenvalues of H_m are called the Ritz eigenvalues. Since H_m is a Hessenberg

matrix of modest size, its eigenvalues can be computed efficiently, for instance with the QR algorithm. It is often observed in practice that some of the Ritz eigenvalues converge to eigenvalues of A . Since H_m is m -by- m , it has at most m eigenvalues, and not all eigenvalues of A can be approximated. Typically, the Ritz eigenvalues converge to the extreme eigenvalues of A . This can be related to the characterization of H_m as the matrix whose characteristic polynomial minimizes $\|p(A)v_1\|$ in the following way. A good way to get $p(A)$ small is to choose the polynomial p such that $p(x)$ is small whenever x is an eigenvalue of A . Hence, the zeros of p (and thus the Ritz eigenvalues) will be close to the eigenvalues of A . There is active research on improving the efficiency of the Krylov subspace method, and recent algorithms propose a restarted approach Afanasjew et al. [2008] that requires constant work and constant storage space per restart cycle. The aim of restarted Krylov methods is to improve the convergence. Theoretical results have shown that convergence improves with an increase in the Krylov subspace dimension m . However, an a-priori value of m which would lead to optimal convergence is not known. We will benefit of the described routine when computing the transition probability kernels from approximated infinitesimal generators of large dimension.

2.5 Numerical example: stochastic volatility

In this section we are going to apply the results developed so far to build the generator of two correlated diffusions and compute the price of a European call option on an asset with stochastic volatility as in the Heston model Heston [1993]. According to this model the spot asset at time t follows the diffusion

$$dS(t) = \mu S dt + \sqrt{v(t)} S dW_1(t) \quad (2.62)$$

where $W_1(t)$ is a Wiener process. The volatility follows a square-root process

$$dv(t) = \kappa(\theta - v(t))dt + \sigma\sqrt{v(t)}dW_2(t) \quad (2.63)$$

where $W_2(t)$ has correlation ρ with $W_1(t)$. The Heston's PDE for an option value $U(S, v, t)$ with interest rate r and market price of volatility risk λv absorbed into κ and σ , is

$$\frac{\partial U}{\partial t} = rU - \frac{1}{2}vS^2\frac{\partial^2 U}{\partial S^2} - rS\frac{\partial U}{\partial S} - \rho\sigma vS\frac{\partial^2 U}{\partial S\partial v} - \frac{1}{2}\sigma^2 v\frac{\partial^2 U}{\partial v^2} - \kappa(\theta - v)\frac{\partial U}{\partial v}. \quad (2.64)$$

The terminal condition for a call option is $U(S, v, T) = \max(S - K, 0)$. Boundary conditions are $U(0, v, t) = 0$, $U(S, v, t) \approx S - Ke^{-r(T-t)}$ for large S . In line with the previous section we study the problem on the rectangle $(S, v) = [0, \bar{S}] \times [0, \bar{v}]$, where $[0, \bar{S}]$ and $[0, \bar{v}]$ are discretized with units h_s and h_v respectively and we denote by $\mathbf{S} = [S_1, S_2, \dots, S_n] \in \mathbb{R}^{n \times 1}$ and $\mathbf{v} = [v_1, v_2, \dots, v_m] \in \mathbb{R}^{m \times 1}$ the resulting discretized sets.

We build the two matrices, conditional on the realizations of v and S ,

$$\begin{aligned} L_{s,k}^{(c)} &= L_s^{(c)}(\mathbf{S}, r, \sqrt{v_k S^2 - \rho\sigma v_k S}) \in \mathbb{R}^{n \times n} \quad k = 1, \dots, m \\ L_{v,i}^{(c)} &= L_v^{(c)}(\mathbf{v}, \kappa(\theta - v(t)), \sqrt{\sigma^2 v - \rho\sigma v S_i}) \in \mathbb{R}^{m \times m} \quad i = 1, \dots, n \end{aligned}$$

and the two auxiliary cross-space matrices $L_{sv}^{(c)(+)}$ and $L_{sv}^{(c)(-)}$. We obtain:

$$\begin{aligned}
L_{sv}^{(c)} &= (I_{(v)} \otimes L_{s,k}^{(c)} + L_{v,i}^{(c)} \otimes I_{(s)}), \quad \text{for all } i, k \\
L_{sv}^{(c)(+)} &= \left(\left(I_m^{(v)} \text{diag}(\mathbf{v}) \sigma_v \rho \right) \otimes \left(I_m^{(s)} \text{diag}(\mathbf{S}) \frac{\sigma_s}{2h_s h_v} \right) + \left(I_p^{(v)} \text{diag}(\mathbf{v}) \sigma_v \rho \right) \otimes \left(I_p^{(s)} \text{diag}(\mathbf{S}) \frac{\sigma_s}{2h_s h_v} \right) \right) \\
L_{sv}^{(c)(-)} &= \left(\left(I_m^{(v)} \text{diag}(\mathbf{v}) \sigma_v \otimes I_p^{(s)} \text{diag}(\mathbf{S}) \frac{\sigma_s}{2h_s h_v} \right) + \left(I_p^{(v)} \text{diag}(\mathbf{v}) \sigma_v \rho \otimes I_m^{(s)} \frac{\sigma_s}{2h_s h_v} \right) \right)
\end{aligned}$$

with $\sigma_s = 1$ and $\sigma_v = \sigma$. The solution given a starting condition (s_0, v_0) can be obtained by exponentiating the generator L_{sv} computed as in eq.(2.41).

We exponentiate the matrix L_{sv} using the Krylov subspace approximation, described in subsection 3.4.1. This methods allows to exponentiate very large sparse matrices, with size larger than (90000×90000) in less than 1 second, projecting the matrix into a subspace of dimension ranging between 4 and the full rank.

$N \setminus \rho$	-0.4	0	0.4	time (sec)
20	0.2112	0.2121	0.213	0.032
50	0.2158	0.2167	0.2176	0.039
100	0.2164	0.2173	0.2182	0.069
150	0.2167	0.2176	0.2184	0.102
200	0.217	0.2179	0.2188	0.196
300	0.2171	0.218	0.2189	0.342
Closed Form	0.2171	0.2180	0.2189	.068

Table 2.1: Comparison of the at the money price of an European call option computed with the Heston closed-form formula and using our proposed method. We compute the Heston closed-form formula price by inverting the characteristic function (cf. Heston [1993]) with a grid of $N = 4096$ points. The other parameters are as follows: $\kappa = 0.01$, $\theta = 0.15$, $v_0 = 0.2$, $\sigma = 0.13$, $T = 1$, $r = 0.03$. We exponentiate the matrix L_{sv} using the Krylov approximation with subspace dimension of 5 and spectral tolerance of 10^{-7} .

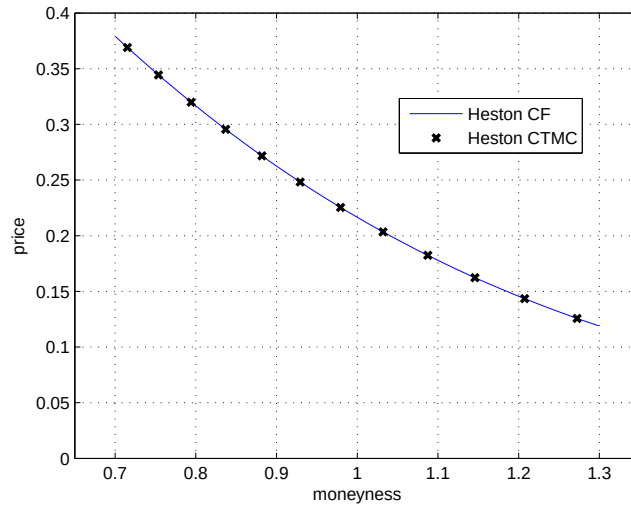


Figure 2.4: Plot of the price of an European call option computed with the Heston closed-form formula and using our proposed method. We compute the Heston closed-form formula price by inverting the characteristic function with a grid of $N = 4096$ points. We use grid of 300 points for both the underlying process and the volatility in our method. The other parameters are as follows: $\kappa = 0.01$, $\theta = 0.15$, $v_0 = 0.2$, $\sigma = 0.13$, $T = 1$, $r = 0.03$, $\rho = -.65$.

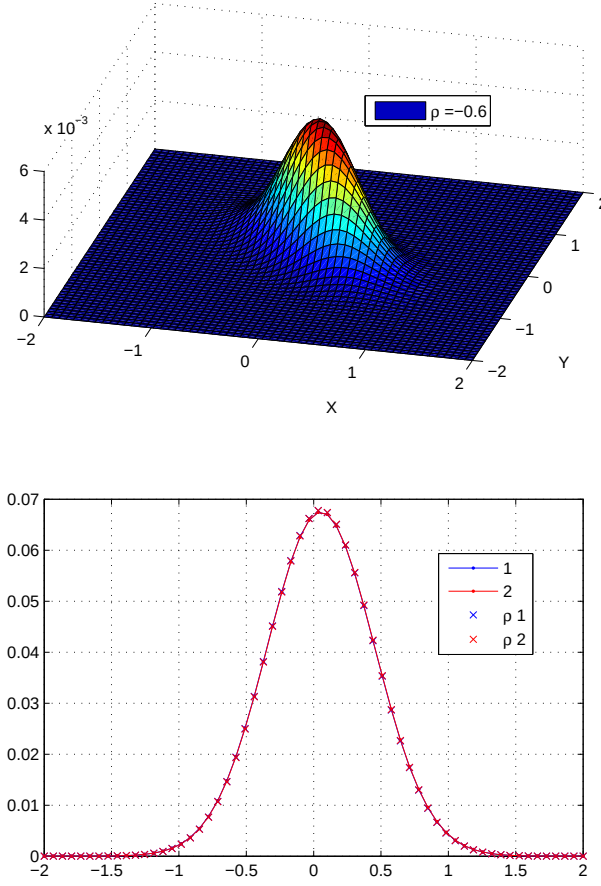


Figure 2.5: The upper plot is the transition density of two correlated Brownian motions with drift. The correlation is $\rho = -0.6$. The other parameters are $t = 1$, $\mu_1 = \mu_2 = 0.05$ and $\sigma_1 = \sigma_2 = 0.25$. The lower plot shows the consistency of the marginal distributions, that are all coincident both when correlation is $\rho = 0$ and $\rho = -0.6$. The number of discretization points is 100 per each dimension. We exponentiate the associated matrix generator of dimension 100×100 using the Krylov approximation with subspace dimension of 4 and spectral tolerance of 10^{-7} in a time of 0.062sec.

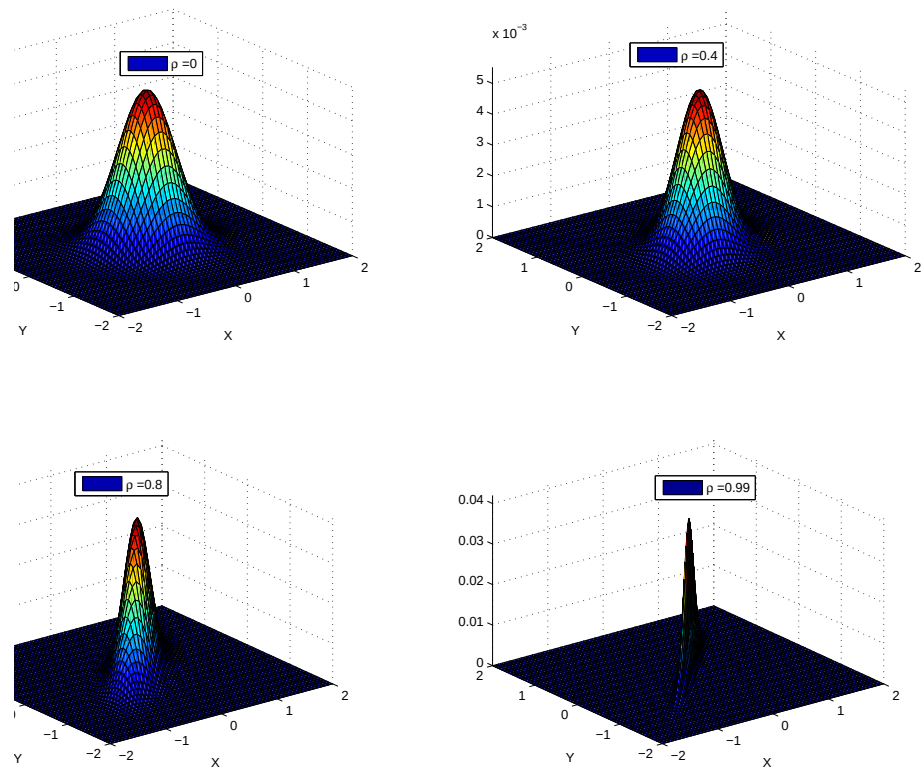


Figure 2.6: Plots of the transition density of two correlated Brownian motions with drift when the correlation changes in value.

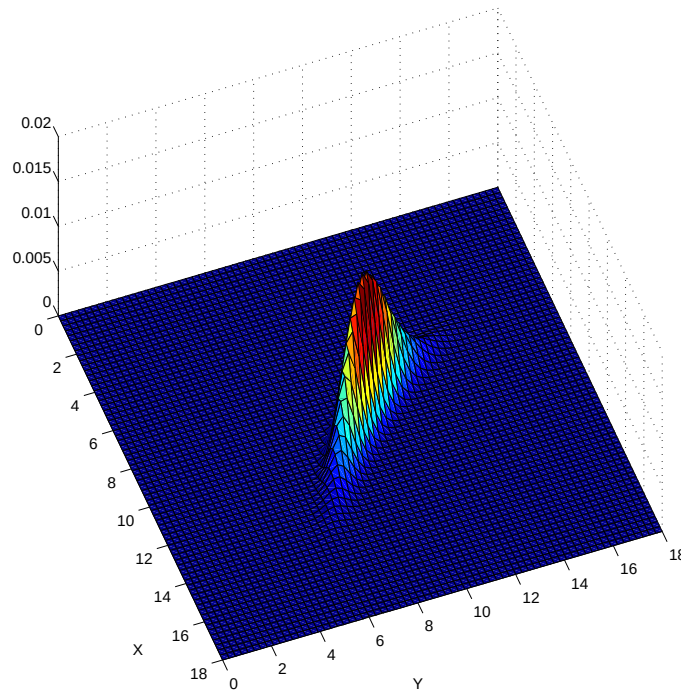


Figure 2.7: Plot of the transition density of two correlated geometric Brownian motions when $\rho = -.93$. The other parameters are $t = 1$, $\mu_1 = 0.035$, $\mu_2 = 0.07$, $\sigma_1 = 0.2$, and $\sigma_2 = 0.25$.

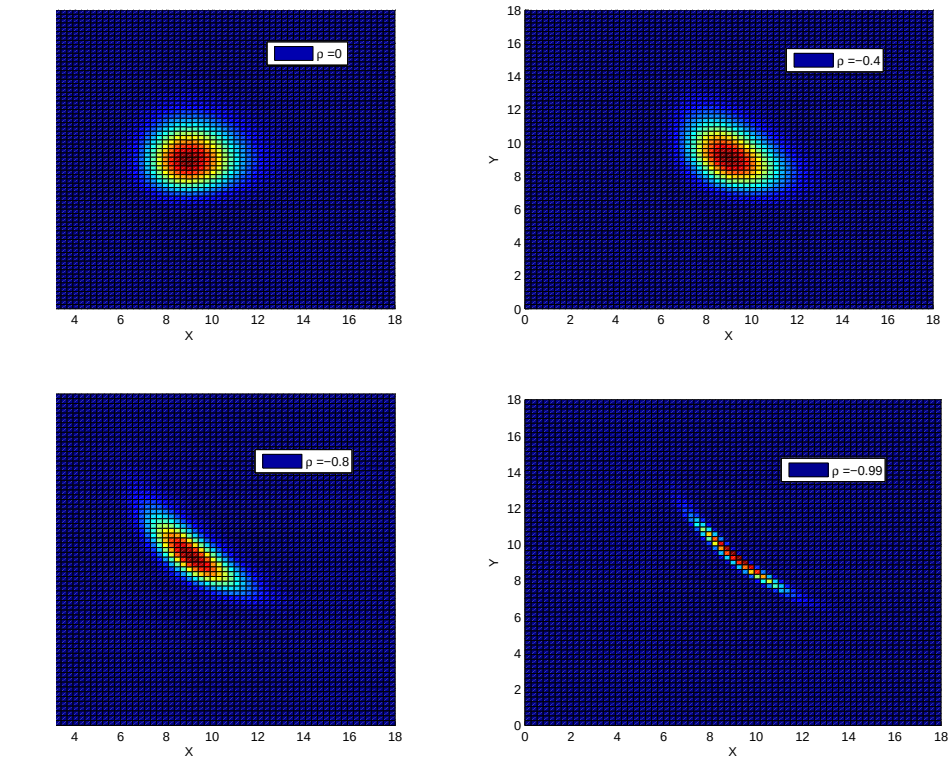


Figure 2.8: Plots of the transition density of two correlated geometric Brownian motions with drift when the correlation changes in value.

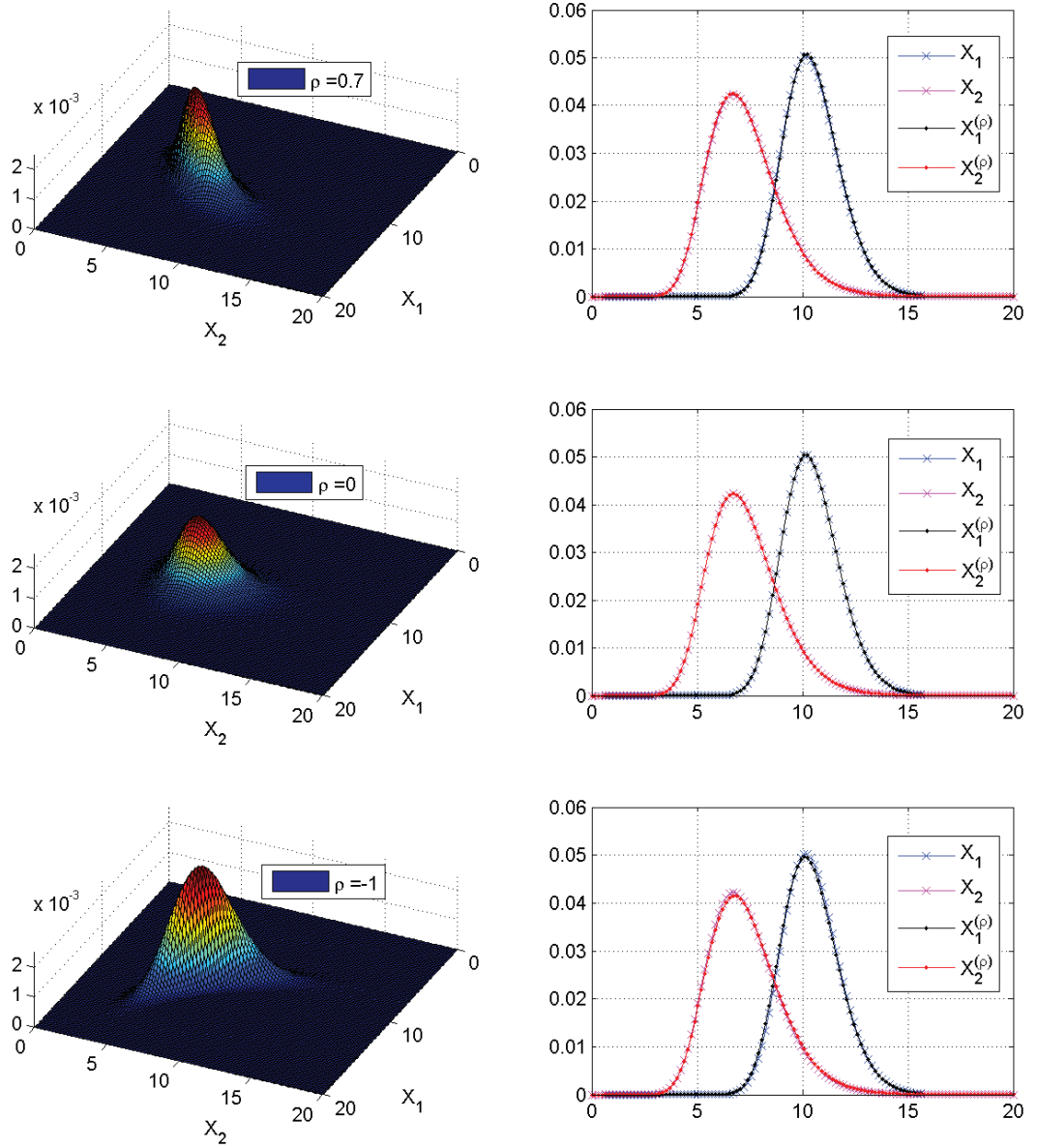


Figure 2.9: Plots of the transition density of two correlated geometric Brownian motions with drift when the correlation changes in value. On the left side the bidimensional transition density is reported. On the right side the marginal distributions. The parameters used are: $\mu_1 = .01, \mu_2 = .015, \sigma_1 = .13, \sigma_2 = .23$. The number of discretization points is 150 and the subspace dimension used for the exponentiation is 5 with tolerance 10^{-7} .

Chapter 3

Credit Migration Risk and Point-in-Time Credit Dynamics

This Chapter presents a credit migration model that aims to consistently capture the point-in-time dynamics of the credit worthiness of debt issuers and their obligations, and a calibration routine that permits the model to effectively fit historical ratings data. Our approach is to view the rating migration matrices as the operator semigroup associated to an approximated parametric infinitesimal generators. Our credit model accounts not only for default risk dynamics but also for the entire transition among states of the rating migration matrix. This modeling feature is fundamental for an efficient risk management of credit derivatives and credit risk portfolios conditionally on a state of the economy or specific macro factors. We fit our model to the historical average rating migration matrices published by Moody's Investors Service, focusing on the banking sector over the period 1920-2005. Our results show that the model can identify the through-the-cycle transition across rating scales and that the point-in-time migration probabilities are only generated by stressed economic conditions and can only be justified by the influence of macro factors on the through-the-cycle unconditional probability values. The great level of modeling details and the accuracy of the produced results is an improvement over those of other available models.

3.1 Literature reviews

In this Chapter we introduce and focus on a new model for rating migration dynamics, migration risk and point-in-time credit quality dynamics, a problem not new in finance but that urges a to be taken into a more relevant consideration by both banks and financial institutions, after the credit crisis and the new Basel III global regulatory standards on bank capital adequacy and liquidity.

The credit rating, either assigned by a major rating agencies or defined within a company credit worthiness internal system, gives a through-the-cycle estimation of the credit quality of a corporation, a country or an individual. The through-the-cycle default probability modeling approach is defined as a long run assessment of the risk associated with a borrower which does not change with the economic cycle, therefore the borrower credit grade will not change due to economic conditions. On the contrary the point-in-time default probability modeling approach is an assessment of borrower risk at that particular point-in-time and the borrower will move up or down credit grades in line with the

economic cycle. Typically, a credit rating tells a lender or investor the probability of the subject being able to pay back a loan. Credit ratings have also been used to adjust insurance premiums, determine employment eligibility, establish the amount of a utility or leasing deposit and calculate a borrower funding costs. However the credit quality of most issuers and their obligations is not fixed and steady over a period of time, but tends to undergo a change. For this reason, changes in ratings occur so as to reflect variations in the intrinsic relative position of issuer and their obligations.

Rating changes are not isolate and rare events in the corporate world and rating triggers seem to regularly affect debt issuers, and as reported in Stumpp and Coppola [2002], they can be of many types, but the most prevalent are provisions which require the issuer of debt to post some sort of extra collateral as a security enhancement if ratings deteriorate or the payment on a swap contract. In fact, as pointed out in Andersson and Vanini [2009], banks not only face loans and debts which are affected by default of their clients but their earnings are also affected by for example annual down- or up-gradings of their client creditworthiness. Such migration of the clients' ratings create losses, opportunity losses or non realizable gains depending on the specific loan contract. Furthermore a rating downgrade has an important consequence in the worsening of the debt issuer funding costs. The downgrade impact a company negatively, thus pushing the company further into distress. These aspects need to be all taken into account to comply with the new rules on capital requirement.

In our approach to credit rating analysis, we don't model directly the rating migration dynamics as in Lando [2004], Schönbucher [2003], Hull and White [2000], but instead a proxy of it, namely the distance to default (DTD) process. This concept was introduced in early works Albanese and Chen [2003], Albanese et al. [2005a] and Dalessandro and Gilli [2007]. We privileged this approach because in structured finance a key role is played by the rating, an ordered system of credit quality levels, and its variation over time that reflect its instantaneous creditworthiness. However, the mathematical framework developed in this Chapter, although similar, differs substantially from Albanese et al. [2005a]. The model dynamic accommodates now for the positive serial correlation in rating changes, known as rating drift in the literature, see Fligewski et al. [2006]. Furthermore, while in Albanese et al. [2005a] the conditional default probabilities were computed by means of an unobservable market factor modeled as a non recombining tree, our approach consists in calculating conditional point-in-time migration probabilities by an explicit specification of the joint density function between two variables, the rating quality and the market/credit cycle, both modeled as continuous time Markov chains (CTMC). Empirical evidence and our calibration results show that the assumptions we made on the conditional dynamics and model setup have a significant economic meaning.

In fact we directly exploit the strong correlation between the rating risk and the business cycle, well documented in, cf. 4.2 Duffie and Singleton [2003], where the authors observe consistent large negative correlation between default rates and GDP growth.

In this way, given a starting credit status, it would be possible to simulate the credit quality deterioration conditional to the market variable and verify whether the current credit risk is underestimated or overestimated. In our framework the rating grades are not limited to the standard number but we model a richer set of rating levels that are useful to estimate the intrinsic creditworthiness within a specific credit category and account for rating outlook levels either positive or negative, as modeled in Lando and Mortensen

[2005]. This is an invaluable tool for relative risk analysis in credit trading. In our model, the DTD contains the survival information or the downgrade frequencies of the defaultable security. In the second part of this Chapter we will mainly focus on empirical analysis of historical migration data published in the form of average transition matrix by the major rating agencies and therefore consider extensively the physical measure $\mathbb{P}$, but all the study can be extended to the $\mathbb{Q}$ pricing measure as in Dalessandro [2010c].

An important contribution of our work is also the formulation of a dedicated calibration procedure that allows us to both identify the model and efficiently fit both the through-the-cycle rating migration matrices and the correlation dynamics between the market factor and rating movements. Our model is flexible and applicable also to data generated by proprietary rating systems, or corporate bond prices or credit default swaps prices based on ratings and can be conditioned to some market scenario to obtain a more accurate assessment of the point-in-time creditworthiness. We will show how our model is able to quantify the impact of the economic cycle on rated credit assets and the migration risk.

The findings within this Chapter have their natural collocation, within the existing financial literature, among the popular and widespread research works applied to the problem of finding a representation for a stochastic matrix in the context of credit migration such as Schuermann and Jafry [1997a], Jarrow et al. [1997], Israel et al. [2001] and Andersson and Vanini [2009], where a continuous time Markov framework is usually assumed for the rating process.

There exists a research area on modeling the effect of macroeconomic factors on corporate defaults and credit rating transitions, and important contributions have been given by the works of Fligewski et al. [2006] Koopman et al. [2005] and Duffie et al. [2007].

The Chapter is organized as following: In Section 2 we introduce the model building blocks and the concept of distance to default. In Section 3 we propose a new way to condition the credit process to a market factor and characterize the conditional model dynamics. Section 4 illustrates the calibration to historical rating migration matrices and to market data and results are discussed.

3.1.1 Model Uniqueness and its Applications

The aim of the proposed framework is to allow practitioners and investors to better model, calibrate and understand the impact of market cyclicity on rating migration. The assessment of the point-in-time credit risk, through an intuitive and accurate modeling framework, is a fundamental tool to allocate diversified risk and for hedging purposes.

In fact the main feature that makes our model different and unique within the existing available literature, is the way conditional default probabilities are computed with respect to a macro variable, and therefore the ability to obtain intuitive parametric point-in-time credit curves that can be easily calibrated to the volatile credit market data (see fig. 3.9). Another important feature is that we model an extended credit scale, and not just the standard rating scale of 21 notches as defined by the major rating agencies. In our model each main rating class is characterized by a number of further 10 or more subclasses, and this allows to better capture the credit quality of the debts and their rating outlooks. The dense rating scale gives an accurate assessment of the distance to default of the rated debts. Some authors tried to incorporate this feature in their models and, among others, it is important to mention the works of Lando and Mortensen [2005] and Metz [2007].

The parametric rating migration model we propose is able to fit market data through a new spectral calibration routine that guarantees the identifiability of the model itself. Financial institutions will benefit of our framework and findings within this Chapter for better quantify their capital requirement and corresponding liquidity facilities in order to comply with the new rules from Basel III. The calibrated model can be used as a tool to forecast cumulative rating migration probabilities and define rating outlooks.

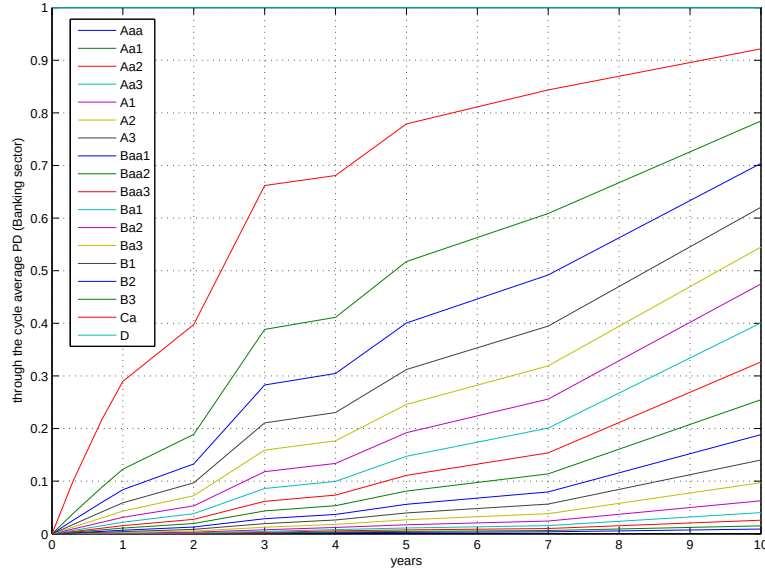


Figure 3.1: Fitted average through-the-cycle default probabilities across rating categories for the Banking sector. The model is able produce meaningful and non-intersecting through-the-cycle credit curves across all the rating scale. Note that the default ‘D’ curve is at the top of the plot area, while lower rating categories are in the bottom part of the plot area.

3.1.2 Rating Migration Matrices and Data source Description

Definition 3.1 : *(Rating) Moody's long term ratings are intended to be measures of expected loss and therefore incorporate elements of both probability of default and severity of loss in the event of default. The system of rating securities was originated by John Moody in 1909. The purpose of Moody's ratings is to provide investors with a simple system of gradation by which relative creditworthiness of securities may be noted.*

Definition 3.2 *Rating Symbols: Gradations of creditworthiness are indicated by rating symbols, with each symbol representing a group in which the credit characteristics are broadly the same. There are nine symbols, from that used to designate least credit risk to that denoting greatest credit risk:*

$$Aaa, Aa, A \quad Baa, Ba, B \quad Caa, Ca, C$$

Moody's appends numerical modifiers 1, 2 and 3 to each generic rating classification from Aa through Caa.

We consider in our analysis, without loss of generality, the rating scale provided by one of the major rating agencies, that includes 18 levels of creditworthiness.

$$\{Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3, B1, B2, B3, Ca, D\}$$

where *Aaa* represents the best possible credit quality, *Ca* the worst non defaulted credit quality, and *D* represents the state of default.

The migration data, used for our empirical analysis, consists in a term structures of average-years rating migration matrices for the years $P = \{1, 2, 3, 4, 5, 7, 10\}$ using historical Moody's 12-sectors data from 1920 until 2005. The data are reported in Notes 4.4.1 at the end of the Chapter. Focusing mainly on the Banking sector dataset, we notice that the entries of the average migration matrices don't represent neither the through-the-cycle dynamic nor the point-in-time probabilities in a clear way. In fact, as clearly represented in fig. 3.2, the rating default probabilities appear very noisy, far from a through-the-cycle default representation of the Banking sector, but almost like conditional point-in-time default frequencies.

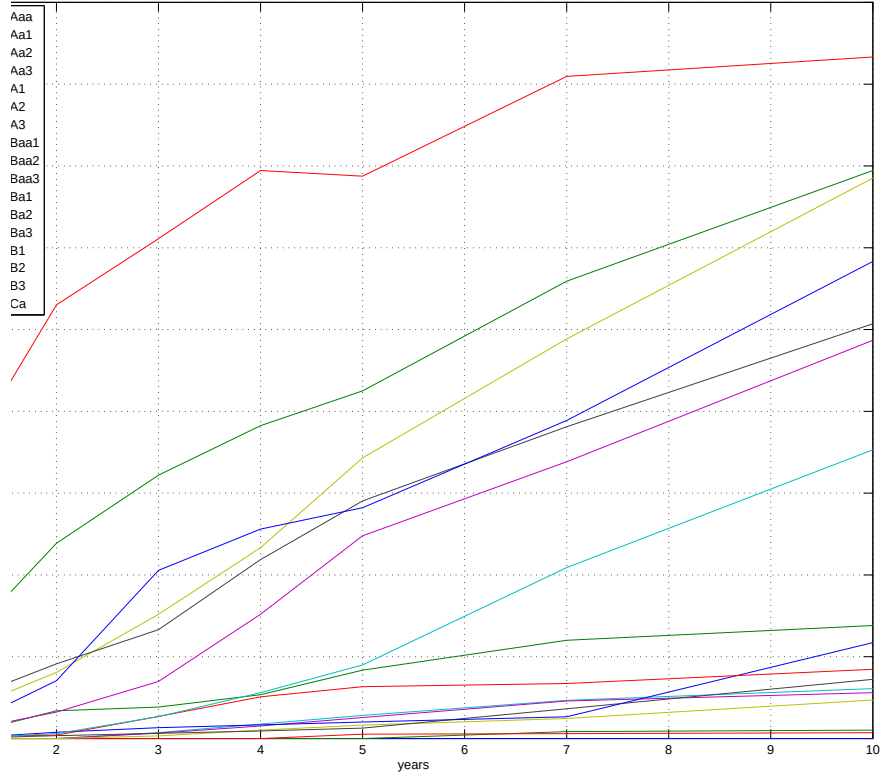


Figure 3.2: Historical average through-the-cycle default probabilities across rating categories for the Banking sector. The curves lack of economic intuition and in some cases there are evident conditions that can lead to arbitrage if those curves are used as input to credit pricing models.

3.2 Model building blocks and Distance to Default Process

In this section we present the main concepts and building blocks of our rating migration model. It is instructive to start with a description of the rating scale and rating matrix to make clear the properties of the objects we are going to model.

The above rating scale is mapped in our model, into an equivalent, more dense numerical scale, i.e.

$$\{Aaa, Aa, A, Baa, \dots, D\} \rightarrow \mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n] \quad (3.1)$$

with $n > 17$, and with typical value of 200, that will represent the states of our Markov chain. Every main rating level, except default, has a subset of 12 subclasses, or more. In this way we have available a richer set of rating levels that are going to be very useful when calibrating our model parameters to observed debt instruments and migration frequencies. It is very common in the literature, see Lando [2004], Schönbucher [2003], to assume that the rating transition is a time-invariant Markov process, and that the entries in the average transition matrix, obtained perhaps from a rating agency, describe the Markov

transition probabilities. Using the Markov assumption is equivalent to saying that the current rating is all that is necessary to predict the future ratings, and that historical data is irrelevant for predicting the future. The time-invariance assumption implies that the transition probabilities do not change with time and are constant over the entire horizon of interest.

More formally, given n rating categories and the rating migration process of a firm for the period $[0, t], t \geq 0$ is a continuous time homogeneous Markov chain (CTMC) with transition matrix $P(0, t) = \{p_{ij}(0, t)\}_{i,j=1,\dots,n}$, in which the ij -th entry represents the probability of migrating from category i to category j during the period $(0, t)$. Similar to the discrete-time Markov process for which the rating transition matrix can be obtained by matrix multiplication from the one-period transition matrix, the matrix $P(0, t)$ can be represented, under the assumption of time homogeneity, through its generator matrix Λ , that is, for any time $t > 0$,

$$P(0, t) = e^{\Lambda t} := \sum_{k=0}^{\infty} \frac{\Lambda^k t^k}{k!} \quad (3.2)$$

in which $\Lambda = \{\lambda_{ij}\}$ satisfies $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$ for $1 \leq i \leq n$, and $\lambda_{ij} \geq 0$ for $1 \leq i \neq j \leq n$. The elements in Λ can be estimated by their maximum likelihood estimators $\hat{\lambda}_{ij} = \frac{N_{ij}}{\int_0^t Y_i(s) ds}$, in which $Y_i(s)$ is the number of firms in rating class i at time s and N_{ij} is the total number of transitions from i to j ($j \neq i$) over the period $[0, t]$.

However the Markov framework has some limitations. In fact the data suggest, Wilson [1997a,b], that the rating transition depends not only on industry category and macroeconomic conditions, but also on the financial trend of the particular company as reflected in its recent rating changes. In other words, the future ratings of a company depend not only on its current rating, but also on changes in its rating in the recent past.

Therefore it would more appropriate to assume that the Markov chain representing the rating migration is time-inhomogeneous and also to relax the Markov property in order to account for the empirical phenomenon of ratings momentum.

Our model consists in specifying a process for the credit quality, that we name the distance to default (DTD) process, a variable not directly observable in the market, but fully defined within the credit information carried by all credit derivatives. We confine such a process on the finite credit quality space S_X , that are the states of the CTMC. We discard the standard intensity based approaches to build CTMC from rating information as the ones used in the works Schuermann and Jafry [1997a], Jarrow et al. [1997], Israel et al. [2001], Lando [2004], Schönbucher [2003], Hull and White [2000] based on MLE of Poisson processes or logit regression of default frequencies, and we privilege the use functional calculus. The setup for the DTD is similar to Albanese et al. [2005a] and it is as follows. Let $(\Omega \subseteq S, \mathcal{F}_t, \mathbb{P})$ be a probability space where is defined the $\mathbb{R}_+$ valued Feller process of the DTD, $X = \{X_t : 0 \leq t \leq T\}$ which is right continuous with left limits and N_t a standard Poisson jump process which is independent of X . Let $\{\lambda_X(s) : s \geq 0\}$ be an intensity process that we assume inhomogeneous and non-negative and that is associated to the chosen DTD dynamics. We define the default time τ as follows:

$$\tau = \inf\{t \in \mathbb{R}_+ : \int_0^t \lambda_X(s) ds \geq N_1\} \quad (3.3)$$

where N_1 is the first jump time of the poisson process. The survival probability expression

is derived using the definition of a Cox process and is as following:

$$\begin{aligned}\mathbb{P}(\tau > t | \mathcal{F}_s) &= \mathbb{E}_s^{\mathbb{P}}[\mathbb{P}(\tau > t | \mathcal{F}_s \cup \{\lambda_X(t) : t \geq 0\})] \\ &= \mathbf{1}_{\{\tau > s\}} \mathbb{E}_s^{\mathbb{P}}\left[\exp\left(-\int_s^t \lambda_X(u) du\right)\right], \quad s \in [0, t].\end{aligned}\quad (3.4)$$

We regard the rating scale commonly adopted by rating agencies to express a judgement of the quality of a certain debt as an ordered set, where the two extremes for the credit-worthiness interval are respectively the risk free rate and default, both absorbing states. The model we propose consists in expressing the underlying DTD dynamics, associated to the credit risky intensity λ_X , by a unique continuous time-inhomogeneous Markov chain with a parametric approximated infinitesimal generator. The intensity function $\lambda_X(s)$ ¹ previously introduced, will be locally modeled by a parametric infinitesimal generator matrix $\mathcal{L}(s) := \mathcal{L}_X(s)$, and the time to default is modeled as the absorption time in default state of our CTMC. Therefore the default probability p_t^d at time $t > 0$ is represented by the default transition probability entry of the matrix $P_X(t)$, given a starting status p_0 , such that $P_X(t)$ satisfies the Kolmogorov's forward equation

$$\frac{\partial P_X(t)}{\partial t} = \mathcal{L}(t)P_X(t) \quad (3.5)$$

Our analysis consists in the derivation of a CTMC that approximates the dynamic of the DTD jump process. The approach is to view the rating migration matrices as the operator semigroup P_X associated to a parametric infinitesimal generators. We choose the DTD dynamic to be of the form

$$dX_t = \kappa(X_t)(\theta(X_t) - X_t)dt + \sigma(X_t)dW_t + \text{'Jumps'} \quad (3.6)$$

or,

$$dX_t = \mu(X_t)X_t dt + \sigma(X_t)X_t dW_t + \text{'Jumps'} \quad (3.7)$$

where by 'Jumps' we mean a jump component specified from the large class of Lévy processes. In fact we need a DTD process that can accommodate for a local mean reverting behavior, local drift and also to exhibit jumps in its paths, in order to have sufficient flexibility to calibrate the historical rating migration data. In the following subsections we concentrate our analysis on the representation of the local approximated infinitesimal generator $\mathcal{L}(t)$ for jump processes, in particular the compound Poisson and variance gamma processes.

Markov chain states specification for the Rating Migration problem

In the section above, we defined the state space $\mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n]$ for our Markov chain, where the number of states n is generally much larger than the $K = 18$ standard rating classes number. This means that the transition kernel P_X solution of the eq.(3.5) is a stochastic matrix of dimension $n \times n$, larger than a standard rating migration matrix $M \in \mathbb{R}^{18 \times 18}$ published by the rating agencies. In order to produce with our model a matrix that can be directly compare with the stard rating migration matrices, we subdivide the space $\mathcal{S}_X$ into K subintervals of adjacent states

$$I_k = [x_{k-1}, \dots, x_k], \quad k = 1, \dots, K, \quad \text{where } \underline{x} = x_0 =, x_1, x_2, \dots, x_K = x_n$$

¹The time s dependency is due to the fact that we are dealing with a time inhomogeneous Markov chain.

each of them corresponding to the k -th rating class, with $\sharp(x_k - x_{k-1}) = \frac{n}{K}$, for $k = 1, \dots, K$, and x_K the state of default.

Let's denote by $\bar{x}_k \in I_k$, the starting state, where I_k is the k -th rating class. The conditional transition probability $\hat{p}_{ij}(t)$ that an obligor with a given initial rating i at time 0 will have a rating j at a later time $t > 0$ can be calculated as follows:

$$\hat{p}_{ij}(t) = \sum_{z=k_j-1}^{k_j-1} P_X(\bar{x}_i, z), \quad j = 1, \dots, K.$$

where $\bar{x}_k$ is a point in the interval I_k which represent the entire k -th rating class and is part of the model specification. For simplicity $\bar{x}_k$ can be chosen as the midpoint of the interval I_k . In this way it will be possible to define through our model a parametric rating matrix, with parameters' set θ , that can directly be compared to one published by an agency.

$$L_t(\theta) = \{\hat{p}_{ij}(t)\} \quad i, j = 1, \dots, K \quad (3.8)$$

We detail the construction of a parametric CTMC for the DTD process in the following sections.

3.2.1 DTD as a process with jumps

Infinitesimal generators of jump processes and more generally Lévy processes are commonly categorized as pseudo differential operators (PDO). Please refer to the Chapter 2 for the necessary background and introduction on PDO.

If $\{X_t\}_{t \geq 0}$ is a Lévy process with characteristic exponent $\Psi(s)$, then its transition operator is a pseudo-differential operator with symbol $\phi(s, t) = e^{t\Psi(s)}$ and its generator is a pseudo-differential operator with symbol $\Psi(s)$.

In probability we can derive such symbols as following

$$\begin{aligned} \phi(s, t) &= \mathbb{E}[e^{isX(t)} | X(0) = x(0)] \\ \Psi(s) &= \left. \frac{\partial}{\partial t} \phi(s, t) \right|_{t=0} = \lim_{t \downarrow 0} \frac{(\mathbb{E}[e^{isX(t)} | X(0) = x(0)] - 1)}{t} \end{aligned}$$

where the symbol $\Psi(s)$ in its most general form is given by the Lévy-Khintchine formula of eq.(??).

$$\Psi(s) = is\mu + \frac{1}{2}s\sigma^2 + \int_{\mathbb{R}} (e^{isx} - 1 - ish(x))\nu(dx). \quad (3.9)$$

We will focus our attention on two processes with jumps, namely the compound poisson and the variance gamma process, and we want to characterize them in term of their approximated infinitesimal generator. We will then exploit the PDO representation of their infinitesimal generator in order to numerically prove some desirable weak convergence results of the approximated generators.

Diffusions

Let's denote by $\mathcal{A}$ the approximated generator associated to the diffusive part of DTD process in eq.(3.6) on the discretized state space S_X .

$$\mathcal{A}(x, y) = \{a(x_i, x_j)\}, \quad i, j = 1, \dots, n \quad (3.10)$$

is a tridiagonal matrix specified by a state dependent drift $\mu(x)$ and volatility $\sigma(x)$. Each matrix entry can be calculated by solving the following system:

$$\begin{cases} a(x_1, x_2) = a_t(x_n, x_{n-1}) = 0 \\ a(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{\mu(x_i)}{h} + \frac{\sigma(x_i)^2}{h^2} \right) \\ a(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma(x_i)^2}{h^2} - \frac{\mu(x_i)}{h} \right) \\ a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})) \end{cases}$$

for all $i = 2, \dots, n-1$, with $h = x_i - x_{i-1}$, and $\frac{\sigma(x_i)^2}{h} \leq \mu(x_i) \leq \frac{\sigma(x_i)^2}{h}$.

Compound Poisson

The first possible way to add jumps to the DTD process in eq.(3.6) is by means of a compound poisson process. We denote with $\{X(t)\}_{t \geq 0}$ a compound poisson process with Lévy measure ν , taking values in the discrete space S .

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

where $N(t)$ is a Poisson random variable with parameter λt and it gives the total number of jumps on the interval $[0, t]$.

The associated infinitesimal generator is

$$\begin{aligned} \mathcal{A}f(x) &:= \lambda(\mathcal{L} - I)f(x) \\ &= \lambda \int (f(x+y) - f(x)) \nu(dy) \end{aligned} \quad (3.11)$$

Given the discrete space of the credit quality $S_X := \{x_i\}$, $i = 1, \dots, n$ we can derived and algebraic scheme to approximate the generator in eq.(3.11) and build the matrix $\mathcal{A}_{cp} = \{a(i, j)\}$, $i, j = 1, \dots, n$ that approximates the infinitesimal generator in eq.(3.11).

The entries of $\mathcal{A}_{cp}$ can be calculated as

$$a(i, j) = \begin{cases} \lambda[\Phi(u, \kappa, \delta) - \Phi(l, \kappa, \delta)], & \text{if } i \neq j \\ -\sum_j a(i, j) & \text{if } i = j \end{cases}$$

and assuming the following Lévy measure,

$$\begin{aligned} \nu(dx) &= \frac{1}{\delta\sqrt{2\pi}} \exp \left\{ -\frac{(dx - \kappa)^2}{2\delta^2} \right\} & p(dx) \sim N(\kappa, \delta^2) \\ \Phi(u, \kappa, \delta) &= \int_{-\infty}^u \nu(dx) \end{aligned}$$

denoting with $N(\kappa, \delta^2)$ the normal distribution with mean κ and variance δ^2 and for $i = 1, \dots, n$,

$$\begin{aligned} u &= \begin{cases} \frac{x_j + x_{j+1}}{2} - x_i, & \text{if } j \neq n \\ x_j - x_i & \text{if } j = n \end{cases} \\ l &= \begin{cases} x_j - x_i & \text{if } j = 1 \\ \frac{x_j + x_{j-1}}{2} - x_j, & \text{if } j \neq 1 \end{cases} \end{aligned}$$

The associated transition probability matrix, can be computed through matrix exponentiation

$$p_t^{(n)} = \exp(t\mathcal{A}_c)p_0, \quad p_0 \text{ starting condition} \quad (3.12)$$

The number of parameters that specify the infinitesimal generator matrix for the compound poisson process is three and are

$$\Theta_{cp} = \{\lambda, \kappa, \delta\} \quad (3.13)$$

It is also possible to specify these parameters as ‘local’, and therefore they will be function of the states of the chain, namely $\lambda(x_i), \kappa(x_i), \delta(x_i)$, for all i . In particular if we want to locally characterize the behaviour of each rating category we would need to specify $K = 18$ local parameters, so that the total number of parameters for the local compound Poisson process would be 54.

The approximated infinitesimal generator for the DTD in eq. (3.6) is

$$\mathcal{A} + \mathcal{A}_{cp}$$

that is the sum of the infinitesimal generator for the diffusive part and the jump component respectively. In the following we verify the convergence of the compound poisson kernel p_t numerically by studying the error term

$$\epsilon_n = \left| p_t^{(n)} - F^{-1}(\overline{T_t \cdot f})(s) \right| \quad (3.14)$$

defined as the distance between the transition kernel density in eq.(3.12) and the transition probability derived from the FFT inversion

$$F^{-1}(\overline{T_t \cdot f})(s) = (T_t \cdot f)(x) = \frac{1}{\sqrt{2\pi}} \int e^{-isx} \bar{f}(s) e^{t\Psi(s)} ds. \quad (3.15)$$

as $n \rightarrow \infty$. The profile of converge error ϵ_n is reported in fig. 3.4 and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. For the derivation of the characteristic function of the Brownian motion with drift and compound poisson process, see appendix A.1.

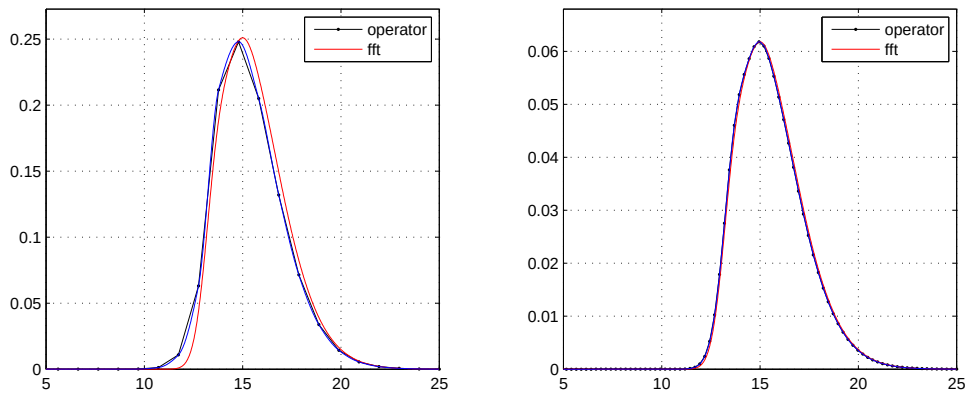


Figure 3.3: Comparison between the Brownian motion with drift and compound poisson transition kernel generated by the operator and the one by inverting the FFT. The inverse FFT is computed over 4096 points, while the operator is computed on 50 states in the left plot and on 200 states on the right one.

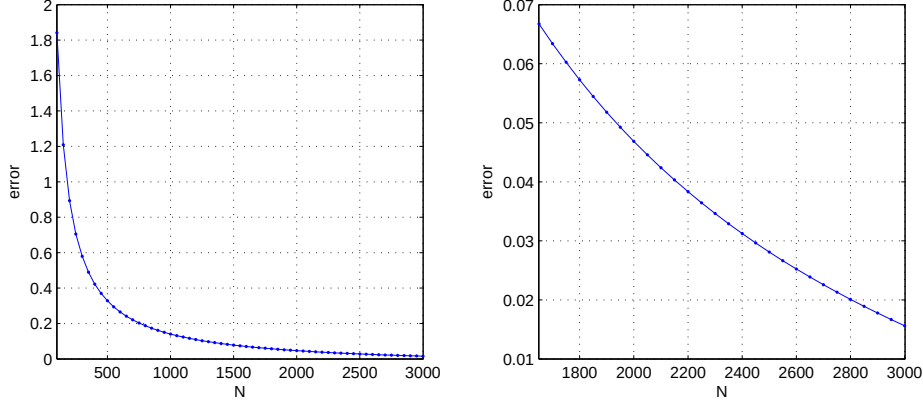


Figure 3.4: Convergence error between the density obtained via FFT inversion over 4096 point and the operator computed over N states.

Variance Gamma

Another possible alternative to add jumps to the DTD process of eq.(3.6) is through subordination using a Gamma process. The resulting DTD process is called variance Gamma (VG). The advantage of using a VG process rather than a compound poisson is that it can require a reduced number of local parameters. Therefore the user of our model can use a DTD process with different fat tailed distribution type according to the quality and nature of the historical migration matrices that need to be calibrated. For further details on the VG process, refer to appendix A.2.

The approximated generator of a VG process is denoted as $\mathcal{A}_{\text{VG}}^{(n)} \in \mathbb{R}^{n \times n}$, and can be computed through the time subordination of a diffusive process, as described in the seminal work of Madan et al. [1998b]. If $\mathcal{A}$ is a diagonalizable matrix representing the approximated infinitesimal generator of the diffusive process as in eq.(3.10), the following equation hold:

$$\mathcal{A} = V D V^{-1}$$

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & & \\ 0 & \lambda_2 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & \lambda_{n-1} & 0 \\ & & 0 & 0 & \lambda_n \end{bmatrix}$$

We then compute the Laplace transform of the Gamma subordinator $\Gamma(x; \frac{t}{\nu}, \nu)$, and apply the resulting map to the spectrum of the original diffusive process to obtain the VG generator.

$$\begin{aligned} \mathcal{L}\{\Gamma(x; \frac{t}{\nu}, \nu)\} &= \int_0^\infty e^{-sx} \Gamma(x; \frac{t}{\nu}, \nu) dx = (1 + \nu)^{-\frac{t}{\nu}} = e^{-\frac{t}{\nu} \log(1+\nu s)} = e^{\Psi(s)} \\ \bar{D} &= \Psi(D) \\ \mathcal{A}_{\text{VG}}^{(n)} &= V \bar{D} V^{-1} \end{aligned}$$

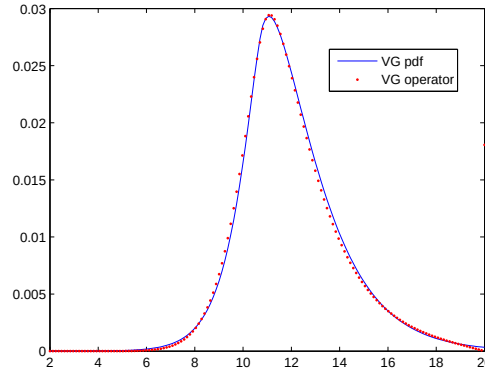


Figure 3.5: Comparison between the two transition kernels of the VG process obtained by inverting the characteristic function over $n = 4096$ points and through the exponentiation of the corresponding approximated infinitesimal generator with $n = 200$ states.

In the above specification of the gamma distribution $\Gamma\left(x; \frac{\hat{\mu}^2}{\nu}, \frac{\nu}{\hat{\mu}}\right)$, the mean parameter is $\hat{\mu} = 1$. The number of parameters that specify the infinitesimal generator matrix for the variance gamma process is four and are

$$\Theta_{vg} = \{\mu, \sigma, \hat{\mu}, \nu\} \quad (3.16)$$

where μ and σ are respectively the constant specifications for the drift and volatility of the underlying diffusion. In the case of a local characterization of the behaviour of all $K = 18$ rating categories we can specify a local drift $\mu(x)$ and a local volatility $\sigma(x)$, and in such a case the total number of parameters would be $36 + 2$. We repeat a similar analysis as in eq.(3.14) and eq.(3.15), for the approximated VG infinitesimal generator, and fig.(3.5) reports a qualitative comparison of the VG transition kernels obtained.

DTD with asymmetric dynamics

In order to have more flexibility in calibrating both transition probabilities and implied default probabilities we can specify the DTD process to be asymmetric. We suppose that the DTD process at any time t may follow two transition function, namely an upward and downward transition, related to two different parameters' specifications. In order to define this specific model we first build two infinitesimal generators, $\mathcal{A}_{\text{DOWN}}^{(n)} = \{a_{\text{DOWN}}(i, j)\}_{i,j=1,\dots,n}$ for the downward transitions and $\mathcal{A}_{\text{UP}}^{(n)} = \{a_{\text{UP}}(i, j)\}_{i,j=1,\dots,n}$ for the upward transitions respectively. The resulting approximated generator matrix $\mathcal{A}^{(n)} = \{a(i, j)\}_{i,j=1,\dots,n}$ can be calculated as,

$$\begin{aligned} a(i, j) &= \begin{cases} a_{\text{DOWN}}(i, j), & i > j \\ a_{\text{UP}}(i, j), & i < j \end{cases} \\ a(i, i) &= -\sum_{j \geq i} a_{\text{UP}}(i, j) - \sum_{j \leq i} a_{\text{DOWN}}(i, j) \end{aligned}$$

In this way it is possible to build the approximated infinitesimal generator for the asymmetric variance gamma process, see Madan et al. [1998b]. This generator can be build to

accommodate the local feature of each rating category and in such a case the total number of parameters would be 40, divided into 18 for the local drift $\mu(x)$, 18 for the local volatility $\sigma(x)$, 2 for the upward jump component ν_u and $\hat{\mu}_u$, 2 for the downward jump component ν_d and $\hat{\mu}_d$.

3.3 Point-in-time Model Dynamics

In this section we explain how we specify the joint dynamic probability density among the rating processes and be able to produce point-in-time rating migration matrices. Consistently with the explanation given in the introduction we assume that the DTD marginal is joint to the marginal of the market factor. When the DTD process will be independent from the market factor process the rating dynamics implied by the DTD will represents a thought-the-cycle rating dynamics. In case of non zero correlation between the market factor and the DTD process, the conditional DTD process will represents a point-in-time rating dynamics.

We assume that the market factor process is the same as the DTD and therefore these two variables will be modeled as identical Markov chains and therefore by just one CTMC.

Let $X^{(z)} = \{X_t^{(z)}\}_{t \geq 0}$, $z = 1, 2$ be two identical continuous time Markov chains² with discrete state space $\mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n]$ and built as illustrated in the previous section or cf. Dalessandro [2010a], with n a positive integer number representing the number of states. We consider the probability transition matrices,

$$\begin{aligned} P_t^{(z)} = P(X_t^{(z)}) &:= \{p_t^{(z)}(x_i, x_j)\} \\ &= \{p_t^{(z), x_i}(x_j)\} \quad \text{For } i, j = 1, \dots, n, \quad z = \{1, 2\}, \quad t \geq 0. \end{aligned}$$

for the chains $X^{(n)}$ and we would like to define a joint distribution that correlates the dynamic of the marginals $p_t^{(z), x_i}$ for all i . Let $F_t^{(z)} := \{F_t(x_i, x_j)\}_{i, j=1, \dots, n}$ the corresponding distribution function matrices.

We would like to generate an arbitrary partition of the probability mass of each transition matrix row, and define its corresponding partitioned state space. Given a row $p_t^{(z), x_i}$ of the matrix $P_t^{(z)}$, we have that $\sum p_t^{(z), x_i} = 1$. We then subdivide the entire probability mass into a sequence of probability masses $q_1, q_2, \dots, q_m$, such that $\sum_{j=1}^m q_j = 1$, and define a corresponding increasing sequence of cumulative probability masses $0 \leq c_1, c_2, \dots, c_m \leq 1$, where

$$c_k = \sum_{j=1}^k q_j, \quad k = 1, \dots, m$$

and the sequence of corresponding percentiles

$$b_k = F_t^{-1(z)}(c_k), \quad k = 1, \dots, m. \quad (3.17)$$

²Please note that where the index (z) is not specified, the value under consideration is assumed to be the same.

This means that the support of each probability mass will be as following

$$\begin{aligned} q_1 &\in s_1 = [\underline{x} = x_1, b_1], \\ q_2 &\in s_2 = [b_1, b_2], \\ &\vdots \\ q_m &\in s_m = [b_{m-1}, b_m = \bar{x} = x_n] \end{aligned}$$

Note that the value of every local probability mass $q_k(x_j)$ is fully defined by $p_t^{(z),x_i}(x_j)$, such that

$$q_{k(z)}(x_j) = p_t^{(z),x_i}(x_j) \quad \text{for all } x_j \in s_{k(z)}$$

We define as $\beta(x) : S_x \mapsto [0, 1]$ the local correlation function.

Definition 3.3 *Given the marginals matrices $P_t^{(z)}$, and their corresponding sequences of probability masses $(q_1^{(z)}, \dots, q_m^{(z)}) = (q_1, \dots, q_m)$, we can define a joint local density for all row x_i , with $\bigcup_{k=1}^m s_k = S_X$, with local marginals $p_t^{(z),x_i}$ by the following set of equations*

$$J_t^{x_i}(x_j^{(1)}, x_j^{(2)}) = \begin{cases} \beta(x_i)q_{k(1)}(x_j^{(1)}) + (1 - \beta(x_i))q_{k(1)}(x_j^{(1)})q_{k(2)}(x_j^{(2)}) & \text{if } k^{(1)} = k^{(2)} \\ (1 - \beta(x_i))q_{k(1)}(x_j^{(1)})q_{k(2)}(x_j^{(2)}) & \text{if } k^{(1)} \neq k^{(2)} \\ \sum_k q_{k(z)} = 1 & z = 1, 2 \end{cases} \quad (3.18)$$

that generates a local correlation equal to

$$\begin{aligned} \rho(x_i) &= \frac{\mathbb{E}_{x_i}[X_t^{(1)} X_t^{(2)}] - (\mu(x_i)t + x_i)}{\sigma^2(x_i)t} \\ &= \frac{\sum_{k(1)}^m \sum_{k(2)}^m s_{k(1)} s_{k(2)} J_t^{x_i}(x_j^{(1)}, x_j^{(2)}) - (\mu(x_i)t + x_i)}{\sigma^2(x_i)t} \end{aligned}$$

with $\mu(x)$ and $\sigma(x)$ the local specifications of the drift and volatility function respectively for the chain.

We provide an illustrative example for the above definition 4.1.

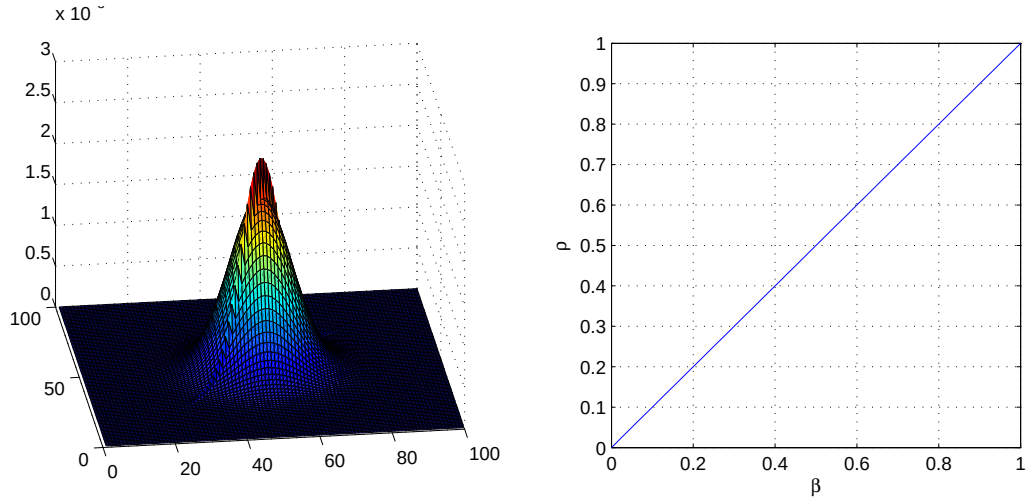
Example 1

We fix a row $x_i = x_i^{(1)} = x_i^{(2)}$ of the two identical approximated transition probability matrices, and we split the two local marginals $p_t^{(z),x_i}$, $z = 1, 2$ in two parts, such that the following equations hold

$$\begin{aligned} p_t^{(1),x_i} &= q_{1(1)} \cup q_{2(1)} \\ p_t^{(2),x_i} &= q_{1(2)} \cup q_{2(2)} \end{aligned}$$

According to eq.(4.9), the joint distribution is

$X^{(2)} \setminus X^{(1)}$	$q_{1(1)}$	$q_{2(1)}$
$q_{1(2)}$	$\beta(x_i)q_{1(2)} + (1 - \beta(x_i))q_{1(2)}q_{1(1)}$	$(1 - \beta(x_i))q_{1(2)}q_{2(1)}$
$q_{2(2)}$	$(1 - \beta(x_i))q_{2(2)}q_{1(1)}$	$\beta(x_i)q_{2(2)} + (1 - \beta(x_i))q_{2(2)}q_{2(1)}$

Table 3.1: Representation of eq.(4.9) for $m = 2$.Figure 3.6: On the left, plot of the joint distribution when $\rho = 1\%$. On the right, plot of β against ρ that confirms the perfect correspondence between the two coefficients.

On each period we can express the marginal transition probability in eq.(4.9) as,

$$\begin{aligned}
 p_t^{(1),x_i} &= \sum_{k=1}^2 J_t^{x_i}(X^{(1)}|X^{(2)} = s_k^{(2)})p_t^{(2),x_i}(x_j \in s_k^{(2)}) \\
 &= J_t^{x_i}(X^{(1)}|X^{(2)} = s_1^{(2)})q_{1(2)} + J_t^{x_i}(X^{(1)}|X^{(2)} = s_2^{(2)})q_{2(2)}
 \end{aligned}$$

We can easily define the analytic expressions for the conditional point-in-time probabilities terms $J_t^{x_i}(X^{(1)}|X^{(2)} = s_1^{(2)})$ and $J_t^{x_i}(X^{(1)}|X^{(2)} = s_2^{(2)})$. In fact

$$\begin{aligned}
 J_t^{x_i}(X^{(1)}|X^{(2)} = s_1^{(2)}) &= \frac{1}{\sum_j q_{1(2)}(x_j \in s_1^{(2)})} \left(\beta(x_i)q_{1(2)} + (1 - \beta(x_i))q_{1(2)}p_t^{(1),x_i} \right) \quad (3.19) \\
 J_t^{x_i}(X^{(1)}|X^{(2)} = s_2^{(2)}) &= \frac{1}{\sum_j q_{2(2)}(x_j \in s_2^{(2)})} \left(\beta(x_i)q_{2(2)} + (1 - \beta(x_i))q_{2(2)}p_t^{(1),x_i} \right) \quad (3.20)
 \end{aligned}$$

where the function $\beta(x)$ provides the local sensitivity of $X_t^{(1)}$ to the process $X_t^{(2)}$. The limiting cases of $\beta(x) = 0$ and $\beta(x) = 1$ correspond to zero and full correlation between a name with a given credit quality $X_t^{(1)}$ and the cycle variable. There is a strong economic intuition behind our model settings. In fact the states $X^{(2)} = s_1^{(2)}$ and $X^{(2)} = s_2^{(2)}$ can be associated to the condition that the market is a ‘bad’ and ‘good’ good state of the economy respectively.

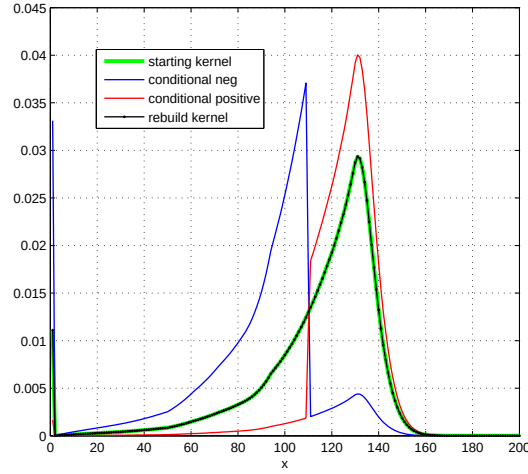


Figure 3.7: Conditional kernels as described by eq.(3.19).

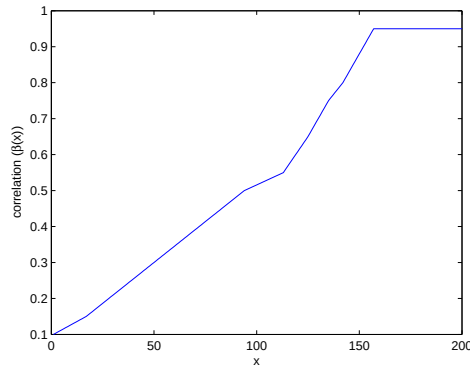


Figure 3.8: Typical calibrated profile of the local correlation coefficient $\beta(x)$. The coefficient tend to zero for DTD approaching zero, hence in case of default. This means that ‘usually’, in standard market condition, investment grade debts (debts with rating from Aaa to Baa) are more correlated than non-investment debts (debts with rating from Baa to Ca).

Three-states Market factor. The privileged assumption in our model to define point-in-time migration probabilities is to consider a three states market factor. We believe that this assumption is the optimal compromise between having a parsimonious model in terms on parameters’ number and the capacity of those parameters to effectively capture and explain the empirical market dynamics.

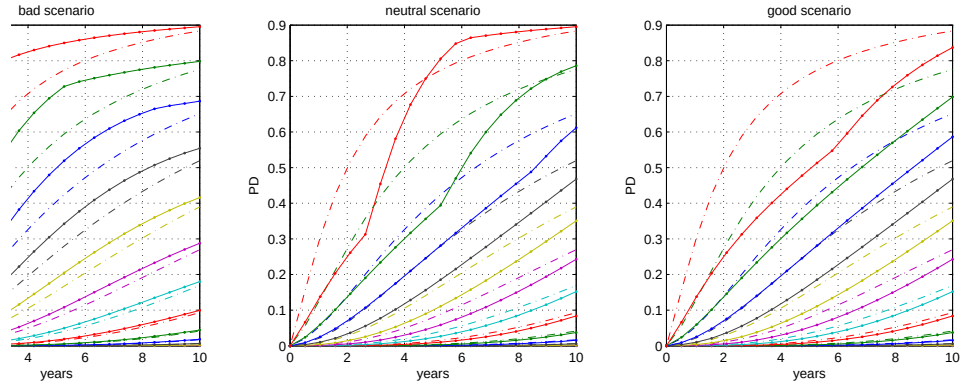


Figure 3.9: Point-in-time (conditional) default probability curves, given that the market is in a bad, neutral or good state. The dashed lines in each plot represent the through-the cycle (unconditional) default probability curves per rating categories, ranging from Ca for the top red curve to Aaa for the bottom one.

n -states chain for the Market factor The number of states of the market factor can be increased if we can assign a solid economic meaning to each partition of the state space. In the following figure we show the conditional transition probability with respect to a 5-states market factor.

3.4 Calibration to Historical Rating Migration Matrices

We are given a term structure of rating migration matrices as described in section 3.1.2 and appendix 4.4.1. Let denote by M the given average migration matrix ('target') and by $L(\theta)$ the unknown parametric one computed with our model, with parameters set θ , as defined in eq.(3.8). The number of parameters θ depends on the model choice and its specification according to equations 3.6, or 3.7 and being the jump component a compound Poisson or a variance gamma. The model choice depends on the historical data quality and properties. In our empirical analysis we found that a convenient model parametrization is by choosing the DTD to be an asymmetric VG process in which case the number of parameters are 40, namely 18 for the local drift, 18 for the local volatility and 4 for the jumps at each point in time in a time-inhomogeneous setting.

At this stage our aim is to introduce a suitable metric able to capture a distance between the two matrices, and then search for the parameters that minimize this distance such that our model can optimally represent and fit the historical market data. In Schuermann and Jafry [1997b] are presented several different metrics for comparing matrices including L^1 and L^2 (Euclidean) distance metrics, and eigenvalue and eigenvector analysis such as singular value decomposition. We adopt a spectral metric that is able to accurately compare both matrices eigenvalues and eigenspaces features and fully capture all the data information content. We believe that this metric has never been used in the context of rating migration matrices estimation and it is by far the most accurate way to compare matrix structure and information. At this end, we first decompose the

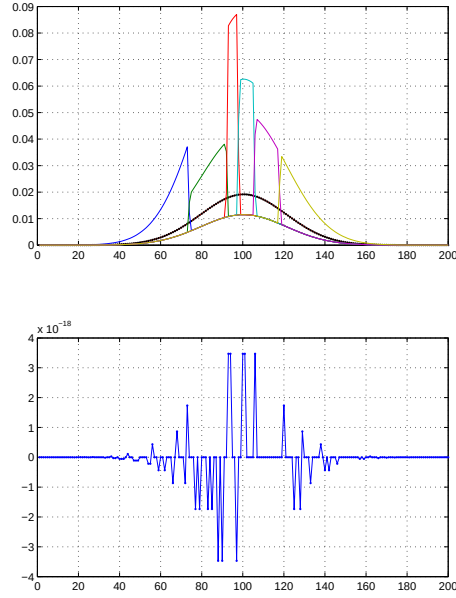


Figure 3.10: It is possible to extend the number of states of the market factor and create several corresponding credit curve scenarios. The above top plot shows the conditional kernels produced by a 5-states market factor. In the bottom plot we show the error produced when rebuilding the through-the-cycle (unconditional) transition kernel, from the 5 conditional ones.

given migration matrix M as following,

$$\begin{aligned} M &= VDV^{-1} \\ \Phi &= V^{-1} \\ \alpha &= \frac{1}{\|VD\|^2} \quad \beta = \frac{1}{\|D\Phi\|^2} \quad \gamma = \frac{1}{\|V\Phi\|^2} \end{aligned}$$

and then we define the spectral metrics $m(M, L)$ as

$$m = \gamma \|V\Phi - I\|^2 + \alpha \|LV - VD\|^2 + \beta \|\Phi L - D\Phi\|^2 \quad (3.21)$$

The calibration routine consists in a sequence of minimization problems of the form

$$\begin{aligned} \min_{\theta_t} m(M_t, L_1(\theta)) & \quad \text{if } t \in [0, 1yr] \\ \min_{\theta_t} m(M_t, L_1 L_2(\theta)) & \quad \text{if } t \in [1yr, 2yr] \\ \min_{\theta_t} m(M_t, L_1 L_2 L_3(\theta)) & \quad \text{if } t \in [2yr, 3yr] \\ & \vdots \\ \min_{\theta_t} m(M_t, L_1 \cdots L_{10}(\theta)) & \quad \text{if } t \in [7yr, 10yr] \end{aligned}$$

and the ultimate result will be a term structure of parameters $\{\theta_t\}$ that fully describes the dynamic of our time-inhomogeneous model. Every nonlinear optimization problem is solved through a global optimization routine directly derived from differential evolution and described in the seminal work of Storn and Price [1995]. See appendix A.3 for details on the solver implementation. We would like to remark that our model $L(\theta)$ is perfectly identifiable. As a demonstrative example, we set some random but meaningful parameters θ_0 and we produce the transition matrix $B = L(\theta_0)$, for $t = 2$ years. We then want to retrieve the parameters by solving the problem

$$\min_{\theta} m(B, L(\theta)) \quad (3.22)$$

Figure 3.11 shows the result of our model identifiability test.

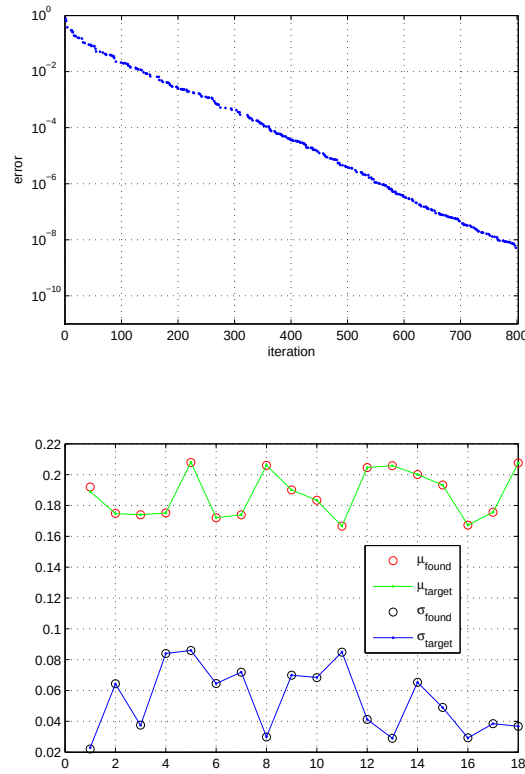


Figure 3.11: The model identifiability is reported by the above plots. In fact the plots report the calibration result of the model local parameters μ_i and σ_i , $i = 1, \dots, 18$. The parameters are randomly selected and the solver stopping criteria is $\min\{800 \text{ iteration, err} = 10^{-10}\}$. Due to the proposed spectral metric of eq.(3.21) the parameters are perfectly retrieved by the heuristic solver, differential evolution.

3.4.1 Results - Fitted model parameters

The following figures and tables report the fitting results for the average through-the-cycle rating migration matrices of the Banking sector provided by Moody's Investor Service. We fit our model, consisting of the asymmetric DTD process with VG dynamics, to the historical rating data for the maturities $T = \{1, 2, 3, 4, 5, 7, 10\}$ years. The results show the profiles of the fitted rating migration matrix, of the diagonal migration probabilities that represent the rating persistency in a specific category and the fitted default curves across rating categories. The fitting errors expressed by the metric in eq.(3.21) are

1yr	2yr	3yr	4yr	5yr	7yr	10yr
0.0113	0.0261	0.0893	0.01352	0.0597	0.0382	0.0989

Table 3.2: Fitting errors. The error is express in terms of the spectral distance as in eq.(3.21)

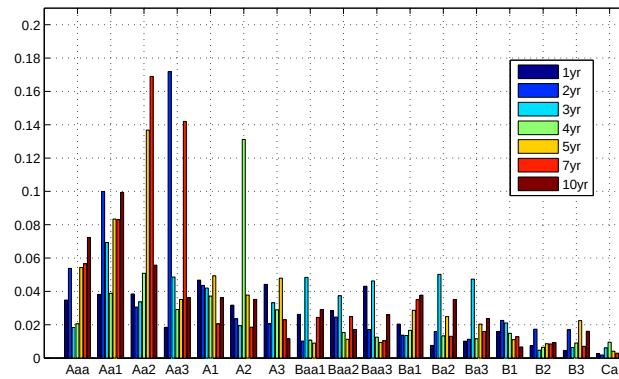


Figure 3.12: Fitted local drift parameters.

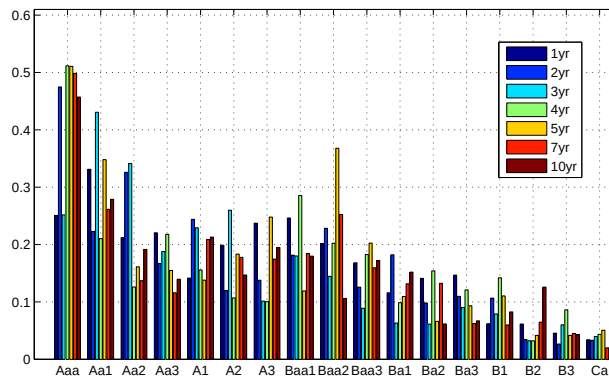


Figure 3.13: Fitted local volatility parameters.

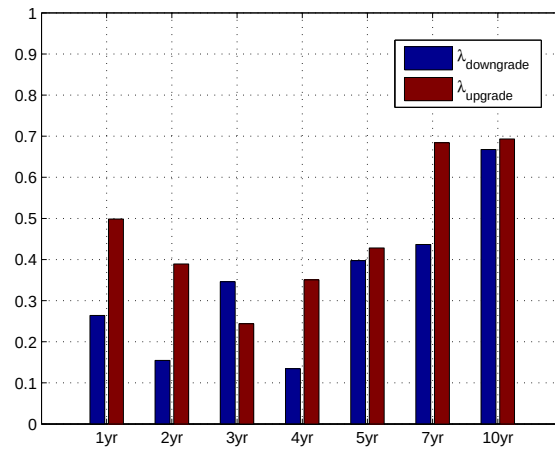


Figure 3.14: Fitted local jumps parameters.

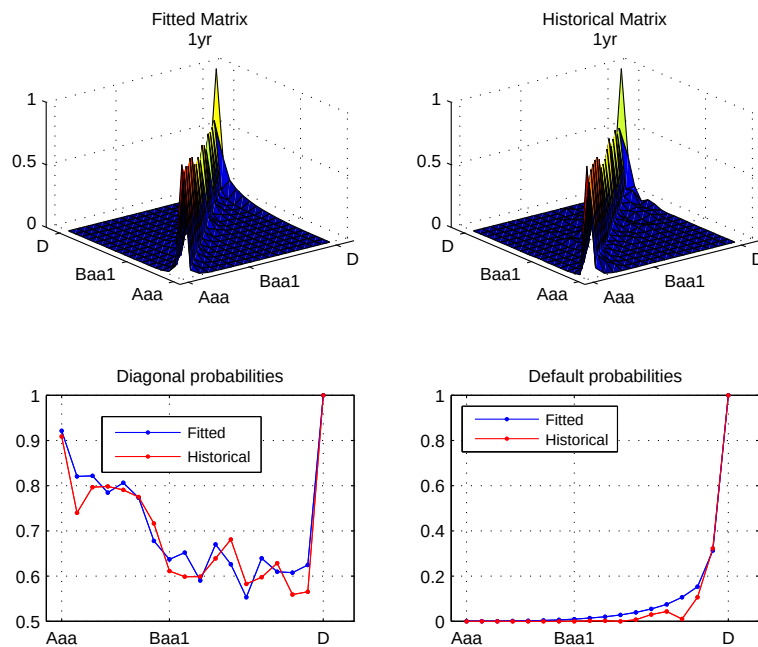


Figure 3.15: Fitted Banking sector average 1yr rating migration matrix.

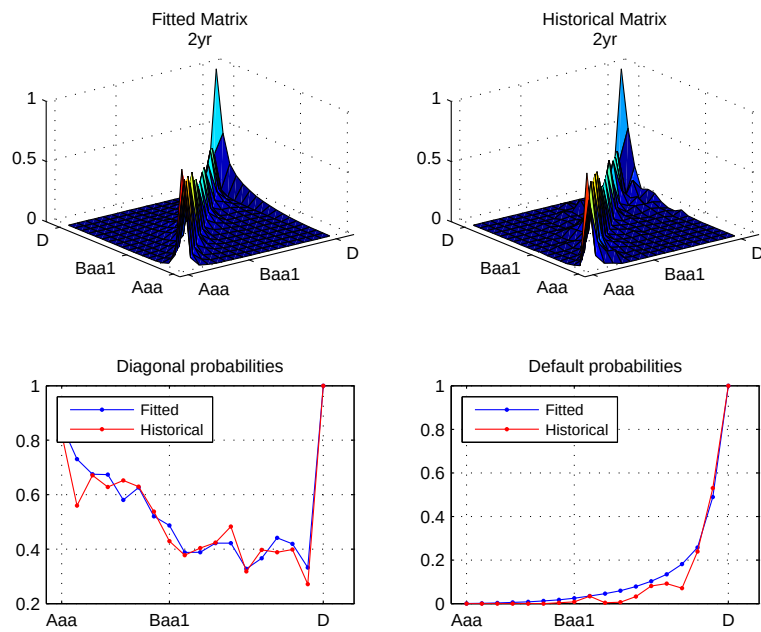


Figure 3.16: Fitted Banking sector average 2yr rating migration matrix.

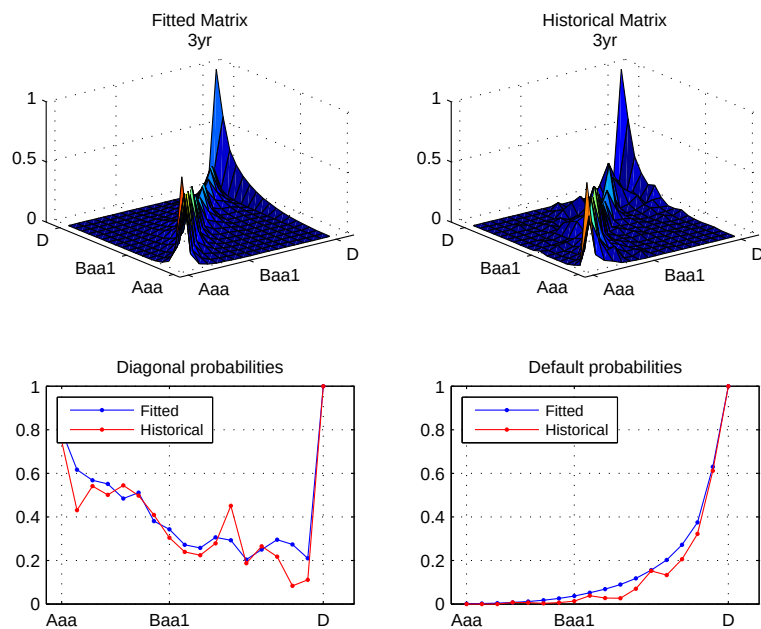


Figure 3.17: Fitted Banking sector average 3yr rating migration matrix.

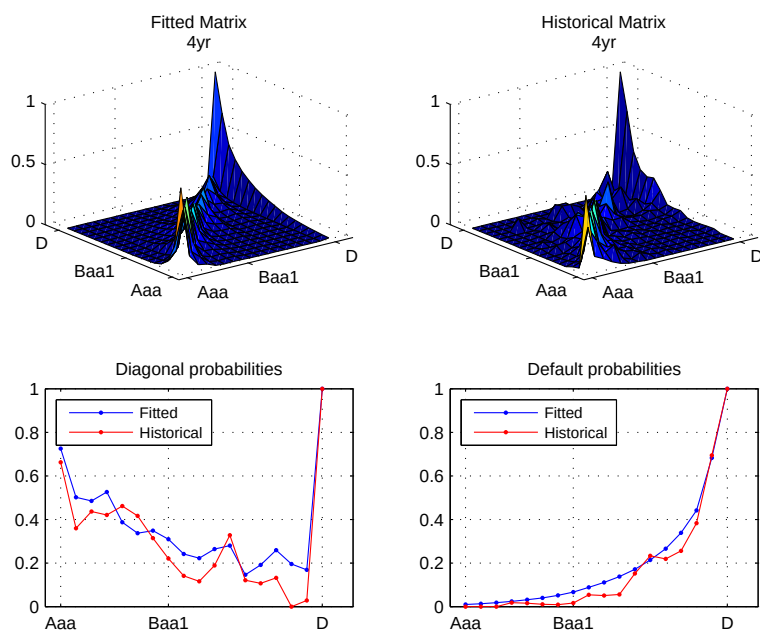


Figure 3.18: Fitted Banking sector average 4yr rating migration matrix.

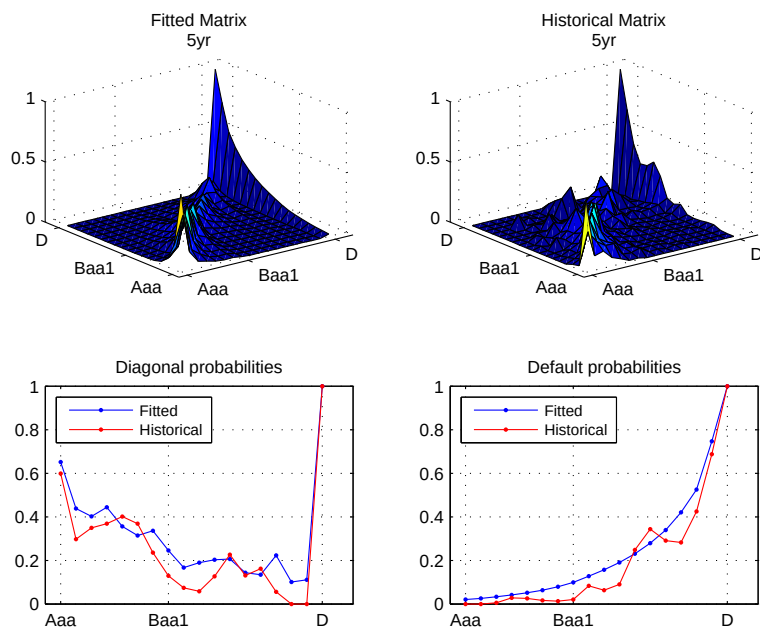


Figure 3.19: Fitted Banking sector average 5yr rating migration matrix.

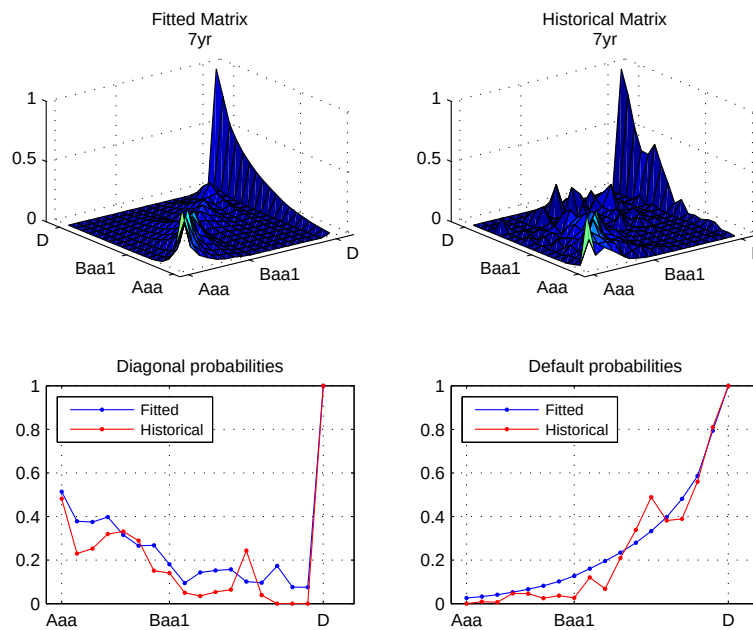


Figure 3.20: Fitted Banking sector average 7yr rating migration matrix.

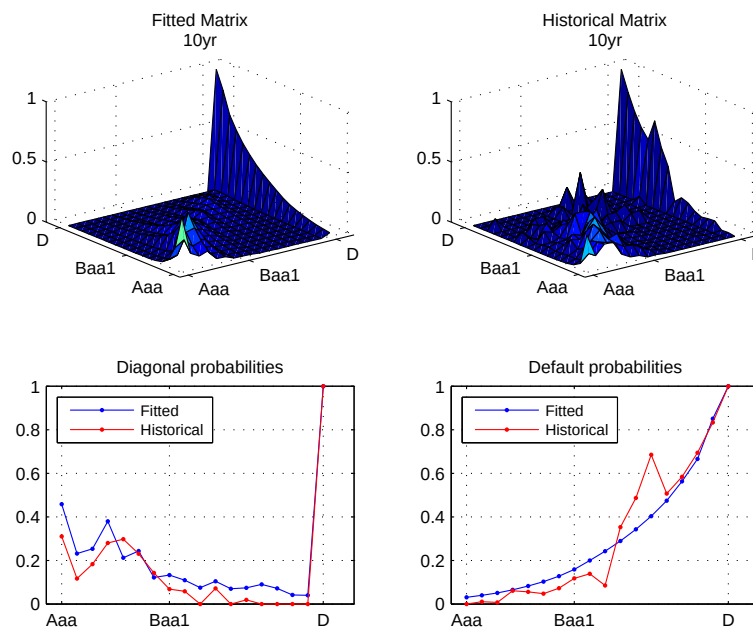


Figure 3.21: Fitted Banking sector average 10yr rating migration matrix.

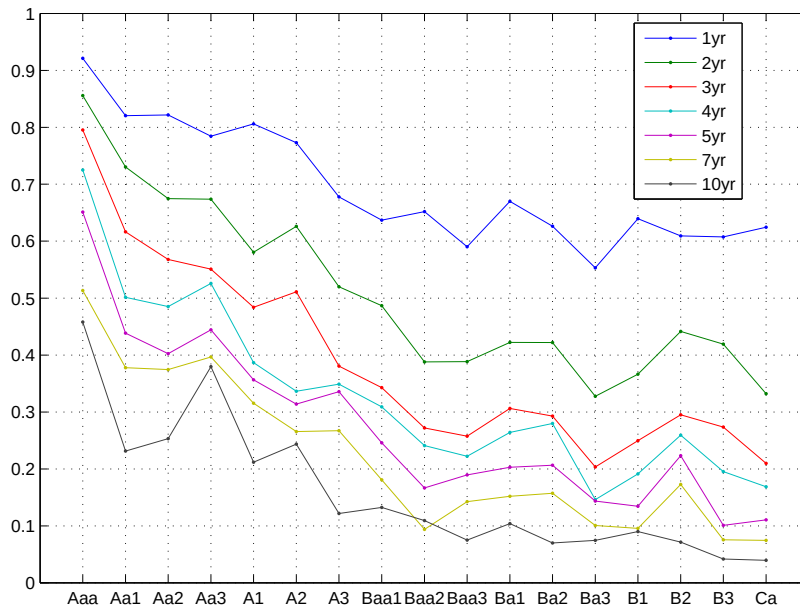


Figure 3.22: Average fitted rating persistency.

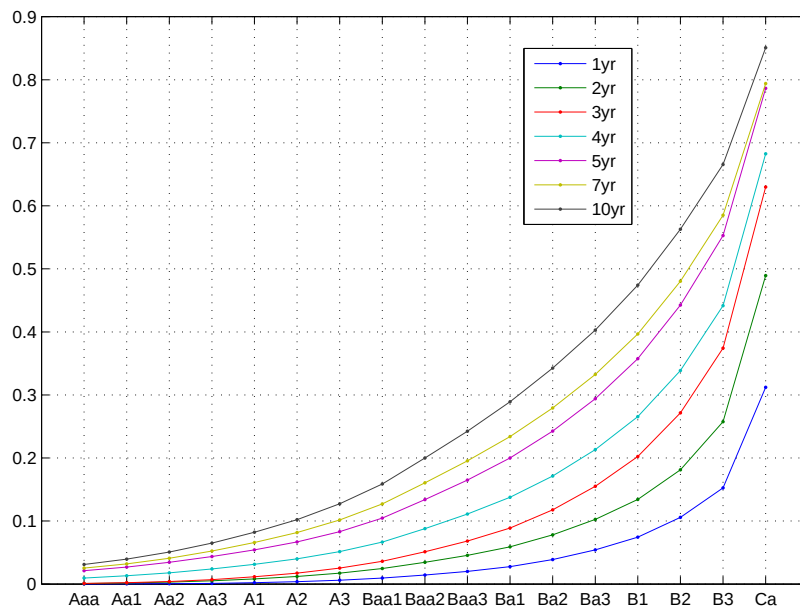


Figure 3.23: Average fitted default probability across rating categories.

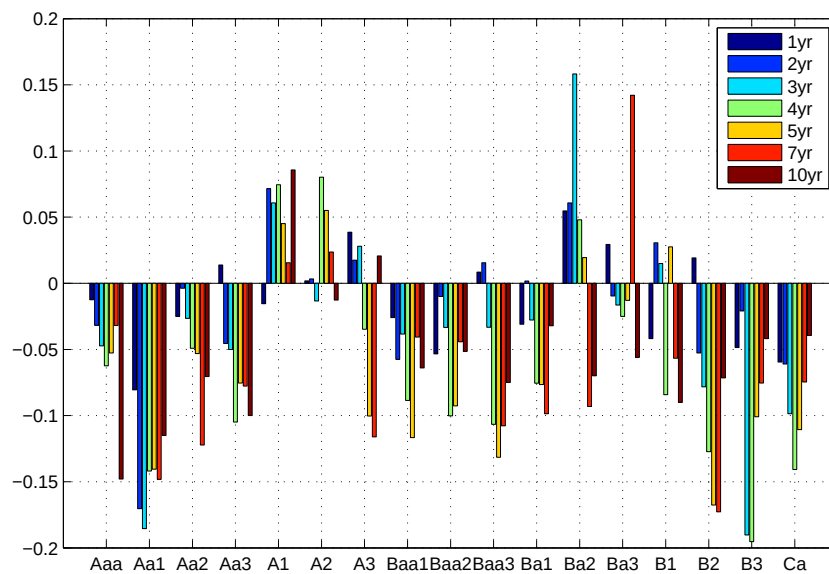


Figure 3.24: Difference between average and average fitted rating persistency. By rating persistency probabilities are denoted the diagonal probabilities of the rating migration matrix.

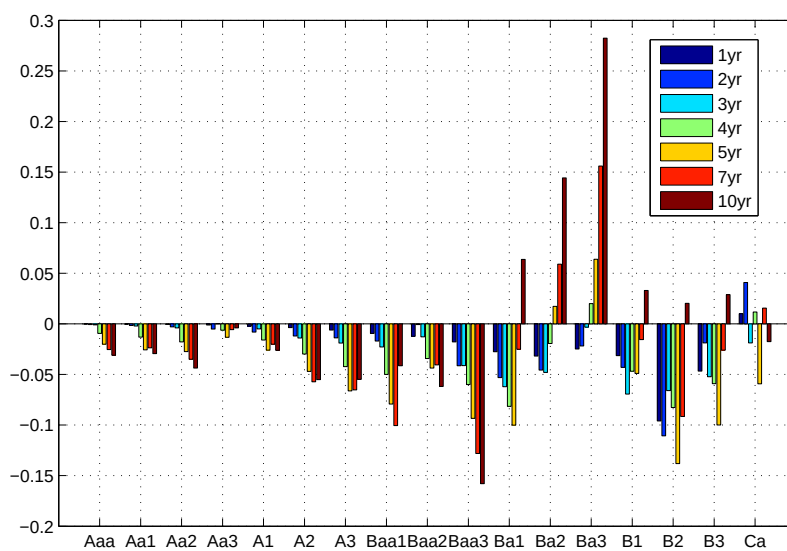


Figure 3.25: Difference between average default probability and average fitted default probability across rating categories.

Notes

3.4.2 Market Data

The Market Data³ consist of historical average migration matrices for Moody's 12 Industry classification, namely. However our analysis mainly focuses on the Banking sector data.

Moody's 12: Sector Names
Banking
Financial (Non-Bank)
Media
Technology
Industrial
Transportation
Hotel, Gaming, & Leisure
Consumer Products
Energy
Utilities
Miscellaneous
Retail
Sovereign

Table 3.3: Moody's 12: Sector Names.

The market data are very noisy and in particular the matrices are not perfectly stochastic, i.e. their rows do not sum to 1. In order to achieve this, all rows have been normalized by their sum. The standard deviation between the initial average migration matrices and the adjusted ones is reported in the following table.

1yr	2yr	3yr	4yr	5yr	7yr	10yr
0.000285	0.000267	0.000215	0.000168	0.00016	0.000145	0.000137

Table 3.4: The standard deviation between the initial and adjusted average migration matrices for the Banking sector.

It is also important to perform some local descriptive analysis of the matrices and at this purpose we report in figures 3.27 and 3.28 the average rating persistency and the average default probability across rating categories respectively. Due to the fact that the average rating matrices have been computed over a long time interval, this implies that the total number of banking corporations has been subject to changes, causing the resulting transition probabilities to exhibits irregularities and sudden jumps.

³The data are dated 2005 and were kindly provided to the author by Moody's Investors Service for research and non-commercial purposes.

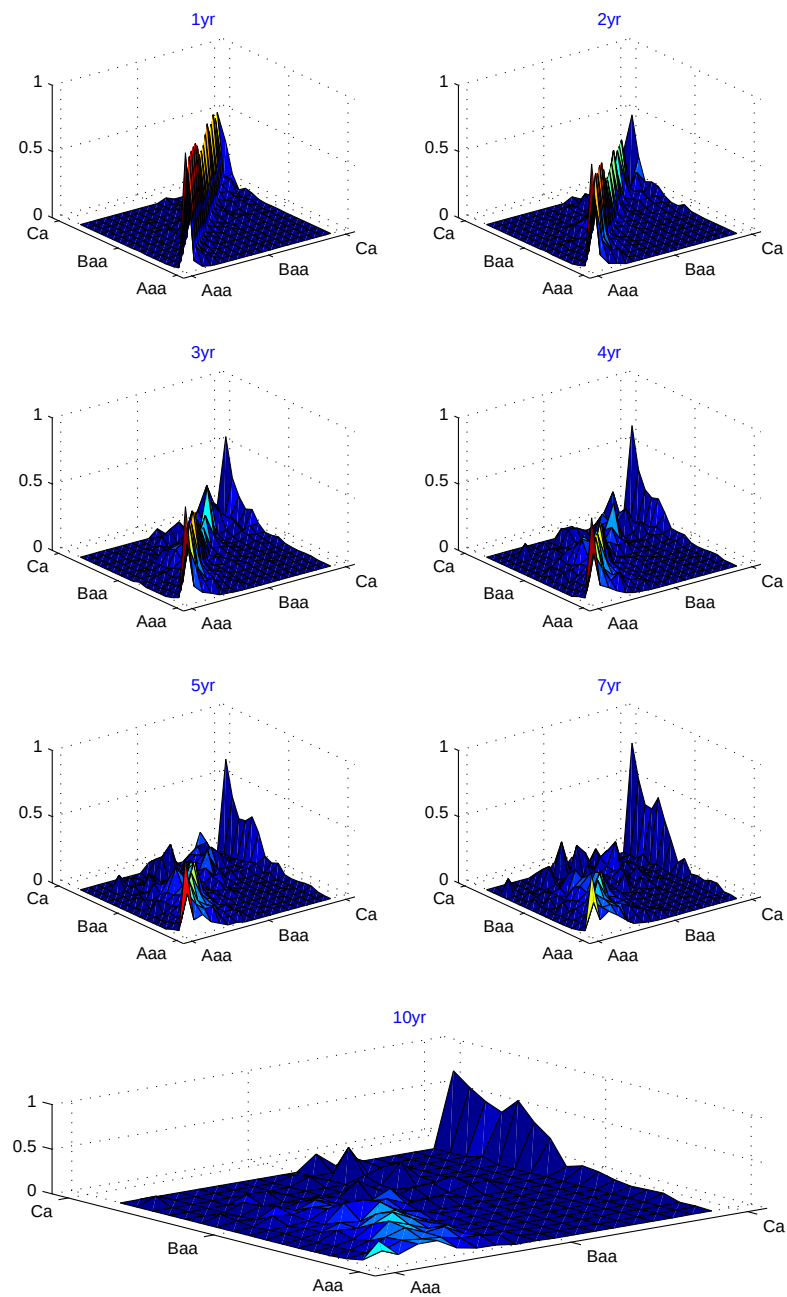


Figure 3.26: Historical average rating migration matrices computed using Banking sector migration data from 1920 to 2005.

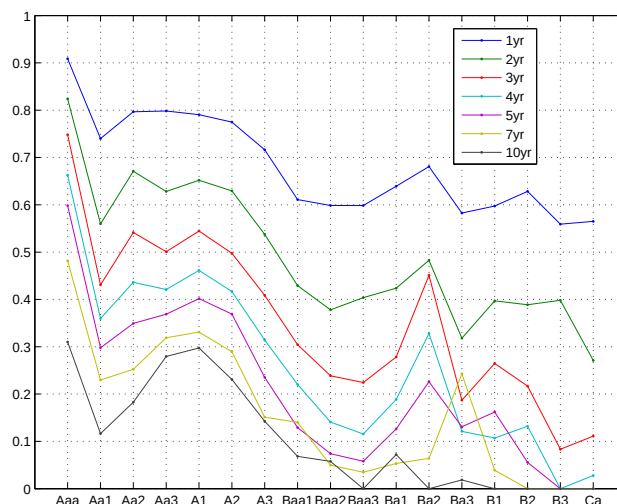


Figure 3.27: Average rating persistency. This data correspond to the diagonal probabilities of the historical average rating migration matrices computed using Banking sector migration data from 1920 to 2005.

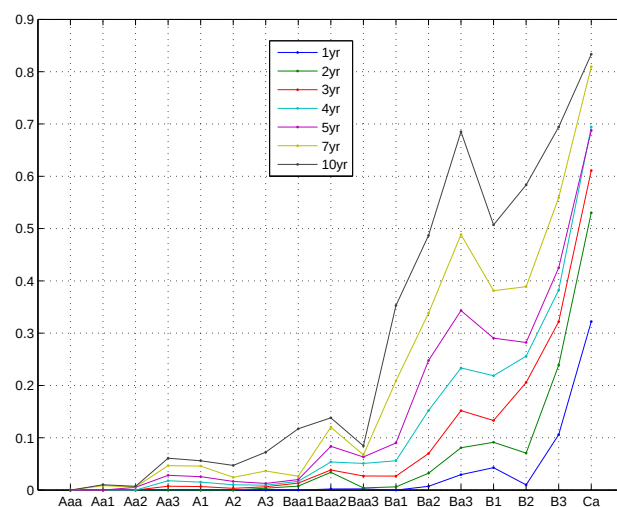


Figure 3.28: Average default probability across rating categories. This data correspond to the last column of the historical average rating migration matrices computed using Banking sector migration data from 1920 to 2005.

Chapter 4

Dynamic Pricing of Index and Bespoke CDO

This Chapter presents a framework to model correlated default events that can be used to price and hedge standard and exotic credit baskets whose values depend on the realized losses of a default portfolio. The model consists of parametric continuous time Markov chain and aims to accurately capture the point-in-time dynamics of the credit worthiness of debt issuers, the single name dynamics, and the dynamics of the correlation among the names underlying the credit portfolio. Key features of the model are the implied credit rating and the local correlation as a function of the implied rating and time. We illustrate through an example how to price and hedge with this model standard CDO index tranches. This method is computationally fast and parsimonious because uses just one matrix. The framework is simple and at the same time very rich in information and this can easily accommodate for the evaluation of MBSs, CMOs and CDO of ABS tranches.

4.1 Literature reviews

The credit market started to grow around 1998 and expanding very fast until the end of 2006 when the credit crunch begun and is currently undergoing a change due to the stringent regulations on capital requirement. A good introduction to the origin of the credit market is in Schönbucher [2003].

This market is still in need for a robust model to price Collateralized Debt Obligations (CDOs). Furthermore a good credit pricing model would help to restore some confidence among investors hit by the credit crunch and by the subsequent volatile market period. Over the last decade the most popular instruments on which the modeling efforts were mainly concentrated proved to be credit default swaps (CDS) and CDOs. The latter offer a challenging complexity to deal with and is the one of the hottest topic in financial research and correlation trading desks. The availability of CDO data for multiple time horizons is still encouraging researcher to develop dynamic models able to fit market data and to track the evolution of the credit risk of a portfolio.

Pricing of a credit portfolio involves either directly or indirectly modeling default dependency among the credit names. There are standard and non-standard market models for valuing portfolio credit derivatives and Hull and White [2008] contains a good overview of such techniques.

The market standard among such models is still the Gaussian copula model Li [2000], a one-factor model that enables easy quotation of market prices. But, it is well-known that the model is inadequate to price nonstandard products. Major attraction of the Gaussian copula model is its simplicity and tractability. It can easily be calibrated to individual hazard rate curves. It can be formulated in closed form providing a semi-analytical framework for pricing. Issues with the Gaussian copula model have been addressed before. Many researchers and practitioners have presented models going beyond the Gaussian copula, and the following is a short overview of the related literature.

In Brigo et al. [2007b] the authors provide an account of the developments in this field and propose a generalized Poisson cluster loss model that is able to fit prices across tranches and maturities. Hull and White [2007] introduces an implied copula model that can calibrate consistently across all CDO tranches for a given maturity. Balakrishna [2009] models the jump distribution in an intensity based model. A doubly stochastic framework for intensity modeling is introduced in Duffie and Garleanu [2001]. Joshi and Stacey [2005] considers an intensity based modeling of business time as a gamma process. Chapovsky et al. [2006] model default intensity as a jump-diffusion process. Errais et al. [2010] present a model based on affine point processes. Some extension of the Intensity-Based framework are Mortensen [2006] and Eckner [2007]. Bielecki et al. [2009] is a comprehensive survey of various methods and results regarding application of the theory of continuous time Markov chains (CTMC) to the valuation of credit derivatives. A CDO pricing model based on Markov chains is presented in Digraziano and Rogers [2005]. Putyatin et al. [2005] use a Markovian approach combined with a Poisson mixing distribution. Bennani [2006] presents a Markov functional approach to instantaneous loss rate. Sidenius et al. [2005] consider a no-arbitrage approach to modeling future loss distribution. Baxter [2006] proposes a threshold model based on time-changed Wiener processes.

In this Chapter we propose a new model to price and hedge correlated credit portfolios, taking into consideration that the correlations among a collection of single names, the creditworthiness of the single credits, the recoveries and the payoffs can undergo a change along the life of the deal. In what follows we present a time-inhomogeneous, fully dynamic, structural and bottom-up model based on CTMC, that is built on the concept of distance to default and enjoys the properties of conditional market recovery and local time-dependent correlation. The model is developed along the lines of Dalessandro [2010b] and the other relevant references are Dalessandro [2010a], Eckner [2007], Albanese et al. [2005a], Sidenius et al. [2005], Schönbucher [2005], Brigo et al. [2007a] and Bielecki et al. [2009].

Our work continues this line of research by proposing a new way to condition all the marginal single name credit names in the portfolio, and generate a multivariate time dependent distribution in a more intuitive way and that better replicates the market movements. Our model calibrates under the objective $\mathbb{P}$ -measure and the risk neutral $\mathbb{Q}$ -measure and this can lead to natural considerations on the risk premia and the change of measure $d\mathbb{Q}/d\mathbb{P}$. We also give empirical evidence of the phenomenon of default clustering among the portfolio constituents along the life of the trades.

The Chapter is organized as following: In Section 2 we introduce the notation used and the model specification for the single credit names and also for their multivariate joint distribution. Section 3 gives more details about the physical and risk-neutral measure and the risk premia that characterize the passage between the two measures. In section 4 we apply the model to the fitting of CDO index tranches, illustrate the greeks computation

and give details of the model calibration through an example.

4.1.1 Model Uniqueness and its Applications

The model proposed in this Chapter presents features that make it one of the most comprehensive frameworks, among the existing bottom-up credit models in the literature, for pricing and risk managing tranching portfolios of credit derivatives.

Among the relevant features, the model provides a very rich credit scale, denser than the standard 18 notches rating scale of the agencies, and allows to better locate the creditworthiness of a debt that can be at the center as well as at the top or at the bottom of a rating category. This feature allows to identify the relative value of a debt with respect to the market assigned rating. The model can, in fact, determine the debt implied rating (see fig.4.4). Some credit names can trade at a spread that doesn't reflect the rating assigned by the rating agency. This is a clear relative value opportunity and the fact that a credit name rating can undergo through an imminent public change of its rated creditworthiness.

Another important feature is that the model allows to compute and visualize the implied local correlation among names in the credit portfolio as a function of the implied creditworthiness and time (see fig.4.7). There is no model that can provide such accurate information and a trader can use it to better justify the loss distribution movements and the clustering of defaults. This model gives a better intuition to tranches' price levels, and their relative movements can be explained through the price sensitivities that are readily available after the initial calibration to the market quotes.

The aim of the proposed model is to allow practitioners and investors to better understand the impact of market information on the degree of credit and defaults events clustering that determines the price of credit derivatives. Especially during this turbulent and very volatile period a sound and reliable model would help to restore some confidence among investors with respect to the quantitative valuation of the price and risk of structured credit derivatives. The assessment of the point-in-time credit risk premia through an intuitive and accurate framework is a fundamental tool for bank that back the performance of portfolios of corporate debts retaining capital designed to withstand default clustering at extreme high confidence levels such as 99.9%. Financial institutions will benefit of the model and findings within this Chapter to better quantify their capital requirement and corresponding liquidity facilities in order to comply with the new rules from Basel III. The proposed framework consists of a parametric continuous time-inhomogeneous Markov chain and is computationally parsimonious because uses just one matrix. The model can be readily calibrated to some given market prices and used as a trading and investment tool to hedge existing credit positions.

4.2 Notation and Model Specification

In this section we present the main concepts and building blocks of our credit model, that is specified under both the physical and risk neutral measure, and this feature is fundamental to quantify risk premia.

Our aim is to formulate a pricing model for single name assets subject to default risk, and also to characterize a dynamic correlation among them and be able to price portfolio products like collateralized debt obligations (CDO). At this purpose we don't represent di-

rectly the credit spread evolution as in Lando [2004], Schönbucher [2003], Hull and White [2000], but instead a proxy of it, namely the distance to default (DTD) process. This concept was introduced in early works Albanese and Chen [2003], Albanese et al. [2005a]. We privileged this approach because in structured finance a key role is played by the rating, an ordered system of credit quality levels, and its variation over time that reflect its instantaneous creditworthiness. The DTD contains the survival information of the defaultable security.

Formally, let $(\Omega \subseteq S, \mathcal{F}_t, \mathbb{P})$ be a probability space where is defined the $\mathbb{R}^+$ valued Feller process process of the DTD, $X = \{X_t : 0 \leq t \leq T\}$ which is right continuous with left limits and filtration $\mathcal{G}_t = \sigma\{X_s; 0 \leq s \leq t\}$. Let N_t be a standard Poisson jump process which is independent of X and filtration $\mathcal{H}_t = \sigma\{N_s; 0 \leq s \leq t\}$. The state variable X and the jump process N are independent and the filtration $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$ contains the information of both processes. Let $\{\lambda_X(s) : s \geq 0\}$ be an intensity process that we assume inhomogeneous and non-negative and that is associated to the chosen DTD dynamics. We define the default time τ as follows:

$$\tau = \inf\{t \in \mathbb{R}^+ : \int_0^t \lambda_X(s) ds \geq N_1\} \quad (4.1)$$

where N_1 is the first jump time of the poisson process. The risk-neutral survival probability expression is derived using the definition of a Cox process and is as following:

$$\begin{aligned} \mathbb{Q}(\tau > t | \mathcal{F}_s) &= \mathbf{1}_{\{\tau > s\}} \mathbb{E}_s^{\mathbb{Q}}[1_{\tau > t} | \mathcal{F}_s] \\ &= \mathbf{1}_{\{\tau > s\}} \mathbb{E}_s^{\mathbb{Q}}\left[\exp\left(-\int_s^t \lambda_X(u) du\right) \middle| \mathcal{G}_s\right], \quad s \in [0, t]. \end{aligned} \quad (4.2)$$

For $0 \leq t \leq T$, the unique arbitrage-free price at time t for a security that pays one currency unit at time t is

$$\begin{aligned} V(0, t) &= \mathbb{E}^{\mathbb{Q}}\left[\exp - \int_0^t r_s ds 1_{\{\tau > t\}}\right] \\ &= \mathbb{E}\left[\mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^t r_s ds\right) 1_{\{\tau > t\}} \middle| \mathcal{G}_t\right]\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^t r_s ds\right) \mathbb{E}\left[1_{\{\tau > t\}} \middle| \mathcal{G}_t\right]\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^t r_s ds\right) \exp \int_0^t \lambda_X(s) ds\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^t r_s + \lambda_X(s) ds\right)\right] \end{aligned}$$

The price of the above risky bond at time 0 with zero recovery can be easily extended to any contingent claim with a promised payment of $f(X_t)$ and an actual payment of $f(X_t)1_{\{\tau > t\}}$. All we need is the state variable information, and the information that default has not yet occurred to compute the price of a defaultable claim.

The DTD process X represents our state variable and is specified by an asymmetric variance gamma (VG) process, see Madan et al. [1998a]. The VG process is defined as following:

$$d \log X_t = m dt + \mu(X_t) dg_t + \sigma(X_t) dW(g_t), \quad X(0) = X_0 \quad (4.3)$$

where $W(t)$ is a standard Brownian motion, $m \in \mathbb{R}$, and $\mu(x)$ and $\sigma(x)$ are a drift function and volatility function respectively, satisfying the usual regularity conditions, see Shreve [1993], and g_t a gamma process. In fact, the VG process is obtained on evaluating the Brownian motion W at a time given by a two-parameter gamma process $g_t(\hat{\mu}, \nu)$, independent of W , and with increments gamma distributed with the mean $\hat{\mu}$ and variance ν both multiplied by the size of the increment; that is, for any $t \geq 0$ and time increment $\delta > 0$,

$$g_{t+\delta} - g_t \sim \Gamma\left(\delta; \frac{\hat{\mu}^2}{\nu}, \frac{\nu}{\hat{\mu}}\right)$$

with $\hat{\mu}, \nu \in \mathbb{R}^{++}$. The gamma process is also often written as $\Gamma(t; \gamma, \lambda)$, therefore parametrized in terms of $\gamma = \frac{\hat{\mu}^2}{\nu}$ that controls the rate of jump arrivals and the scaling parameter $\lambda = \frac{\nu}{\hat{\mu}}$ that inversely controls the jumps size, see Applebaum [2004]. The VG process in eq.(4.3) can be also expressed as the difference of two gamma processes

$$d \log X_t = m dt + dg(\mu_u, \nu_u) - dg(\mu_d, \nu_d), \quad X(0) = X_0 \quad (4.4)$$

The explicit relation between the parameters of the gamma processes differenced in (4.5) and the original parameters of the VG process (4.3) is detailed in Madan et al. [1998b]. In particular, we specify the DTD process with an asymmetric gamma subordination, as

$$d \log X_t = m dt + \mu \overline{dg}_t + \sigma dW(\overline{g}_t), \quad X(0) = X_0 \quad (4.5)$$

where $\overline{dg}_t = dg_t(\mu_u, \nu_u)$ for upward jumps and $\overline{dg}_t = dg_t(\mu_d, \nu_d)$ for downward jumps. More generally the VG process is a pure jump process of infinite activity and finite variation and able to capture frequent small changes and rare small jumps. The VG process can be alternatively be approximated, see S. Kotz [2001], by a compound Poisson process that leads to a representation with explicitly given size of jumps and their locations. Further details on the VG process are in and its risk neutral characterization are in Madan and Seneta [1990]. The process X is then approximated by an equivalent n -states continuous time-inhomogeneous Markov chain $X^{(n)}$ and confined on the discrete state space

$$\mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n] \quad (4.6)$$

that maps the rating scale

$$\{Aaa, Aa, A, Baa, \dots, D\} \rightarrow \mathcal{S}_X$$

where $\{Aaa, \dots, D\}$ are the $K = 18$ symbols of the extended Moody's rating system and $n > 18$, usually $n = 200$. The state $\underline{x}$ represents the credit risk free state while $\bar{x}$ is the default state and we assume that are both absorbing states of the chain. The state space $\mathcal{S}_X$ is subdivided into K subintervals of adjacent states

$$I_k = [x_{k-1}, \dots, x_k], \quad k = 1, \dots, 18, \quad x_0 = \underline{x}, \quad x_{18} = x_n$$

each of them corresponding to the k -th rating class. In this way we have available a richer set of rating levels that are going to be very useful when calibrating our model parameters to observed debt instruments, and if a process is in I_k at time t , then is said to have a credit rating of k . Our model relies on the construction of the approximated infinitesimal generator matrix $\mathcal{L}(t)$ of $X^{(n)}$. In order to build $\mathcal{L}(t)$ for the asymmetric VG process in

eq.(4.5) we do as following, at each point $t > 0$, used to calibrate the time-inhomogeneous chain.

The first step is to build the tridiagonal matrix $\mathcal{A}(x, y; t) = \{a_t(x_i, x_j)\}$, $i, j = 1, \dots, n$ representing the approximated generator for the discretized diffusion with state dependent drift $\mu(x; t)$ and volatility $\sigma(x; t)$. Each matrix entry can be calculated by solving the following system:

$$\begin{cases} a_t(x_1, x_2) = a_t(x_n, x_{n-1}) = 0 \\ a_t(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{\mu(x_i; t)}{h} + \frac{\sigma(x_i; t)^2}{h^2} \right) \\ a_t(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma(x_i; t)^2}{h^2} - \frac{\mu(x_i; t)}{h} \right) \\ a_t(x_i, x_i) = -(a_t(x_i, x_{i-1}) + a_t(x_i, x_{i+1})) \end{cases}$$

for all $i = 2, \dots, n-1$, with $h = x_i - x_{i-1}$, and $\frac{\sigma(x_i; t)^2}{h} \leq \mu(x_i; t) \leq \frac{\sigma(x_i; t)^2}{h}$. Then we build the approximated generator of a VG process that is denoted by $\mathcal{A}_{vg}(x, y; t) \in \mathbb{R}^{n \times n}$, and can be derived as the time subordination of a diffusive process. If $\mathcal{A}(x, y; t)$ is a diagonalizable matrix, the following equation hold:

$$\mathcal{A} = VDV^{-1}$$

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & & \\ 0 & \lambda_2 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & \lambda_{n-1} & 0 \\ & & 0 & 0 & \lambda_n \end{bmatrix}$$

We then compute the Laplace transform of the Gamma subordinator, and apply the resulting map to the spectrum of the original diffusive process to obtain the VG generator.

$$\begin{aligned} \tilde{\mathcal{L}}\{\Gamma(x; \frac{\hat{\mu}^2(t)}{\nu(t)}, \frac{\nu(t)}{\hat{\mu}(t)})\} &= \int_0^\infty e^{-sx} \Gamma(x; \frac{\hat{\mu}^2(t)}{\nu(t)}, \frac{\nu(t)}{\hat{\mu}(t)}) dx = e^{-\frac{\hat{\mu}^2(t)}{\nu(t)} \log(1 + \frac{\nu(t)}{\hat{\mu}(t)} s)} = e^{\Psi(s)} \\ \bar{D} &= \Psi(D) \\ \mathcal{A}_{vg} &= V\bar{D}V^{-1} \end{aligned}$$

where $\mathcal{A}_{vg} = \mathcal{A}_{vg}(x, y; t, \mu(t), \sigma(t), \hat{\mu}(t), \nu(t))$. Finally we can compute the infinitesimal generator for the asymmetric VG process.

At this purpose we first build two infinitesimal generators, $\mathcal{A}_d(t) = \{a_{t,d}(i, j)\}_{i,j=1,\dots,n}$ equal to $\mathcal{A}_{vg}(x, y; t, \mu(t), \sigma(t), \hat{\mu}_d(t), \nu_d(t))$ for the downward transitions and $\mathcal{A}_u(t) = \{a_{t,u}(i, j)\}_{i,j=1,\dots,n}$ equal to $\mathcal{A}_{vg}(x, y; t, \mu(t), \sigma(t), \hat{\mu}_u(t), \nu_u(t))$ for the upward transitions respectively. The resulting approximated generator matrix $\mathcal{L}(t) = \{l_t(i, j)\}_{i,j=1,\dots,n}$ can be calculated as,

$$\begin{aligned} l_t(i, j) &= \begin{cases} a_{t,d}(i, j), & i > j \\ a_{t,u}(i, j), & i < j \end{cases} \\ l_t(i, i) &= -\sum_{j \geq i} a_{t,u}(i, j) - \sum_{j \leq i} a_{t,d}(i, j) \end{aligned}$$

The infinitesimal generator $\mathcal{L}(t)$ will be a time dependent parametric generator $\mathcal{L}(t; \Theta) \in \mathbb{R}^{n \times n}$ fully specified by the set of time dependent parameters

$$\Theta = \left\{ \mu(x; t), \sigma(x; t), \hat{\mu}_u(t), \hat{\mu}_d(t), \nu_u(t), \nu_d(t) \right\}$$

Given the generator $\mathcal{L}(t)$, it is immediate to generate a transition kernel via the Kolmogorov's forward equation

$$\frac{\partial p_t(x, y)}{\partial t} = \mathcal{L}(t; \Theta) p_t(x, y), \quad \text{with } p_0 \text{ starting probability} \quad (4.7)$$

The probability migration matrix containing the survival information of the defaultable security is the operator semigroup associated to the parametric infinitesimal generator $\mathcal{L}(t)$. If we denote by $p_t(x, y)$ the transition probability matrix generated by the DTD chain over the time interval $[0, t]$, with $t \geq 0$, the default probability is represented by the transition probability entry of the matrix $p_t(x, y)$ into the default state $\bar{x}$, given a starting status x_0 , namely $p_t(x_0, \bar{x})$.

4.2.1 Conditional Model Dynamics

Having characterized the process for the credit quality X_t , the next step is to introduce correlations. We do so by conditioning to a market factor that we assume to be a joint and identical process to the DTD.

Without loss of generality, we assume that the distribution of the market factor is the same as the DTD and these two processes will be modeled as identical Markov chains, therefore by just one CTMC. But this last specification is equivalent to state that the market factor is the calibrated DTD process. Let $X^{(z)} = \{X_t^{(z)}\}_{t \geq 0}$, $z = 1, 2$ be two identical continuous time Markov chains¹ with discrete state space $\mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n]$ and built as illustrated in the previous section, with n a positive integer number representing the number of states. We furthermore denote the discrete state space for both chain by $\mathcal{S}_X = [\underline{x} = x_1, x_2, \dots, \bar{x} = x_n]$. $X^{(n)}$ and X are close to each other during the finite time interval of interest. We consider the probability transition matrices,

$$\begin{aligned} P_t^{(z)} = P(X_t^{(z)}) &:= \{p_t^{(z)}(x_i, x_j)\} \\ &= \{p_t^{(z), x_i}(x_j)\} \quad \text{For } i, j = 1, \dots, n, \quad z = \{1, 2\}, \quad t \geq 0. \end{aligned}$$

for the chains $X^{(n)}$ and we would like to define a joint distribution that correlates the dynamic of the marginals $p_t^{(z), x_i}$ for all i . Let $F_t^{(z)} := \{F_t(x_i, x_j)\}_{i, j=1, \dots, n}$ the corresponding distribution function matrices.

We would like to generate an arbitrary partition of the probability mass of each transition matrix row, and define its corresponding partitioned state space. Given a row $p_t^{(z), x_i}$ of the matrix $P_t^{(z)}$, we have that $\sum p_t^{(z), x_i} = 1$. We then subdivide the entire probability mass into a sequence of probability masses $q_1, q_2, \dots, q_m$, such that $\sum_{j=1}^m q_j = 1$, and define a corresponding increasing sequence of cumulative probability masses $0 \leq c_1, c_2, \dots, c_m \leq 1$, where

$$c_k = \sum_{j=1}^k q_j, \quad k = 1, \dots, m$$

and the sequence of corresponding percentiles

$$b_k = F_t^{-1(z)}(c_k), \quad k = 1, \dots, m. \quad (4.8)$$

¹Please note that where the index (z) is not specified, the value under consideration is assumed to be the same.

This means that the support of each probability mass will be as following

$$\begin{aligned} q_1 &\in s_1 = [\underline{x} = x_1, b_1], \\ q_2 &\in s_2 = [b_1, b_2], \\ &\vdots \\ q_m &\in s_m = [b_{m-1}, b_m = \bar{x} = x_n] \end{aligned}$$

Note that the value of every local probability mass $q_k(x_j)$ is fully defined by $p_t^{(z),x_i}(x_j)$, such that

$$q_{k(z)}(x_j) = p_t^{(z),x_i}(x_j) \quad \text{for all } x_j \in s_{k(z)}$$

Definition 4.1 We define $\beta(x_i)$, for all i , to be the local correlation function. Given the marginals matrices $P_t^{(z)}$, and their corresponding sequences of probability masses $(q_1^{(z)}, \dots, q_m^{(z)}) = (q_1, \dots, q_m)$, we can define a joint local density for all row x_i , with $\bigcup_{k=1}^m s_k = \mathcal{S}_X$, with local marginals $p_t^{(z),x_i}$ by the following set of equations

$$J_t^{x_i}(x_j^{(1)}, x_j^{(2)}) = \begin{cases} \beta(x_i)q_{k(1)}(x_j^{(1)}) + (1 - \beta(x_i))q_{k(1)}(x_j^{(1)})q_{k(2)}(x_j^{(2)}) & \text{if } k^{(1)} = k^{(2)} \\ (1 - \beta(x_i))q_{k(1)}(x_j^{(1)})q_{k(2)}(x_j^{(2)}) & \text{if } k^{(1)} \neq k^{(2)} \\ \sum_k^m q_{k(z)} = 1 & z = 1, 2 \end{cases} \quad (4.9)$$

that generates a local correlation equal to

$$\begin{aligned} \rho(x_i) &= \frac{\mathbb{E}_{x_i}[X_t^{(1)}X_t^{(2)}] - (\mu(x_i)t + x_i)}{\sigma^2(x_i)t} \\ &= \frac{\sum_{k(1)}^m \sum_{k(2)}^m s_{k(1)}s_{k(2)}J_t^{x_i}(x_j^{(1)}, x_j^{(2)}) - (\mu(x_i)t + x_i)}{\sigma^2(x_i)t} \end{aligned}$$

with $\mu(x)$ and $\sigma(x)$ the local specifications of the drift and volatility function respectively for the chain. Considering n -states for the driving market factor, in each period we can express the marginal transition probability as,

$$\begin{aligned} p_t^{(1),x_i} &= \sum_{k=1}^n J_t^{x_i}(X^{(1)}|X^{(2)} = s_k^{(2)})p_t^{(2),x_i}(x_j \in s_k^{(2)}) \\ &= \sum_{k=1}^n J_t^{x_i}(X^{(1)}|X^{(2)} = s_k^{(2)})q_{k(2)} \end{aligned}$$

4.3 Model Identification under the physical and risk neutral measure

The calibration of the DTD process involves three sequential steps. Firstly we estimate the $\mathbb{P}$ measure in such a way to ensure agreement with rating transition probabilities and historical probabilities of default. Secondly we estimate the $\mathbb{Q}$ measure from the term structure of quoted single name CDS spreads. The third and last stage consists in the evaluation of the CDO tranche prices that includes an adjustment for the market price of correlation risk.

Calibration under $\mathbb{P}$

The model under $\mathbb{P}$ is characterized by the following set of parameters

$$\Theta^{\mathbb{P}} = \left\{ \mu^{\mathbb{P}}, \sigma, \hat{\mu}_u, \hat{\mu}_d, \nu_u, \nu_d \right\} \quad (4.10)$$

where $\mu^{\mathbb{P}}(x; t)$ is the local drift, $\sigma(x; t)$ is the local volatility, $\hat{\mu}_u(t)$, $\nu_u(t)$ specify the upward jumps component, $\hat{\mu}_d(t)$, $\nu_d(t)$ specify the downward jumps component. Note that all the parameters are time dependent, because we are calibrating a time-inhomogeneous model to the available term structure of market data that, in the case of the $\mathbb{P}$ -measure, consists of average rating migration matrices. Let's denote by $\bar{x}_i \in I_i$, the starting state, where I_i is the i -th rating class. The conditional transition probability $\hat{p}_{ij}(t)$ that an obligor with a given initial rating i at time 0 will have a rating j at a later time $t > 0$ can be estimated by matching it with historical averages provided by rating agencies. We model this quantity as follows:

$$\hat{p}_{ij}(t; \Theta^{\mathbb{P}}) = \sum_{z=k_j-1}^{k_j-1} p_t(\bar{x}_i, z; \Theta^{\mathbb{P}})$$

where $\bar{x}_i$ is a point in the interval I_i which represent the entire i -th rating class and is part of the model specification. For simplicity $\bar{x}_i$ can be chosen as the midpoint of the interval I_i . In this way it is possible to calculate from a parametric transition probability kernel matrix $p_t(\bar{x}_i, z; \Theta^{\mathbb{P}}) \in \mathbb{R}^{n \times n}$, $t \geq 0$, a parametric migration matrix of dimensions $K \times K$, that can be directly compared with a term structure of historical average rating matrices. Typical maturities for the rating matrices are $T = \{1, 2, 3, 4, 5, 7, 10\}$ years, and therefore the corresponding dimensionality of the calibrated model parameters would be $\mu^{\mathbb{P}}(x; t) \in \mathbb{R}^{18 \times 7}$, $\sigma(t; x) \in \mathbb{R}_+^{18 \times 7}$, $\hat{\mu}_u(t), \nu_u(t), \hat{\mu}_d(t), \nu_d(t) \in \mathbb{R}_+^{1 \times 7}$. A more extensive characterization of this model under the measure $\mathbb{P}$ and its calibration to rating migration matrices is reported in Dalessandro [2010b].

Calibration under $\mathbb{Q}$

In this section we illustrate the calibration steps necessary to estimate the $\mathbb{Q}$ measure from the term structure of quoted single name CDS spreads and then the evaluation of the standard index CDO tranches market quotes. Details and market conventions on credit default swap index tranches and their pricing can be found in appendix 4.4.1. The proposed model is fully specified, under the measure $\mathbb{Q}$, by the following set of parameters

$$\Theta^{\mathbb{Q}} = \left\{ \mu^{\mathbb{Q}}, \sigma, \hat{\mu}_u^{\mathbb{Q}}, \hat{\mu}_d^{\mathbb{Q}}, \nu_u^{\mathbb{Q}}, \nu_d^{\mathbb{Q}}, \mathbf{q}, \beta \right\} \quad (4.11)$$

that need to be calibrated to the observed prices. The local volatility function $\sigma(t; x)$ is calibrated under the physical measure $\mathbb{P}$, and therefore is not recalibrated under $\mathbb{Q}$. We plot its calibrated profile in fig.(4.1).

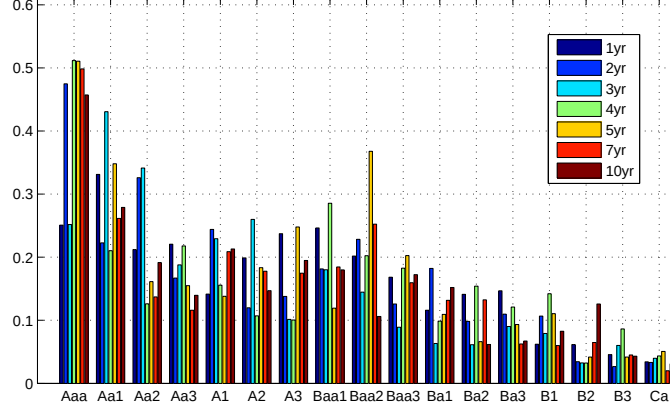


Figure 4.1: The above bars represent the local volatility $\sigma(t; x)$ for the DTD under the $\mathbb{P}$ and $\mathbb{Q}$ measures.

We find the parameters $\Theta^{\mathbb{Q}}$ with a 2-steps calibration procedure where:

1. Firstly, we fit the parameters subset

$$\Theta_{\text{Index Portfolio}}^{\mathbb{Q}} = \left\{ \mu^{\mathbb{Q}}, \sigma, \hat{\mu}_u^{\mathbb{Q}}, \hat{\mu}_d^{\mathbb{Q}}, \nu_u^{\mathbb{Q}}, \nu_d^{\mathbb{Q}} \right\}$$

to the index portfolio of 125 CDSs spread curves in the CDX.IG index, by minimizing the relative root mean square CDS pricing error. Term structure of probability of default for each standard rating class² are given by

$$p_t^{\text{def}}(\bar{x}_i) = p_t(\bar{x}_i, \bar{x}; \Theta_{\text{Index Portfolio}}^{\mathbb{Q}}) \quad (4.12)$$

The dataset of 125 CDSs used in our calibration provides CDS spreads at maturities: 1y, 2y, 3y, 4y, 5y, 7y, 10y. Once the set $\Theta_{\text{Index Portfolio}}^{\mathbb{Q}}$ is found we can estimate further the implied starting DTD and produce the levels $\bar{x}_i^q$. In fact although the 125 CDS names have a rating attached, the representative starting rating level $\bar{x}_i$ for the i-th rating category cannot be sufficient to match the exact implied default probability from the credit spreads. In this way we are able to produce an implied DTD representative of the exact risk neutral creditworthiness of each name, $\bar{x}_k^q$, $k = 1, \dots, 125$. The results for the 125 CDSs in the CDX.IG index as on May 15, 2008 are reported in fig.(4.4).

²This means that the starting DTD level is $\bar{x}_i \in I_i$ for the i-th rating class.

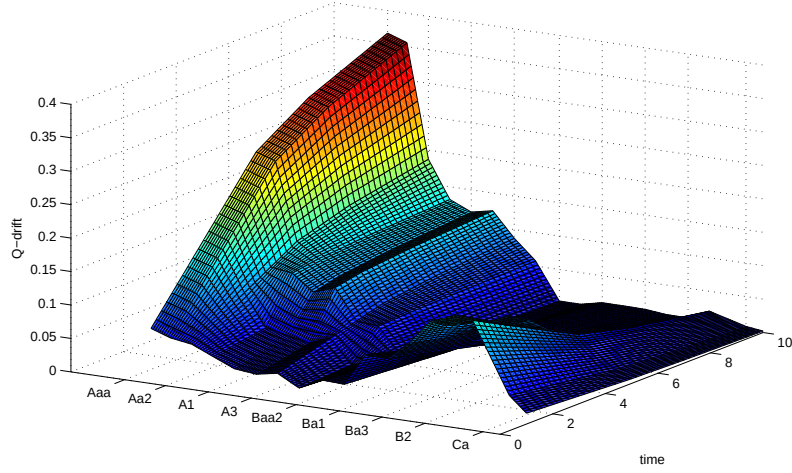


Figure 4.2: The above surface represents the calibrated risk neutral drift, across time and creditworthiness, to quoted CDO Index tranche prices as on May 15, 2008. The drift tends to increase over time for investment-grade names while the opposite trend is observed for non investment-grade names. The drift value toward the default state converges to zero.

The parameters set $\Theta_{\text{Index Portfolio}}^{\mathbb{Q}}$ can be further adjusted to match the index basis (see appendix 4.4.2 for background on index basis and notation used in the rest of this section) as reported in next paragraph.

Index basis calibration Let be $I(t)$ the level of the credit index at time $t \geq 0$. Recalling the definition of fair index spread as the quantity

$$s_F(t) = \frac{\sum_{k=1}^M s_k(t) PV01_k(t)}{\sum_{k=1}^M PV01_k(t)} \quad (4.13)$$

we account in our model for the Index basis

$$b(t) = I(t) - s_F(t), \quad t \geq 0$$

by means of a homogeneous local drift multiplier, $\alpha(X_t) \in \mathbb{R}^{++}$, such that the risk-neutral DTD process can price index tranches. This will affect the market price of the firm specific risk as following

$$\eta^{\text{Diff}, \alpha}(X_t) = \alpha \mu_{\mathbb{U}}^{\mathbb{Q}} - \mu_{\mathbb{U}} \quad (4.14)$$

2. Secondly, we fit the parameters subset

$$\Theta_{\text{Index Tranches}}^{\mathbb{Q}} = \{\mathbf{q}, \beta\}$$

to the quoted index tranches prices, by minimizing the relative weighed root mean square tranche pricing error. In fact once computed the basis-adjusted set $\Theta_{\text{Index Portfolio}}^{\mathbb{Q}}$

able to match single name CDS spreads and Index spread quotes, we can proceed to the calculation of the standardized index tranche premia. At this purpose in the following paragraphs, we specify the behavior of the recovery rate at default and of the portfolio correlation, both key parameters when pricing index tranches.

(a) Market status dependent Recovery Rates

The recovery rate plays an important role when pricing a single name and has an significant impact on the loss distribution in a portfolio of credit derivatives. It is possible in our framework to relax the assumption of a constant level. Using the notation consistently with section 4.2.1, we define a market factor dependent recovery $R(X^{(2)})$ such that

$$R(X^{(2)}) = \sum_k R(X^{(2)} = s_k)P(X^{(2)} = s_k), \quad k = 1, \dots, n \quad (4.15)$$

However the quoted recovery rate is usually 40%, and in the proposed calibration example we will use this value when calibrating the model to the market.

(b) Dynamic Correlation

The correlation structure among single names, introduced in section 4.2.1, is a deterministic function

$$\beta := \beta(x, t) : [\underline{x}, \bar{x}] \times [0, T] \mapsto [0, 1]$$

β is linearly interpolated across the two dimensions, namely the distance to default and time. This behavior reflects the fact that the correlation is not homogeneous across level of creditworthiness but tends to assume local distinct values that lead to the phenomenon of ‘defaults clustering’. We define a correlation surface across time and DTD that takes the following form

$$\beta(x, t) = \bigcup_{i=1}^3 \left(\bigcup_{j=2}^5 \rho_j \mathbf{1}_{\{x \in \chi(t)_j\}} \right) \mathbf{1}_{\{t \in \Delta t_i\}} \quad \bigcup_j \chi(t)_j = \bar{\chi}(t). \quad (4.16)$$

with $\bar{\chi}(t) = [\underline{x}, \underline{x} + h(t), \underline{x} + 2h(t), \dots, \bar{x} - h(t), \bar{x}]$, $h(t) \in \mathbb{R}^{++}$ and $\Delta t_1 = [0, 5yr]$, $\Delta t_2 = [5yr, 7yr]$, $\Delta t_3 = [7yr, 10yr]$.

There are 5 correlation levels ρ_j and 1 unit spacing $\alpha(t)$ for each of the 2 time intervals Δt . The correlation surface is fully specified by 18 parameters

$$\Theta^\beta = \{\rho, \mathbf{h}\} \quad (4.17)$$

The set $\mathbf{q}$ specifies the probability masses associated to the market factor in a particular scenario. The number of market states used in our numerical example is 5 and therefore $\mathbf{q}$ will be a set of 5 probability masses that sum up to 1, as described in section 4.2.1.

Tranche quotes calibration Let $s_{tr}(k, t, T)^z$ for $z \in \{\text{mo}, \text{mk}\}$ denote the spread at time t of the k -tranche with maturity T (usually 5, 7 or 10 years) as implied by the model (mo), and as quoted in the market (mk). The calibration routine consists in minimizing

the mean square tranche pricing error with respect to the only free model parameter left, $\beta(x, t)$,

$$\min_{\beta(x, s)} \sum_{t=1}^T (s_{tr}(k, t, T)^{mo})^\xi \sqrt{\left(\sum_{k=1}^K s_{tr}(k, t, T)^{mo} - s_{tr}(k, t, T)^{mk} \right)^2} \quad \xi < 0. \quad (4.18)$$

The quoted index tranche spreads term structure is calibrated in aggregate by minimizing the objective function 4.18.

Change of measure

Once calculated the parameter sets $\Theta^{\mathbb{P}}$ and $\Theta^{\mathbb{Q}}$ it is possible to make additional considerations on the measure change and on the risk premia produced by our model. The change between the physical measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$, can be characterized by the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$. Closely following Eckner [2007], we define the following change of measure process $(m_t)_{\{t \geq 0\}}$ and its independent components:

$$m_t = \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}_t\right) \quad (4.19)$$

$$= \mathbb{E}\left(\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} \frac{d\hat{\mathbb{Q}}}{d\tilde{\mathbb{Q}}} \frac{d\mathbb{Q}}{d\hat{\mathbb{Q}}} \middle| \mathcal{F}_t\right) \quad (4.20)$$

$$= m_t^{(1)} m_t^{(2)} m_t^{(3)}. \quad (4.21)$$

The first factor $m_t^{(1)}$ reflects the change in drift, while the other two factors $m_t^{(2)}$ and $m_t^{(3)}$ represent the change in the jump intensity (size (JS) and arrival (JA) parameters) and change in correlations under $\mathbb{Q}$ respectively. We can define the risk premia accordingly,

$$\eta^{\text{Market}}(X_t) = \mu_{\mathbb{U}}^{\mathbb{Q}} - \mu_{\mathbb{U}} \quad (4.22)$$

$$\eta^{\text{JA}}(X_t) = \frac{\gamma_{u/d}^{\mathbb{Q}}}{\gamma_{u/d}} \quad \eta^{\text{JS}}(X_t) = \frac{\lambda_{u/d}^{\mathbb{Q}}}{\lambda_{u/d}} \quad (4.23)$$

$$\eta^{\text{Corr}}(X_t) = \frac{\beta^{\mathbb{Q}}}{\beta} \quad (4.24)$$

This very similar characterization of the change of measure $d\mathbb{Q}/d\mathbb{P}$ in terms of risk premia in the context of CDO index tranches pricing has been proposed among others by Eckner [2007]. In appendix 4.4.4 we report details of the change of measure for a VG process and through the reported calculations it is possible to explicitly write the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$ in terms of the risk premia of equations (4.22), (4.23) and (4.24). In figure 4.3 we represent the pure jump-to-default risk premium produced by our model.

4.3.1 Price sensitivities

The most common approach for hedging CDO tranches is to hedge against small changes in the single-name CDS spreads. In our framework the computation of the hedging position in CDS i , known as spread-delta, is equal to

$$\phi_t^{xi} = \frac{\partial_{x_i} s^{[a,d]}(t, X_t)}{\partial_{x_i} s^i(t, X_t)} \quad (4.25)$$

where $\partial_{x_i} s^{[a,d]}(t, X_t)$ and $\partial_{x_i} s^i(t, X_t)$ are the partial derivatives with respect to x_i , a single name CDS, which can be approximated by finite differences. There is no recalibration involved when computing (4.25). It is possible to delta hedge also against global movements in the CDS spreads and take index spread-delta position in the index, that is computed in our framework as

$$\phi_t^{\text{Idx}} = \frac{\partial_x^2 s^{[a,d]}(t, X_t)}{\partial_x s^{\text{Idx}}(t, X_t)} \quad (4.26)$$

where $\partial_x s^{[a,d]}(t, X_t)$ and $\partial_x s^{\text{Idx}}(t, X_t)$ are the partial derivatives with respect to all CDSs. Computation of the index spread-delta is the same as the spread-deltas by finite differences, except that we need to shift all CDS spreads by 1bps.

Regarding the delta hedging of CDO in the base correlation framework, see Cousin and Laurent [2008] we can furthermore define two common approaches that can be denoted as ‘sticky strike’ and ‘sticky delta’ to parallel the terminology used in equity derivatives markets. In the sticky strike approach, the base correlations are kept unchanged when bumping the credit curves. When computing ‘sticky deltas’, one takes into account the change in base correlations due to the change in the moneyness of the tranche when credit spreads move up: the equity tranche becomes more junior, which actually leads to using a smaller base correlation. In other words, in the sticky delta approach an increase in credit spreads is associated with a smaller dependence between default events. As a consequence, the sticky delta of an equity tranche is lower than the delta computed under the sticky strike approach. In our example, see fig.4.5, the equity tranche deltas are slightly larger than those computed under the Gaussian copula, as in a ‘sticky delta’ approach. Similar results are produced also by Eckner [2007]. Such a result is consistent with a market where an increase in the average credit spread is the outcome of some idiosyncratic shifts and an increase in the dispersion of credit spreads. This is typical of the correlation crisis in May 2005, which was actually associated with smaller correlations on the equity tranches.

In the same manner it is possible to compute the second order sensitivities. We refer to idiosyncratic Gamma (iGamma) the convexity to single names, namely

$$\Gamma_t^{x_i} = \frac{\partial_{x_i} s^{[a,d]}(t, X_t)}{\partial_{x_i}^2 s^i(t, X_t)} \quad (4.27)$$

or to the market gamma (mGamma) the convexity to market moves and is equal to

$$\Gamma_t^{\text{Idx}} = \frac{\partial_x^2 s^{[a,d]}(t, X_t)}{\partial_x^2 s^{\text{Idx}}(t, X_t)} \quad (4.28)$$

All sensitivities computed in our framework do not require recalibration and their value reflect an adjusted spot DTD for the names underlying the credit portfolio.

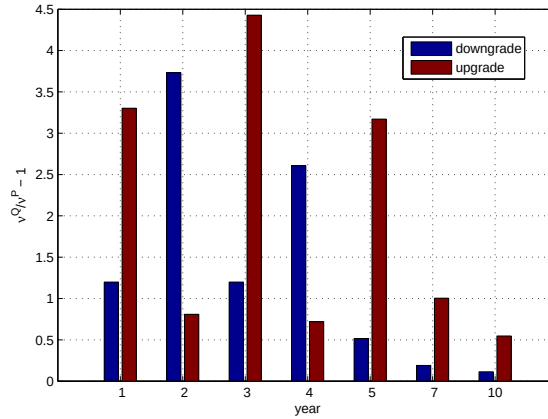


Figure 4.3: The bars represent the quantity $(\frac{\lambda^Q}{\lambda^P} - 1)$. The blue bars can be referred as pure jump-to-default risk premium, and are calibrated to quoted CDO Index tranche prices as on May 15, 2008. The behavior of these premiums in time is very intuitive. They are higher in the first five years and decrease afterwards. The jump to default is perceived and priced in the market as an event more likely to happen within few years of the trade date.

4.4 Model performance

In this section we examine in details how the proposed model calibrates to quoted tranche prices and generates risk sensitivities by a direct comparison with the gaussian copula model. We chose the CDX.IG index of May 15, 2008 when the market exhibited high volatility and high spread levels.

The market quotes are as it follows:

CDX.IG Market Quotes as 15 May 2008						
Mat.	5 Years		7 Years		10 Years	
Index Level (bps)	93		91		91	
Det.	Spread	Corr	Spread	Corr	Spread	Corr
3%	48.0%	36.4%	54.5%	33.4%	58.8%	33.0%
7%	375	56.1%	455	50.1%	545	46.3%
10%	193	64.0%	220	57.6%	271	53.1%
15%	97	75.4%	118.5	68.6%	133	64.4%
30%	48.5	96.5%	51.5	91.9%	61	89.1%

Table 4.1: CDX.IG Market quotes on May 15, 2008, as from Markit. The quotes include implied tranche base correlation. The equity tranche is quoted as upfront percentage.

The following table reports the calibrated implied base correlations and the index spread deltas ϕ_t^{Idx} .

The following table reports the calibrated index levels, index-tranches spreads levels and the index spread deltas ϕ_t^{Idx} generated with the proposed model.

Calibration Results for CDX.IG on 15 May 2008									
Mat.	5 Years			7 Years			10 Years		
Index Level	95.92 (3.05%)			92.3 (0.69%)			95.93 (3.06%)		
Tranche	Spread	error (%)	ϕ_t^{Idx}	Spread	error (%)	ϕ_t^{Idx}	Spread	error (%)	ϕ_t^{Idx}
0-3%	48.16%	.33	38.57	54.48%	.03	42.3	59.1%	0.4	47.58
3-7%	369	1.54	4.77	461.17	1.34	5.42	569.1	4.23	8.17
7-10%	205	6.18	2.79	218.3	0.78	4.58	257.29	5.33	4.92
10-15%	107.6	9.40	2.07	126.43	6.27	2.85	151.5	12.26	3.34
15-30%	50.13	3.25	1.00	50.8	1.38	1.47	63.8	4.39	1.86

Table 4.3: Calibration results for CDX.IG Index, May 15, 2008. The model is able to perfectly replicate the quoted market spread levels. The maximum spread fitting error of the model is produced in correspondence of the 10yr-tranche 10 – 15%, and is of 18.5bps.

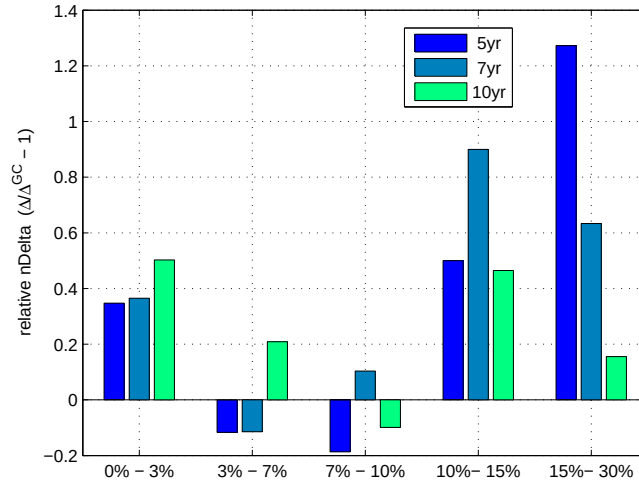


Figure 4.5: The DV01 results are very meaningful. They are in line with the market and the mezzanine delta shows a higher level due to the presence of high correlation in a group of names in the portfolio. The correlation clustering generates a multimodal loss distributions that is the main cause of this observed behavior for the mezzanine tranche delta.

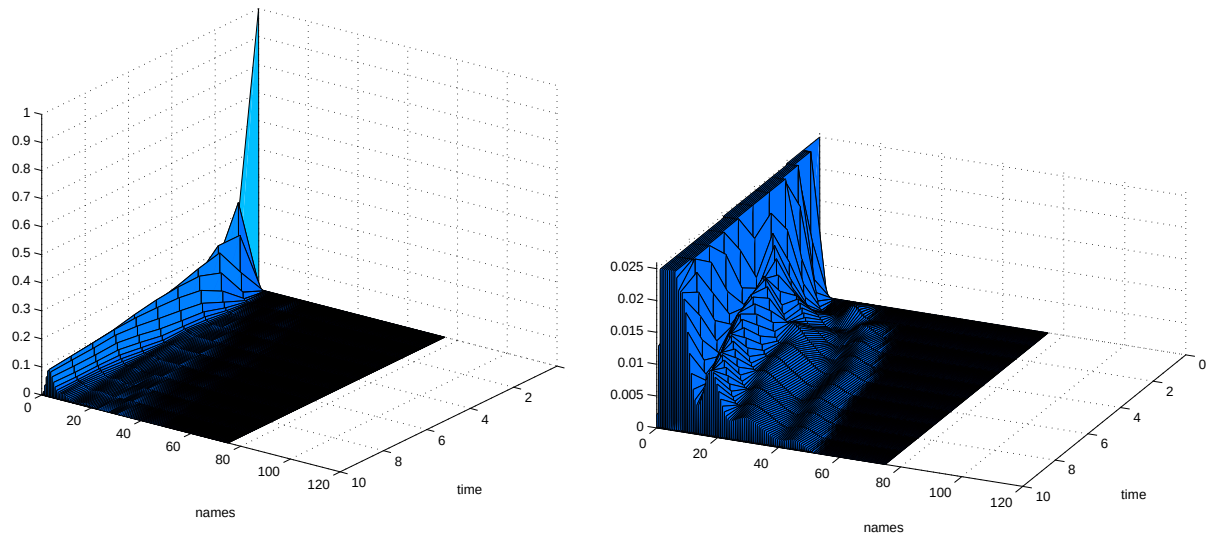


Figure 4.6: Portfolio default loss distribution for the names underlying the index CDX.IG on 15 May 2008. The right plot is a magnification of the loss distribution.

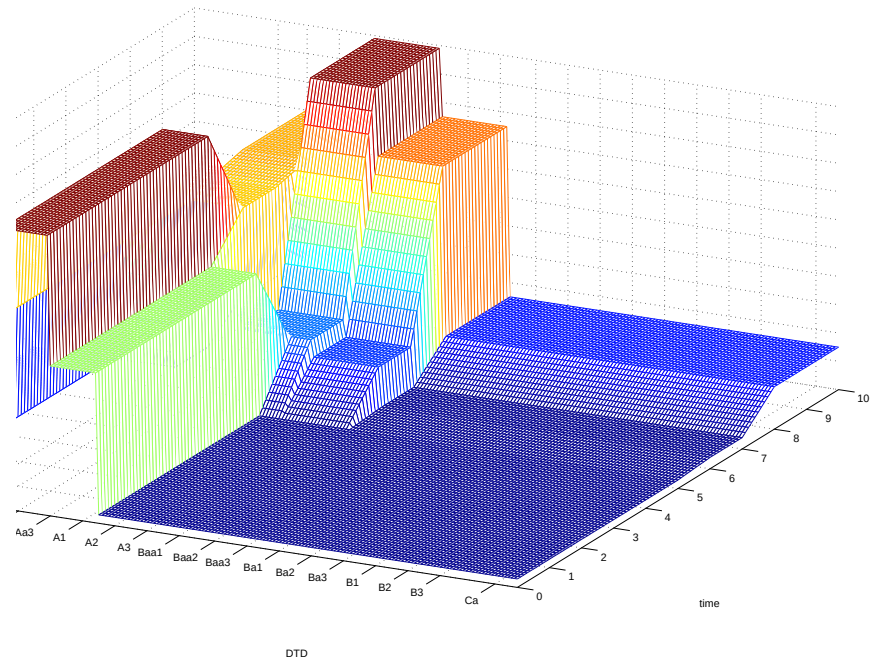


Figure 4.7: Calibrated parametric correlation surface as in eq.(4.16) for CDX.IG on 15 May 2008. The correlation levels modulate the portfolio names concentration producing clusters that are local across time and creditworthiness. Such modulation reflects the portfolio composition reported in fig. 4.4 in an intuitive and meaningful way. After the 5-years point very few names with non investment-grade implied creditworthiness cluster and drive the tranche prices toward the quotes observed in the market. The investment grade names seem overall consistently clustered and their correlation slightly decreases after the 5-years maturity.

Notes

4.4.1 Credit default swap index tranches

This note introduces the mechanisms of a credit default swap index (CDX) and of the CDX tranches. These are very standard concepts, see Markit.

A credit default swap index (denoted by CDX) contract provides protection against the credit risk of a standardized basket of reference entities. The mechanics of a CDX are slightly different from that of a single-name CDS. If a credit event occurs, the swap premium payment ceases in the case of a single-name CDS. In contrast, for a CDX the swap premium payment continues to be made by the protection buyer but based on a reduced notional amount since less reference entities are being protected. As of this writing, the settlement for a CDX is physical settlement. Currently, there are two families of standardized indexes: the Dow Jones CDX and the International Index Company iTraxx. The former includes reference entities in North America and emerging markets, while the latter includes reference entities in Europe and Asia markets. Both families of indexes are standardized in terms of the index composition procedure, premium payment, and maturity.

The two most actively traded indexes are the Dow Jones CDX NA IG index and the iTraxx Europe index. The former includes 125 North American investment-grade companies. The latter includes 125 European investment-grade companies. For both indexes, each company is equally weighted. Also for these two indexes, CDX contracts with 3-, 5-, 7- and 10-year maturities are available. The composition of reference entities included in a CDX are renewed every six months based on the vote of participating dealers. The start date of a new version index is referred to as the roll date. The roll date is March 20 and September 20 of a calendar year or the following business days if these days are not business days. A new version index will be 'on-the-run' for the next six months. The composition of each version of a CDX remains static in its lifetime if no default occurs to the underlying reference entities, and the defaulted reference entities are eliminated from the index. There are two kinds of contracts on CDXs: unfunded and funded. An unfunded contract is a CDS on a portfolio of names. This kind of contract is traded on all the Dow Jones CDX and the iTraxx indexes. For some CDXs such as the Dow Jones CDX.NA.HY index and its sub-indexes and the iTraxx Europe index, the funded contract is traded. A funded contract is a credit-linked note (CLN), allowing investors who because of client imposed or regulatory restrictions are not permitted to invest in derivatives to gain risk exposure to the CDX market. The funded contract works like a corporate bond with some slight differences. A corporate bond ceases when a default occurs to the reference entity. If a default occurs to a reference entity in an index, the reference entity is removed from the index (and also from the funded contract). The funded contract continues with a reduced notional principal for the surviving reference entities in the index. Unlike the unfunded contract which uses physical settlement, the settlement method for the funded contract is cash settlement. The index swap premium of a new version index is determined before the roll day and unchanged over its life time, which is referred to as the coupon or the deal spread. The price difference between the prevailing market spread and the deal spread is paid upfront. If the prevailing market spread is higher than the deal spread, the protection buyer pays the price difference to the protection seller. If the prevailing market spread is less than the deal spread, the protection seller pays the price difference to the protection buyer. The index premium

payments are standardized quarterly in arrears on the 20th of March, June, September, and December of each calendar year.

The CDXs have many attractive properties for investors. Compared with the single-name swaps, the CDXs have the advantages of diversification and efficiency. Compared with basket default swaps and collateralized debt obligations, the CDXs have the advantages of standardization and transparency. The CDXs are traded more actively than the single-name CDSs, with low bid-ask spreads.

CDS index tranches

With the innovation of CDXs, the synthetic CDO technology is applied to slice CDXs into standardized tranches with different subordinate levels to satisfy investors with different risk favorites. The tranches of an index provide the layer protections to the underlying portfolio risk. Each tranche has an attachment percentage and a detachment percentage. When the cumulative percentage loss of the portfolio reaches the attachment percentage, investors in the tranche start to lose their principal, and when the cumulative percentage loss of principal reaches the detachment percentage, the investors in the tranche lose all their principal and no further loss can occur to them.

Both of the most actively traded indexes, the Dow Jones CDX.NA.IG and the iTraxx Europe, are sliced into five tranches: equity tranche, junior mezzanine tranche, senior mezzanine tranche, junior senior tranche, and super senior tranche. The standard tranche structure of the Dow Jones CDX.NA.IG is 0-3%, 3-7%, 7-10%, 10-15%, and 15-30%. The standard tranche structure of the iTraxx Europe is 0-3%, 3-6%, 6-9%, 9-12%, and 12-22%. For both indexes, the swap premium of the equity tranche is paid differently from the non-equity tranches. It includes two parts: (1) the upfront percentage payment and (2) the fixed 500 basis points premium per annual. The market quote is the upfront percentage payment. For all the nonequity tranches, the market quotes are the premium in basis points, paid quarterly in arrears. Just like the indexes, the premium payments for the tranches (with the exception of the upfront percentage payment of the equity tranche) are made on the 20th of March, June, September, and December of each calendar year.

Following the commonly accepted definition for a synthetic CDO, CDX tranches are not part of a synthetic CDO because they are not backed by a portfolio of bonds or CDSs. In addition, CDX tranches are unfunded and they are insurance contracts, while synthetic CDO tranches are funded and they are CLNs. However, the net cash flows of index tranches are the same as synthetic CDO tranches and these tranches can be priced the same way as a synthetic CDO.

4.4.2 Index Basis

We define **average** index spread level the quantity

$$s_A(t) = \frac{1}{M} \sum_{k=1}^M s_k(t) \quad (4.29)$$

Where $s_k(t)$ is the spread of the k-name at time t and M is the total names in the pool.

We define **fair** index spread level the quantity

$$s_F(t) = \frac{\sum_{k=1}^M s_k(t) PV01_k(t)}{\sum_{k=1}^M PV01_k(t)} \quad (4.30)$$

The Index **fair** spread $s_F(t)$ could differ significantly from the average index spread, and it is usually lower. This is only due to the difference in credit duration of each underlying CDS and its representation is in the form of weighted average of the CDS spreads $s_k(t)$ with their risky annuity $PV01_k(t)$. Eq.(4.30) provides the definition of the dispersion of each CDS spread with respect to the Index.

Remark 4.1 *When the $s_k(t)$, $k = 1, \dots, M$ increase, $s_A(t)$ increases but $s_F(t)$ decreases.*

From the **fair** index spread value it is possible to derive the **intrinsic** index spread $s_I(t)$ that is calculated using a flat curve instead of the spread curves of the index constituents.

$$(s_I(t) - S)PV01_I(t) = (s_F(t) - S) \frac{1}{M} \sum_{k=1}^M PV01_k(t)$$

where S is the fixed index coupon. Given the term structure of the CDSs in the pool we can calculate the levels for all the indices as in Fig.(4.8).

We define **index basis relative to the average** index the quantity

$$b_A(t) = s_m(t) - s_A(t) \quad (4.31)$$

where $s(m)(t)$ is the market quote for the index at time t .

Remark 4.2 $b_A(t)$ is usually negative.

We define **index basis** the quantity

$$b_I(t) = s_m(t) - s_I(t) \quad (4.32)$$

Remark 4.3 $b_I(t)$ is usually positive.

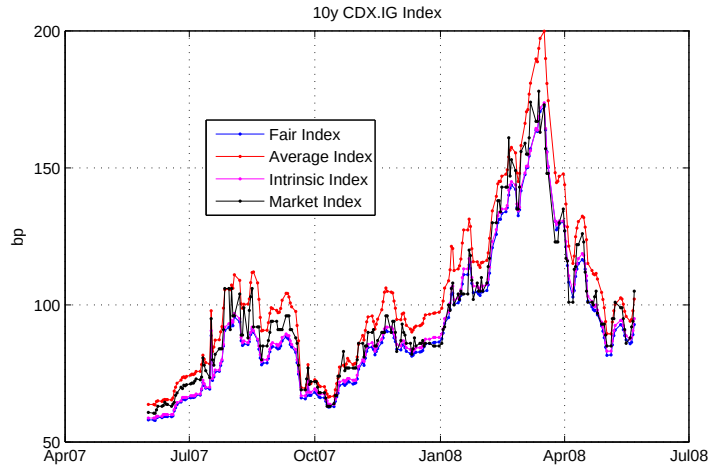


Figure 4.8: Comparison of the 10 years CDX.IG Indices. The average index has higher levels. Its departure from the other corresponding indices tends to increase with high volatility and consequently the ‘basis’ widens.

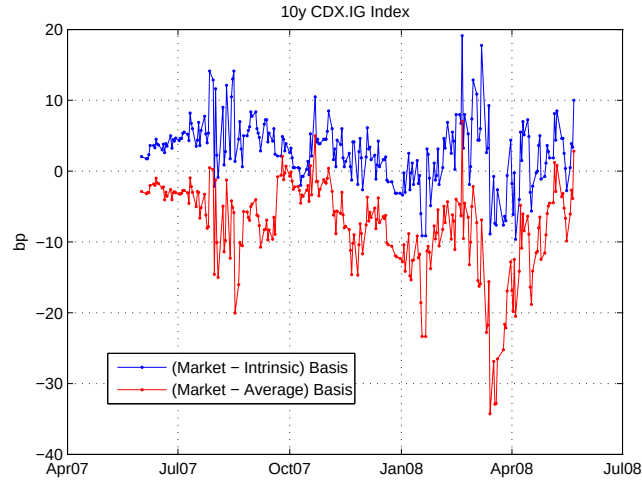


Figure 4.9: Basis for the 10 years CDX.IG Index. In March 2008 the basis reached its high.

4.4.3 Loss distribution and Index Tranches pricing

Lets M denote the number of obligors. The portfolio is represented by a vector of notionals w_m , and we assume they are normalized to 1, $\sum_{m=1}^N w_m = 1$. If name m defaults the portfolio is reduced by $w_m(1 - R_m)$, where R_m is the recovery rate of obligor m . The

cumulative loss is given by

$$L_t := \sum_{m=1}^M w_m (1 - R_m) \mathbf{1}_{\{\tau_m < t\}} \in [0, 1] \quad (4.33)$$

where τ_m is the default time of the m -th obligor. Given a tranche $[a, d]$ where a is the attachment and d is the detachment point in percentage, the loss of the tranche is defined as

$$\bar{L}_t^{[a,d]} := \min \left([\bar{L}_t - a]^+, d - a \right) \in [0, d - a] \quad (4.34)$$

The normalized tranche loss is

$$L_t^{[a,d]} = \frac{\bar{L}_t^{[a,d]}}{d - a} \in [0, 1] \quad (4.35)$$

and the corresponding normalized expected tranche loss is

$$l_t^{[a,d]} = \mathbb{E}[L_t^{[a,d]}]$$

Let $\phi_t(x) = \frac{\partial \bar{L}_t}{\partial x}$ be the loss density function. Then the expected cumulative tranche loss is

$$\bar{l}_t^{[a,d]} := \mathbb{E}[\bar{L}_t^{[a,d]}] = \int_0^1 \min \left([x - a]^+, d - a \right) \phi_t(x) dx \quad (4.36)$$

We can rewrite the above expression as

$$\begin{aligned} \bar{l}_t^{[a,d]} &= \int_0^1 \min \left([x - a]^+, d - a \right) \phi_t(x) dx \\ &= \int_0^1 \left(\min(x, d) - \min(x, a) \right) \phi_t(x) dx \\ &= \mathbb{E}[\bar{L}_t^{[0,d]}] - \mathbb{E}[\bar{L}_t^{[0,a]}] \\ &= \bar{l}_t^{[0,d]} - \bar{l}_t^{[0,a]} \end{aligned}$$

which implies that the cumulative loss up to a certain detachment point d_z can be expressed as

$$\bar{l}_t^{[0,d_z]} = \sum_{i=1}^z \bar{l}_t^{[d_{i-1}, d_i]} \quad (4.37)$$

for some given detachment points $a_i \quad i = 0, \dots, z$.

Loss computation

Given the independence of the unconditional default probabilities produced by the calibrated model, we use Fourier transform methods in order to build the unconditional loss distribution of the credit pool. The portfolio unconditional characteristic function for N names with conditional probability of default $p_k(X(t)|y; \beta(x))$, $k = 1, \dots, N$ is

$$\phi(t; u, \beta(x)) = \prod_{k=1}^N [1 + p_k(X(t)|y; \beta(x))(e^{-ius_k} - 1)] \quad (4.38)$$

where s_k is the normalized loss fraction for the k -th name taking into account the recovery $r_k(t)$ and $\beta(x)$ is the local correlation function previously introduced. The loss distribution at each time t may be calculated by numerical inverting the unconditional characteristic function using the Fractional Fourier transform Bailey and Swartztrauber [1991], Carr and Madan [1999].

$$L(x, t; \beta(x)) = \frac{1}{\pi} \int_0^\infty \text{Re}[e^{ixu} \phi(t; u, \beta(x))] du$$

Once the term structure of the Loss probability densities have been determined we can proceed to the index tranche prices evaluation using the following standard equations.

Expected cumulative loss properties

We compute the probability that the loss exceeds the detachment point, by differentiating $\bar{l}_t^{[0,d]}$, with respect to the detachment point d

$$\begin{aligned} \frac{\partial \bar{l}_t^{[0,d]}}{\partial d} &= \frac{\partial \int_0^1 \min(x, d) \phi_t(x) dx}{\partial d} \\ &= \int_d^1 \phi_t(x) dx \\ &= P_t(x > d) \end{aligned}$$

and for small dx

$$l_t^{[a, a+dx]} = \frac{\bar{l}_t^{[a, a+dx]}}{dx} \approx P_t(x > a) \quad (4.39)$$

By differentiating $\bar{l}_t^{[0,d]}$, twice with respect to the detachment point d , we obtain the loss distribution density $\phi_t(d)$

$$\phi_t(d) = -\frac{\partial^2 \bar{l}_t^{[0,d]}}{\partial d^2} \quad (4.40)$$

Thus in order to have a positive density and a monotonously increasing loss distribution in time we must ensure the conditions

$$\frac{\partial^2 \bar{l}_t^{[0,d]}}{\partial d^2} \leq 0 \quad (4.41)$$

$$\frac{\partial^2 \bar{l}_t^{[0,d]}}{\partial d \partial t} \geq 0 \quad (4.42)$$

The normalized expected tranche loss is a function of the type

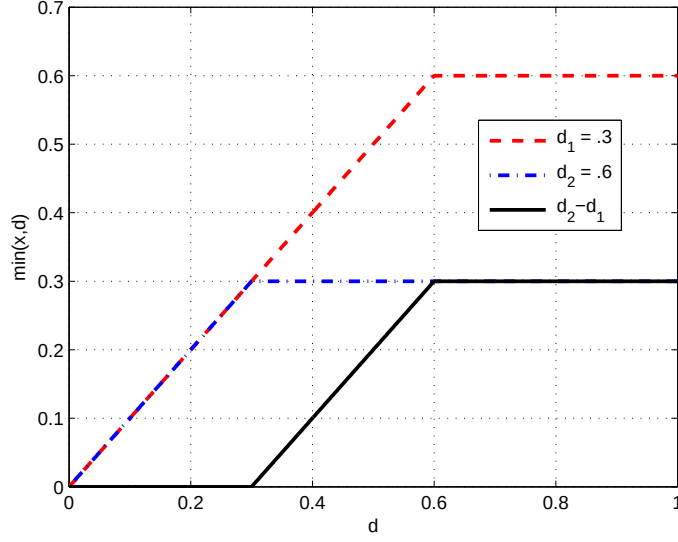
$$f(d, t) : [0, +\infty) \times [0, +\infty) \mapsto [0, 1]$$

and if $x \in [0, l_{max}]$ and $d \in [0, +\infty)$, although the maximum detachment point is usually set to the maximum possible loss

$$f(d, t) = \mathbb{E}[L_t^{[0,d]}] = l_t^{[0,d]} = \int_0^1 \min(x, d) \phi_t(x) dx \quad (4.43)$$

The ETL function $f(d, t)$ is

1. monotonically increasing with $f(0, t) = 0$ and $f(x, t) \leq t$

Figure 4.10: $\min(x, d)$ functions.

2. continuous
3. concave
4. constant outside the interval $[0, l_{max}]$

An ETL function $f(d, t)$ is right differentiable. The right derivative $d \rightarrow f'(d+, t)$ is monotonically decreasing and right continuous, hence cadlag. For any $a < d$ we have

$$\frac{f(d, t) - f(a, t)}{d - a} \leq f'(a+, t) \leq 1 \quad (4.44)$$

hence $f(d, t) - f(a, t) \leq d - a$. This implies that $f(d, t)$ is absolutely continuous, and therefore differentiable everywhere.

CDO Loss and Tranche spread Let $L_t = (1 - R) \frac{n - \text{defaults}(t)}{M}$ be the loss of the pool of n defaulted names over a total number of M at time t . The tranching Loss at two points a and d with $0 \leq a \leq d \leq 1$ is as in Eq.(4.35)

$$L_t^{[a, d]} = \frac{1}{d - a} \left[(L_t - a) 1_{\{a < L_t \leq d\}} + (d - a) 1_{\{L_t > d\}} \right] \quad (4.45)$$

The tranche $[a, d]$ spread is given by

$$s^{[a, d]} = \frac{\mathbb{E} \left[\sum_{i=1}^N D(0, T_i) \left(L_t^{[a, d]} - L_{t-1}^{[a, d]} \right) \right]}{\mathbb{E} \left[\sum_{i=1}^N D(0, T_i) (T_i - T_{i-1}) \left(1 - \frac{L_t^{[a, d]} + L_{t-1}^{[a, d]}}{2} \right) \right]} \quad (4.46)$$

Therefore given the CDS spreads $s_t^{\text{cds}} = \mathbf{s}^{\text{cds}} = (s^1, s^2, \dots, s^{125})$, the unconditional default indicators $D_t = \mathbf{D}$ and the set of correlations $\rho_t = \rho$, the tranche- $[a, d]$ premium can be expressed as,

$$s_t^{[a, d]} = s^{[a, d]}(t, \mathbf{s}^{\text{cds}}, \mathbf{D}, \rho)$$

Tranche and Index prices as a function of the Expected tranche loss

The NPV of the premium leg and the default leg for the Index are respectively

$$PL^{\text{Idx}} = s^{\text{Idx}} \sum_{i=1}^N D(0, T_i)(T_i - T_{i-1}) \left(1 - \frac{\mathbb{E}[L_t] + \mathbb{E}[L_{t-1}]}{2(1-R)} \right) \quad (4.47)$$

$$DL^{\text{Idx}} = \sum_{i=1}^N D(0, T_i) \left(\mathbb{E}[L_t] - \mathbb{E}[L_{t-1}] \right) \quad (4.48)$$

In these equations the summations run over the payments dates. The NPV of the premium leg and the default leg for the a Tranche $[a, d]$ are respectively

$$PL^{[a,d]} = s^{[a,d]} \sum_{i=1}^N D(0, T_i)(T_i - T_{i-1}) \left(1 - \frac{\mathbb{E}[L_t^{[a,d]}] + \mathbb{E}[L_{t-1}^{[a,d]}]}{2} \right) \quad (4.49)$$

$$DL^{[a,d]} = \sum_{i=1}^N D(0, T_i) \left(\mathbb{E}[L_t^{[a,d]}] - \mathbb{E}[L_{t-1}^{[a,d]}] \right) \quad (4.50)$$

The tranche (usually the equity tranche $[0, 3\%]$) may be quoted with an upfront $U_{[a,d]}$, then

$$PL^{[a,d]} = U^{[a,d]} + .05 \sum_{i=1}^N (T_i - T_{i-1}) \left(1 - \frac{\mathbb{E}[L_t^{[a,d]}] + \mathbb{E}[L_{t-1}^{[a,d]}]}{2} \right)$$

The expected tranche Loss is

$$\mathbb{E}[L_i^{[a,d]}] = \frac{\mathbb{E}[(\bar{L}_i - a)^+] - \mathbb{E}[(\bar{L}_i - d)^+]}{d - a}$$

The relation between ETL and EL is the following

$$\sum_{i=1}^k L_t^{[a_i, d_i]} (d_i - a_i) = \bar{L}_t \Rightarrow \sum_{i=1}^k \mathbb{E}[L_t^{[a_i, d_i]}] (d_i - a_i) = \mathbb{E}[\bar{L}_t] \quad (4.51)$$

The relation between ETL and Expected base tranche Loss is

$$\mathbb{E}[L_t^{[a,d]}] = \frac{d}{d-a} \mathbb{E}[L_t^{[0,d]}] - \frac{a}{d-a} \mathbb{E}[L_t^{[0,a]}] \quad (4.52)$$

Remark 4.4 Given t , Eq.(4.52) ensures consistency across ETL values. See above Expected cumulative loss properties section, in particular Eq.(4.37). This means that a requirement is that the ETL function $f(t, a_i, a_{i+1})$, $i = 1, 2, \dots, z$, with $a_1 = 0$ and $a_z = 100\%$ is a non-increasing function in a .

4.4.4 Change of measure for the VG process

In this appendix we examine the properties of the change of measure for the variance gamma (VG) process. We refer for background on Lévy processes to Sato [1999].

Without any loss of generality, the DTD process X in eq.(4.5) is a VG, characterized by the triplet (μ, σ, v) and can be expressed as

$$\begin{cases} X_t = X_0 e^{L_t} \\ L_t = mt + Z_t & m \in \mathbb{R} \\ Z_t = \mu G_t + \sigma W(G_t), & \mu \in \mathbb{R} \quad \sigma \in \mathbb{R}^{++} \end{cases} \quad (4.53)$$

where $G_t \sim \Gamma\left(t, \frac{\hat{\mu}^2}{\nu}, \frac{\nu}{\hat{\mu}}\right)$ is a gamma process with mean rate $\hat{\mu}$ and variance rate ν , with $\hat{\mu}, \nu \in \mathbb{R}^{++}$. In what follows we assume $\hat{\mu} = 1$, leading to $G_t \sim \Gamma\left(t, \frac{1}{\nu}, \nu\right)$. The VG process is fully specified by the parameters $X^{VG} := X^{VG}(t; m, \mu, \sigma, \nu)$. The VG process lacks of a diffusion component and is a pure jump process of infinite activity and finite variation and generally able with its dynamic to capture both frequent small changes and rare large jumps. The following result governs the change of measure for Lévy processes, see Sato [1999], Cont and Tankov [2004].

Proposition 4.1 *Let $(X_t, \mathbb{P})$ and $(X_t, \mathbb{Q})$ be two Lévy processes on $\mathbb{R}$ with characteristic triplets $(\mu^{\mathbb{P}}, \sigma^{\mathbb{P}^2}, v^{\mathbb{P}})$ and $(\mu^{\mathbb{Q}}, \sigma^{\mathbb{Q}^2}, v^{\mathbb{Q}})$. Then $\mathbb{P}$ and $\mathbb{Q}$ are equivalent for all $t > 0$, if and only if the following conditions are satisfied:*

1. $\sigma^{\mathbb{P}} = \sigma^{\mathbb{Q}}$
2. The Lévy measures are equivalent with

$$\int_{-\infty}^{\infty} (e^{\frac{\phi(x)}{2}} - 1)^2 v^{\mathbb{P}}(dx) < \infty \quad (4.54)$$

where $\phi(x) = \ln\left(\frac{dv^{\mathbb{Q}}}{dv^{\mathbb{P}}}\right)$

3. If $\sigma^{\mathbb{P}} = 0$, we must have

$$\mu^{\mathbb{Q}} - \mu^{\mathbb{P}} = \int_{-1}^1 x(v^{\mathbb{Q}} - v^{\mathbb{P}})dx \quad (4.55)$$

When $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, the Radon-Nikodym derivative is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{U_t} \quad (4.56)$$

$$U_t = \eta X_t^c - \frac{\eta^2 \sigma^{\mathbb{P}^2} t}{2} - \eta \mu^{\mathbb{P}} t + \lim_{\epsilon \downarrow 0} \left(\sum_{\substack{s \leq t \\ |\Delta X_s| > \epsilon}} \phi(\Delta X_s) - t \int_{|x| > \epsilon} (e^{\phi(x)} - 1)^2 v^{\mathbb{P}}(dx) \right) \quad (4.57)$$

Here (X_t^c) is the continuous part of (X_t) and η is such that

$$\mu^{\mathbb{Q}} - \mu^{\mathbb{P}} - \int_{-1}^1 x(v^{\mathbb{Q}} - v^{\mathbb{P}})dx = \sigma^{\mathbb{P}} \eta, \quad \text{if } \sigma^{\mathbb{P}} > 0 \quad (4.58)$$

U_t is a Lévy process with characteristic triplet (μ_U, σ_U, v_U) given by:

$$\mu_U = -\frac{1}{2}\sigma_U - \int_{-\infty}^{\infty} (e^x - 1 - x1_{|x| \leq 1})^2 v_U(dx) \quad (4.59)$$

$$\sigma_U = \eta^2 \sigma^{\mathbb{P}} \quad (4.60)$$

$$\nu_U = v^{\mathbb{P}} \phi^{-1} \quad (4.61)$$

In some but not all the circumstances we use the Esscher transform to specify the risk neutral equivalent martingale measure (EMM) $\mathbb{Q}$ for the DTD VG process. It is important to bear in mind that the credit market is incomplete due to the presence of jumps and therefore there are many equivalent pricing measures that can be specified using the result above.

Remark 4.5 *The use of the Esscher transform has the advantage to be analytically handy but defines a EMM with a specific form of the investors' preferences.*

Definition 4.2 (*Esscher transform*) *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable X , and a parameter h , the Esscher transform $\mathbb{P}^h$, sometimes also called Essher measure, is defined by*

$$d\mathbb{P}^h = \frac{e^{hX}\mathbb{P}}{\mathbb{E}[e^{hX}]} \quad (4.62)$$

provided the expectation exists.

This transformation depends on the parameter h and the random variable X . In the context of derivative pricing only one particular choice of the parameter h is of interest: the one such that the discounted asset price becomes a martingale under $\mathbb{P}^h$. For an introduction of the Esscher transform for Lévy processes see Cont and Tankov [2004]. We can then make proposition 4.1 specific to the case of the Esscher transform and the following theorem summarizes the result.

Theorem 4.2 *Suppose $T > 0$ and $h \in \mathbb{R}$ such that*

$$\mathbb{E}[e^{hX_1}] < \infty$$

Then

$$\begin{aligned} \frac{d\mathbb{P}^h}{d\mathbb{P}} &= e^{hX_t - \phi(h)T} \\ &= \frac{e^{hX_t}}{\mathbb{E}^h[e^{hX_t}]} \end{aligned}$$

defines a probability measure $\mathbb{P}^h$ such that $\mathbb{P}^h \sim \mathbb{P}$ and $\{X_t\}_{0 \leq t \leq T}$ is a Lévy process under $\mathbb{P}^h$ with triplet (μ^h, σ^h, ν^h) given by

$$\mu^h = \mu + h\sigma + \int_{-\infty}^{\infty} (e^{hx} - 1 - x1_{|x| \leq 1})^2 \nu(dx) \quad (4.63)$$

$$\sigma_h = \sigma \quad (4.64)$$

$$\nu_h(dx) = e^{hx} \nu(dx) \quad (4.65)$$

Proof See p. 685, Shiryaev [1999]. $\square$

In what follows we apply the result of theorem 4.2 to the specific case of VG. Considering the characteristic function of the process L in (4.5),

$$\phi_L(u, t) = e^{i\mu t} \left[\frac{1}{1 - iu\mu\nu + u^2 \frac{\sigma^2}{2}\nu} \right]^{\frac{t}{\nu}} \quad (4.66)$$

The VG process Z can be represented as the difference between two independent gamma processes, which follows from the fact

$$\phi_Z(u, t) = \left[\frac{1}{1 - iu\mu\nu + u^2 \frac{\sigma^2}{2}\nu} \right]^{\frac{t}{\nu}} = \left[\frac{b_+}{b_+ - iu} \right]^{\frac{t}{\nu}} \left[\frac{b_-}{b_- - iu} \right]^{\frac{t}{\nu}} \quad (4.67)$$

where

$$b_+ = \frac{2}{\nu \left(\sqrt{\mu^2 + \frac{2\sigma^2}{\nu}} + \mu \right)} \quad b_- = \frac{2}{\nu \left(\sqrt{\mu^2 + \frac{2\sigma^2}{\nu}} - \mu \right)} \quad (4.68)$$

Consequently the Lévy measure of the process $Z(t)$ is given by

$$\nu_Z(dx) = \frac{1}{\nu} |x|^{-1} (e^{-b_+ x} \mathbf{1}_{\{x>0\}} + e^{-b_- x} \mathbf{1}_{\{x<0\}}) \quad (4.69)$$

and the characteristic triplet of the process L is $(\mu, 0, \nu_Z)$. In virtue of theorem 4.2 and eq.(4.70), it follows that the characteristic function of the VG process under $\mathcal{P}^h$ is

$$\phi_L^h(u, t) = e^{i u m t} \left[\frac{1}{1 - i u \mu^h \nu^h + u^2 \frac{\sigma^2}{2} \nu^h} \right]^{\frac{t}{\nu}} \quad (4.70)$$

where

$$\mu^h = \mu + h\sigma^2 \quad (4.71)$$

$$\sigma^h = \sigma \quad (4.72)$$

$$\nu^h(dx) = \frac{\nu}{1 - h\mu\nu - h^2 \frac{\sigma^2}{2} \nu} \quad (4.73)$$

In order for the Esscher transform to specify a measure under which discounted asset prices behave like martingales, the Esscher parameter h needs to satisfy the following condition

$$r = \ln \phi_L \left(\frac{h+1}{i}, 1 \right) - \ln \phi_L \left(\frac{h}{i}, 1 \right)$$

Therefore the Esscher parameter h is the solution to the equation

$$f(h) = r - \mu - \frac{1}{\nu} \ln \frac{1 - h\mu\nu - h^2 \frac{\sigma^2}{2} \nu}{1 - (h+1)\mu\nu - (h+1)^2 \frac{\sigma^2}{2} \nu} = 0.$$

The risk neutral dynamics of the DTD process under $\mathbb{P}^h$, which in virtue of the Girsanov theorem is given by

$$X(t) = X(0) e^{(r - \ln \phi_Z^h(\frac{1}{i}, 1)t + Z^h(t))} \quad (4.74)$$

where $Z^h(t) \sim VG(\mu^h, \sigma, \frac{1}{\nu}, \nu^h)$.

Chapter 5

Conclusion

We have presented within this thesis properties and results on the continuous-time Markov chain approximation for multivariate diffusions and Lévy processes and their application to derivative pricing. We mainly investigated the problem of approximating the dynamics of correlated processes, that has been the dominant theme throughout this dissertation, by means of a discretization scheme based on continuous-time Markov chains.

In Chapter 3 we presented a method to approximate multidimensional correlated diffusions as continuous time Markov chains. The contributions have been twofold. Firstly the proposed methodology is new in the literature, applies to any multidimensional diffusion and can be seen as an alternative to Alternating Direction Implicit (ADI) schemes for the numerical solution of initial-boundary value problems for convection-diffusion equations with cross derivative terms. We illustrated desired convergence properties of such approximation and provided details of the algorithm based on tensor algebra. Secondly we provided guidance on an effective numerical implementation of our method and suggested approaches to compute the exponential of very large sparse matrices, that is a crucial aspect when finding a solution of the associated initial-boundary value problem.

The contribution of Chapter 4 has been twofold. First we designed and formalized a credit migration model under the physical measure, based on the concept of distance to default, along the lines of Albanese et al. [2005a] and Dalessandro [2010a]. Second, we introduced a joint model for the local migration process and a market factor. The model was able to dynamically condition through-the-cycle transition probabilities, and define an intuitive and flexible profile of point-in-time migration that can fully capture the historical market information. We formulated a calibration routine, specifically thought for our model, and conducted an empirical analysis using the historical average rating migration matrices for the Banking sector from Moody's 12. The calibration results showed that our model is able to consistently capture the point-in-time dynamics of the credit worthiness not just for default risk dynamics but for the entire transition among states of the rating migration matrix. This modeling feature is fundamental for an effective risk management of credit derivatives and credit risk portfolios conditionally on a state of the economy. We believe that the proposed model is very intuitive and showed how it can be an important tool for relative risk analysis in credit trading.

In Chapter 5 we presented a credit model for the dynamic pricing of correlated default baskets, that is an evolution of the dynamic model in Albanese et al. [2005a] and inherits great intuition from the framework defined in Dalessandro [2010b].

However the contributions in the proposed model are significative and multiple. We

characterize the representation of all the single credit names and their marginal default transition density as a one parametric time inhomogeneous Markov chain. The market factor is also represented by a Markov chain. Such a single factor induces a meaningful correlation and drives the default clustering phenomenon within the credit portfolio together with a market-status dependent recovery rate. This approach is remarkably more efficient, computational effective and more elegant than the conditional tree adopted in Albanese et al. [2005a,b]. Due to the discrete nature of our modeling approach, price sensitivities can be calculated without any additional computational effort.

The calibration results showed that our model is able to fit standard CDO index quotes, and produce meaningful price sensitivities. We believe that the proposed model is very intuitive, can be used to price and hedge both standard and exotic credit baskets whose values depend on the realized losses of a default portfolio and can represent an important tool in credit trading.

There are several natural open questions and extensions related to this thesis. First the approximation of correlated diffusions introduced in Chapter 3 can be investigated further both theoretically and experimentally. The method can furthermore be easily extended to the case of processes with jumps.

A second interesting aspect of investigation, would be the application of the concept of spectral distance between two matrices, introduced in Chapter 4, to the problem of maximum likelihood estimation defined through a continuous-time Markov chain approximation of stochastic processes.

The third important aspect of future research is the possibility of calibrate the dynamic credit model based on correlated Markov chain approximation, introduced in Chapter 5, to a larger dataset of CDO Index quotes and generate extensive empirical analysis and evidence on how accurate can be the hedging through time when performed with the proposed framework.

Appendix A

Appendices

A.1 Exponential jump diffusion process

We denote by $(S_t)_{t \geq 0}$ the asset price process and its dynamics is modelled using an exponential Lévy process of the type

$$S_t = S_0 e^{L_t}$$
$$L_t = \left(\mu - \frac{\sigma^2}{2} - \lambda\kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} Y_i$$

The fact that an asset price S_t is modelled as an exponential Lévy process L_t means that its log-return $\ln(\frac{S_t}{S_0})$ is modelled as a Lévy process such that:

$$\ln\left(\frac{S_t}{S_0}\right) = L_t = \left(\mu - \frac{\sigma^2}{2} - \lambda\kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} Y_i \quad (\text{A.1})$$

Let's go through the calculations of this model. Suppose in the small time interval dt the asset price jumps from S_t to $y_t S_t$, denoting by $\ln(y_t) \sim N(\kappa, \delta^2)$ the price jump size random variable. The relative jumps size is:

$$\frac{dS_t}{S_t} = \frac{y_t S_t - S_t}{S_t} = y_t - 1$$

and the SDE related to eq.(A.1) takes the form:

$$\frac{dS_t}{S_t} = (\mu - \lambda\kappa)dt + \sigma dB_t + (y_t - 1)dN_t \quad (\text{A.2})$$

The expected relative price change $\mathbb{E}[\frac{dS_t}{S_t}]$ from the jump part dN_t in the time interval dt is $\lambda\kappa dt$ since $\mathbb{E}[(y_t - 1)dN_t] = \mathbb{E}[y_t - 1]\mathbb{E}[dN_t] = \kappa\lambda dt$. this is a predictable part of the jump. This is the reason why the instantaneous expected return on the asset μdt is adjusted by $-\lambda\kappa dt$ in the drift term of the jump-diffusion process to make the jump part of an unpredictable innovation:

$$\mathbb{E}\left[\frac{dS_t}{S_t}\right] = \mathbb{E}[(\mu - \lambda\kappa)dt] + \mathbb{E}[\sigma dB_t] + \mathbb{E}[(y_t - 1)dN_t]$$
$$\mathbb{E}\left[\frac{dS_t}{S_t}\right] = (\mu - \lambda\kappa)dt + 0 + \lambda\kappa dt = \mu dt$$

If the asset price does not jump within the small time interval dt (i.e. $dN_t = 0$), then the jump diffusion process is simply a Brownian motion with drift process.

Let's solve the stochastic differential equation (A.2) using the Itô formula for jump-diffusion Schoutens [2005], Cont and Tankov [2004].

$$\begin{aligned}
d \ln(S_t) &= \frac{\partial \ln S_t}{\partial t} dt + (\mu - \lambda \kappa) S_t \frac{\partial \ln S_t}{\partial S_t} dt + \\
&\quad + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 \ln S_t}{\partial S_t^2} dt + \sigma S_t \frac{\partial \ln S_t}{\partial S_t} dB_t + [\ln y_t S_t - \ln S_t] \\
&= (\mu - \lambda \kappa) S_t \frac{1}{S_t} dt + \frac{\sigma^2 S_t^2}{2} \left(-\frac{1}{S_t^2}\right) dt + \sigma S_t \left(\frac{1}{S_t}\right) dB_t + [\ln y_t + \ln S_t - \ln S_t] \\
&= (\mu - \lambda \kappa) dt - \frac{\sigma^2}{2} dt + \sigma dB_t + \ln y_t \\
\ln S_t - \ln S_0 &= \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)(t - 0) + \sigma(B_t - B_0) + \sum_{i=1}^{N_t} \ln y_i \\
\ln S_t &= \ln S_0 + \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} \ln y_i \\
\exp(\ln S_t) &= \exp \left\{ \ln S_0 + \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} \ln y_i \right\} \\
S_t &= S_0 \exp \left\{ \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t \right\} + \exp \left(\sum_{i=1}^{N_t} \ln y_i \right) \\
&= S_0 \exp \left\{ \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t \right\} \prod_{i=1}^{N_t} y_i = S_0 \exp \left\{ \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} \ln y_i \right\}
\end{aligned}$$

Density function of a Jump diffusion process Given a jump-diffusive process as in eq. (A.3) where the log-return is a Lévy process

$$\ln \left(\frac{S_t}{S_0} \right) = L_t = \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t + \sigma B_t + \sum_{i=1}^{N_t} Y_i$$

where

$$\prod_{i=1}^{N_t} y_i = 1, \quad \sum_{i=1}^{N_t} Y_i = \sum_{i=1}^{N_t} \ln y_i = 0 \quad \text{if } N_t = 0.$$

In the Black-Scholes settings $\ln \frac{S_t}{S_0} \sim N[(\mu - \frac{\sigma^2}{2})t, \sigma^2 t]$, while in a jump-diffusive model returns are not log-normal. Assuming $Y_i \sim N(\kappa, \delta^2)$ the probability density of the log

returns $x_t = \ln(\frac{S_t}{S_0})$ is a quickly converging series of the form

$$\begin{aligned} P(x_t \in S) &= \sum_{k=0}^{\infty} P(N_t = k) P(x_t \in S | N_t = k) \\ P(x_t) &= \underbrace{\sum_{k=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!}}_{P_A} \underbrace{N\left(x_t; \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa\right)t, (\sigma^2 + \lambda \delta^2)t\right)}_{P_B} \end{aligned} \quad (\text{A.3})$$

where P_A is the probability of the asset price to jump k times during the interval t and P_B is the Black-Scholes normal density assuming that the asset price jumps k times in the interval t . The Model is then a weighted average of the Black-Scholes normal density by the probability that the asset jumps k times.

Derivation of the characteristic exponent We consider the characteristic function of the density function (A.3),

$$\begin{aligned} \phi(u) &= \int e^{iux} P(x) dx \\ &= \exp \left\{ \lambda t \exp \left(i\kappa u - u \frac{\delta^2}{2} \right) - \lambda t (1 + iu\kappa) - tu \left(-i\kappa + \frac{\sigma^2}{2} (i + u) \right) \right\} \\ &= \exp [t\psi(u)] \\ \psi(u) &= \lambda \left\{ \exp \left(iu\kappa - \frac{\delta^2 u^2}{2} \right) - 1 \right\} + iu \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa \right) - \frac{\sigma^2 u^2}{2} \end{aligned}$$

where $\kappa = E[y_t]$, being y_t the prize jumps size random variable.

Alternative derivation of the characteristic exponent Consider the Lévy measure of the jump-diffusive process

$$l(dx) = \frac{\lambda}{\delta \sqrt{2\pi}} \exp \left\{ -\frac{(dx - \kappa)^2}{2\delta^2} \right\} = \lambda f(dx) \quad f(dx) \sim N(\kappa, \delta^2) \quad (\text{A.4})$$

Substituting eq. (A.4) into the Lévy-Khinchine formula (of finite variation type, eq.(??)),

$$\begin{aligned} \phi(u) &= ibu - \frac{\sigma^2 \mu^2}{2} + \int (e^{iux} - 1) l(dx) \\ &= ibu - \frac{\sigma^2 u^2}{2} + \int (e^{iux} - 1) \lambda f(dx) \\ &= ibu - \frac{\sigma^2 u^2}{2} + \lambda \int (e^{iux} - 1) f(dx) \\ &= ibu - \frac{\sigma^2 u^2}{2} + \lambda \left\{ \int e^{iux} f(dx) - \int f(dx) \right\} \end{aligned}$$

where,

$$\int e^{iux} f(dx) = \exp \left\{ iu\kappa - \frac{\delta^2 u^2}{2} \right\}$$

hence,

$$\phi(u) = iu \left(\mu - \frac{\sigma^2}{2} - \lambda \kappa \right) - \frac{\sigma^2 u^2}{2} + \lambda \left\{ \exp \left\{ iu\kappa - \frac{\delta^2 u^2}{2} \right\} - 1 \right\}$$

A.2 Variance Gamma (VG) process

The Variance Gamma (VG) can be thought as a mixture of normal distributions, where the mixing weights density is given by the Gamma distribution of the subordinator. The VG distribution was introduced by Madan and Seneta [1990], Madan et al. [1998b]. We reformulate the model of the DTD in eq. (3.7) as

$$d \log X_t = r dt + \theta dg(t) + \alpha dW(g(t)), \quad X(0) = X_0 \quad (\text{A.5})$$

where r, θ and α are real constants and $\alpha \geq 0$. This model differs from the “usual” notation of GBM mainly in the term g_t . In fact we introduce g_t in order to characterize the market activity time. We can define the ‘market time’ as a positive increasing random process, $g(t)$, with stationary increments $g(u) - g(t)$ for $u \geq t \geq 0$. In probabilistic jargon $g(t)$ is a subordinator. An important assumption is that $E[g(u) - g(t)] = u - t$. This means that the market time has to reconcile with the calendar time between t and u on average. We also note that, conditional on the subordinator g , between two instants t and u :

$$\alpha(W(g(u)) - W(g(t)))|_g \sim \alpha\sqrt{g(u) - g(t)}\varepsilon, \quad \varepsilon \sim N(0, 1) \quad (\text{A.6})$$

VG assumes $g(t) \sim \Gamma(t/\nu, \nu)$, a Gamma process with parameter ν that is independent from the standard Brownian motion $\{W_t\}_{t>0}$. Notice also that the gamma process definition implies $g(u) - g(t) \sim (\frac{u-t}{\nu}, \nu)$ for any u and t . Increments are independent and stationary Gamma random variables. Consistently with the above condition on the expected subordinator, this Gamma assumption implies $E[g(u) - g(t)] = u - t$ and $Var(g(u) - g(t)) = \nu(u - t)$. Through iterated expectations, one has easily that

$$\mathbb{E}[\log(X(u)) - \log(X(t))] = (r + \theta)(u - t), \quad Var(\log X(u) - \log X(t)) = (\alpha^2 + \theta^2 \nu)(u - t). \quad (\text{A.7})$$

The transition density function for Eq.(A.5), where we set $u - t = \Delta t$ and $X_{\Delta t} = \log \frac{X(t+\Delta t)}{X(t)}$, is as following,

$$f_{X_{\Delta t}}(x) = \int_0^\infty f_N(x; x_0 + r\Delta t + \theta g, \alpha^2 g) f_\Gamma(g; \frac{\Delta t}{\nu}, \nu) dg \quad (\text{A.8})$$

The above integral converges and the PDF of the VG process $X_{\Delta t}$ is

$$f_{(X_{\Delta t})}(x) = \frac{2e^{\frac{\theta(x-\bar{\mu})}{\bar{\sigma}^2}}}{\bar{\sigma}\sqrt{2\pi\nu\frac{\Delta t}{\nu}}\Gamma(1/\nu)} \left(\frac{|x - \bar{\mu}|}{\sqrt{\frac{2\bar{\sigma}^2}{\nu} + \bar{\theta}^2}} \right)^{\Delta t/\nu - 1/2} K_{\Delta t/\nu - 1/2} \left(\frac{|x - \bar{\mu}|\sqrt{\frac{2\bar{\sigma}^2}{\nu} + \bar{\theta}^2}}{\bar{\sigma}^2} \right) \quad (\text{A.9})$$

Here $K_\eta(\cdot)$ is a modified Bessel function of the third kind with index η , given for $\omega > 0$ by

$$K_\eta(x) = \frac{1}{2} \int_0^\infty y^{\eta-1} \exp \left\{ -\frac{x}{2}(y^{-1} + y) \right\} dy$$

If $X \sim VG(\mu, \theta, \sigma, \nu)$ and $c > 0$, then

$$cX \sim VG(c\mu, \theta, c\sigma, c\nu) \quad (\text{A.10})$$

The class of VG distributions is closed under convolution in the following sense. If X_1 and X_2 are independent random variable that are VG distributed with $X_1 \sim VG(\mu_1, \theta_1, \sigma_1, \nu_1)$

```

1 function fx = VGdensity(x, theta, nu, sigma, mu, T )
2
3     v1 = 2*exp((theta*(x-mu))/sigma^2)/((nu^(T/nu))*...
4         sqrt(2*pi) * sigma * gamma(T/nu) );
5     M2 = (2*sigma^2)/nu + theta^2;
6     v3 = abs(x-mu)./sqrt(M2);
7     v4 = v3.^(T/nu - 0.5) ;
8     v6 = (abs(x-mu).*sqrt(M2))./sigma^2;
9     K = bessellk(T/nu - 0.5, v6);
10    fx = v1.*v4.*K;

```

Algorithm 2: *MATLAB*[®] code of the variance gamma density function as in equation (??).

and $X_2 \sim VG(\mu_2, \theta_2, \sigma_2, \nu_2)$, where $\sigma_1^2 \theta_1 = \sigma_2^2 \theta_2$ and $\frac{\mu_1}{\sigma_1^2} = \frac{\mu_2}{\sigma_2^2}$ then

$$X_1 + X_2 \sim VG\left(\mu_1 + \mu_2, \frac{\theta_1 \theta_2}{\theta_1 + \theta_2, \sqrt{\sigma_1^2 + \sigma_2^2}}, \nu_1 + \nu_2\right).$$

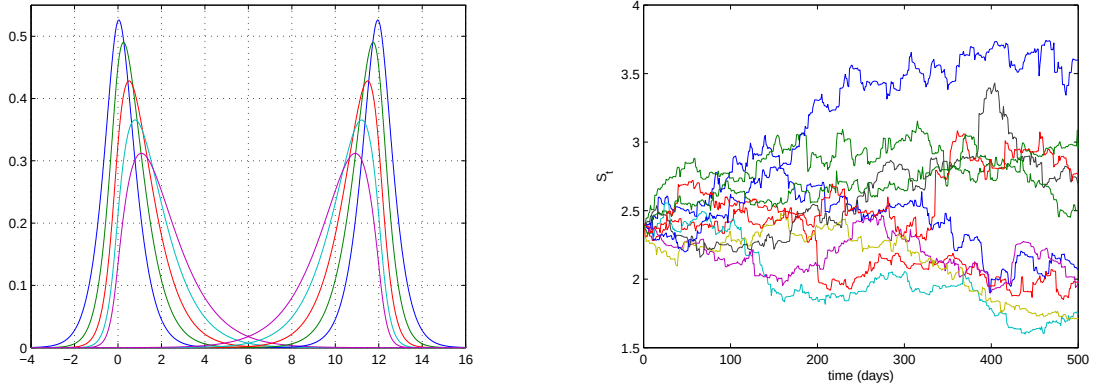


Figure A.1: On the left side, VG densities for $\nu = .4$, $\sigma = .9$, $\mu = 0$ and various values of of the parameter θ . On the right, simulated VG paths samples.

Given an historical time series of independent log-returns $(x_1, x_2, \dots, x_n)$ and denoting by $\pi = \{\mu, \theta, \sigma, \nu\}$ the variance-gamma probability density function parameters that we would like to fit, the maximum likelihood estimation (MLE) of π consists in finding $\hat{\pi}$ that maximize the the logarithm of the likelihood function $L(\pi) = \prod_{i=1}^n f(x_i|\pi)$.

If $X \sim VG(\mu, \theta, \sigma, \nu)$, the moment generating function of X , $E[e^{zX}]$ is given by

$$M_X(z) = e^{\mu z} \left(1 - \theta \nu z - \frac{1}{2} \nu \sigma^2 z^2\right)^{-\frac{1}{\nu}} \quad (\text{A.11})$$

and we obtain the four central moments of the return distribution over an interval of

length t :

$$\begin{aligned}
E[X] &= (\mu + \theta)t \\
E[(X - \bar{X})^2] &= (\nu\theta^2 + \sigma^2)t \\
E[(X - \bar{X})^3] &= (2\theta^3\nu^2 + 3\sigma^2\nu\theta)t \\
E[(X - \bar{X})^4] &= (3\nu\sigma^4 + 12\theta^2\sigma^2\nu^2 + 6\theta^4\nu^3)t + (3\sigma^4 + 6\theta^2\sigma^2\nu + 3\theta^4\nu^2)t^2.
\end{aligned}$$

Now considering the skewness and the kurtosis

$$\begin{aligned}
S &= \frac{(3\nu\theta)t}{((\nu\theta^2 + \sigma^2)t)^{3/2}} \\
K &= \frac{(3\nu\sigma^4 + 12\theta^2\sigma^2\nu^2 + 6\theta^4\nu^3)t + (3\sigma^4 + 6\theta^2\sigma^2\nu + 3\theta^4\nu^2)t^2}{((\nu\theta^2 + \sigma^2)t)^2}.
\end{aligned}$$

and ignoring the terms very small like θ^2 , θ^3 , θ^4 , we obtain

$$\begin{aligned}
S &= \frac{3\nu\theta}{\sigma\sqrt{t}} \\
K &= 3\left(1 + \frac{\nu}{t}\right).
\end{aligned}$$

So the initial guess for the parameters π_0 will be:

$$\begin{aligned}
\sigma &= \sqrt{\frac{V}{t}} \\
\nu &= \left(\frac{K}{3} - 1\right)t \\
\theta &= \frac{S\sigma\sqrt{t}}{3\nu} \\
\mu &= \frac{M}{t} - \theta
\end{aligned}$$

A.3 Differential Evolution

Differential Evolution (DE) is a method of mathematical optimization of multidimensional functions and belongs to the class of evolution strategy optimizers. DE finds the global minimum of a multidimensional, multimodal (i.e. exhibiting more than one minimum) function with good probability. The crucial idea behind DE is a scheme for generating trial parameter vectors. DE adds the weighted difference between two population vectors to a third vector. This way no separate probability distribution has to be used which makes the scheme completely self-organizing. Differential Evolution was first described by Price and Storn in the ICSI technical report Storn and Price [1995]. It is a global optimizer and was developed for continuous functions; as shown in algorithm (3), it assumes that NP solutions are stored in real-valued vectors of length D . In each generation, the algorithm creates a candidate solution for each existing solution $P_{:,i}^{(k)}$. This new solution is created by first adding the difference, weighted by a parameter F , between two other solutions to a third solution. Then an elementwise crossover takes place with probability CR between this ‘auxiliary’ solution $P_{:,i}^{(v)}$ and the existing solution $P_{:,i}^{(k)}$. If this final candidate solution $f(P_{:,i}^{(v)})$ is better than $P_{:,i}^{(k)}$, it replaces it; if not, the old solution is kept. We propose a version of the DE algorithm where the parameters F and CR are self adaptive. Given $0 \leq F_l \leq F_u \leq 1$ and $\tau_1, \tau_2 \in [0, 1]$, at every iteration the coefficients F and CR are updated as,

$$\begin{cases} F = F_l + u \times F_u & \text{if } u < \tau_1 \\ CR = u & \text{if } u < \tau_2 \end{cases} \quad (\text{A.12})$$

where $u \sim U(0, 1)$. In this thesis we propose a constrained, self adaptive version for DE. In fact the temporary vector solution $P_{:,i}^{(k)}$ is a ‘feasible’ solution, if can both minimize the objective function and be included within a given interval, $xMin \leq P_{:,i}^{(k)} \leq xMax$. In our DE algorithm we can also assign a starting population set in order to speed up the local search. This is particularly useful when the solution of the problem is almost known, or in MLE problems where it is possible to guess the order of magnitude or the unknown parameters. In the following we illustrate our DE algorithm with a pseudo-code first and then with its implementation in $F\#$, a multi-paradigm programming language, targeting the .NET Framework, that encompasses functional programming as well as imperative object-oriented programming disciplines. $F\#$ is a strongly typed language that uses type inference and is very indicate for fast prototyping and robust scientific applications.

Input: Objective function $f(x)$, D , NP , VTR, $xMin$, $xMax$, $MaxIters$, Strategy, starting x

Output: Global BestMember x , Best $f(x)$, number of iterations, F , CR

$\tau_1 := .05$, $\tau_2 := .05$, $F_l := .2$, $F_u := .8$, $CR := .85$, $F := .8$;

Starting population $P_{ji}^{(1)}$, $j = 1, \dots, D$, $i = 1, \dots, NP$, sampled using ‘starting x ’;

```

for  $k = 1 : MaxIters$  do
   $P^{(0)} := P^{(1)}$ ;
  for  $i = 1 : NP$  do
    if  $u < \tau_1$  then
       $F := F_l + u \times F_u$  ;
    end
    if  $u < \tau_2$  then
       $CR := u$  ;
    end
    Generate  $z_1, z_2, z_3 \in \{1, \dots, NP\}, i \neq z_1 \neq z_2 \neq z_3$  ;
    Given a ‘Strategy’, Calculate  $P_{\cdot,i}^{(v)} = P_{\cdot,z_1}^{(0)} + F \times (P_{\cdot,z_2}^{(0)} - P_{\cdot,z_3}^{(0)})$ ;
    for  $j = 1 : D$  do
      if  $u < CR$  then
         $P_{j,i}^{(u)} = P_{j,i}^{(v)}$ ;
      else
         $P_{j,i}^{(u)} = P_{j,i}^{(0)}$ ;
      end
    end
    if  $f(P_{\cdot,i}^{(u)}) < f(P_{\cdot,i}^{(0)}) \ \&\& \ xMin \leq P_{\cdot,i}^{(u)} \leq xMax$  then
       $P_{\cdot,i}^{(1)} = P_{\cdot,i}^{(u)}$ ;
    else
       $P_{\cdot,i}^{(1)} = P_{\cdot,i}^{(0)}$ ;
    end
  end
end
  
```

Algorithm 3: Differential Evolution. Self adaptive, constrained with starting sampling interval.

Remark A.1 *The proposed DE algorithm has been successfully used with all the applications and problems of this thesis that required optimization of parameters (MLE) or calibration to market data (volatility smile, matrix migration fitting, CDO tranches prices model calibration).*

```

1  (*)
2  © copyright Antonio Dalessandro 11/2008
3  Imperial College London
4  *)
5

```

```

6  #light
7
8  open Microsoft.FSharp.Math
9  open System
10 open System.Collections.Generic
11
12 let rand = System.Random()
13 let unitRand = rand.NextDouble
14
15 let randDouble min max = min + (max - min)*unitRand()
16
17 module Random =
18     let rand = new System.Random()
19     let int n = rand.Next(n)
20     let float x = x * rand.NextDouble()
21
22 let swap (a: 'a array) i j =
23     let t = a.[i]
24     a.[i] <- a.[j]
25     a.[j] <- t
26
27 let shuffle a =
28     let n = Array.length a
29     for i = 0 to n-1 do
30         swap a i (Random.int n)
31
32 let rem (a:int[]) b =
33     [|for i in 0 .. (Array.length a - 1)
34      -> a.[i] - (int(a.[i]/b))*b |]
35
36 let arraysum (a:int[]) (b:int[]) =
37     [|for i in 0 .. (Array.length a - 1)
38      -> a.[i] + b.[i]|]
39
40 [<AbstractClass>]
41 type BaseConstaints() =
42     abstract MyCon<'a> : 'a[]*('a[]*'a[]) -> bool
43
44 type DE_SearchGlobalMin( objF: float array -> float,
45                          D: int,
46                          NP: int,
47                          VTR:float,
48                          xMin: float,
49                          xMax: float,
50                          maxIters: int,
51                          strategy: int,
52                          startingArr: float array) =
53
54
55 let mutable CR = 0.85
56 let mutable F = 0.8
57
58 let tau1 = 0.05

```

```

59 let tau2 = 0.05
60 let Fl = 0.2
61 let Fu = 0.8
62
63 [<DefaultValue>]
64 val mutable cc:(float[]*float[])
65
66
67 //let mutable pop = Array.init NP (fun _ -> Array.init D (fun _ ->
68   randDouble xMin xMax))
69
70 let mutable pop = [| for j in 0 .. NP - 1 ->
71   startingArr |> Array.map (fun i -> i + randDouble
72     xMin xMax) |]
73
74 let mutable ui = Array.init NP (fun _ -> (Array.zeroCreate D: float array))
75 let mutable bm = Array.init NP (fun _ -> (Array.zeroCreate D: float array))
76 let mutable mui = Array.init NP (fun _ -> Array.init D (fun _ ->
77   randDouble 0. 1.))
78 let mutable mpo = Array.init NP (fun _ -> Array.init D (fun _ ->
79   randDouble 0. 1.))
80
81 // section of variables for population evolution
82 let indPop = [|0..NP-1|]
83 let indRotation = [|0..NP-1|]
84 let ind = [|1..4|]
85
86 //let mutable globalBestMember = Array.init D (fun _ -> randDouble xMin
87   xMax)
88 let mutable globalBestMember = startingArr |> Array.map (fun i -> i +
89   randDouble xMin xMax)
90 let mutable Valuation = Array.init NP (fun i -> randDouble xMin xMax)
91 let mutable BestValue = randDouble xMin xMax
92
93 let CC = {
94   new BaseConstraints() with
95     member x.MyCon (X , (low, high) ) =
96       let temp = [| for k in 0..(Array.length X -
97         1)->
98         ( X.[k] ≤high.[k] && X.[k] ≥
99           low.[k] )
100         |]
101       Array.forall (fun x -> x) temp
102   }
103
104 member this.Cons
105   with get() = this.cc
106   and set(v:float[]*float[]) = this.cc <-v
107
108 // evaluate the best member function after initialization
109 member this.updateGbest =
110   globalBestMember <- Array.minBy objF pop
111   BestValue <- objF globalBestMember

```

```

104 member this.prepareLearning =
105     this.updateGbest
106     for i = 0 to NP - 1 do
107         Valuation.[i] <- (objF pop.[i])
108
109 member this.DoLearning =
110     this.prepareLearning
111     //printfn "starting POPULATION: %A \n" pop
112     let mutable n = 1
113     while (BestValue ≥ VTR && n ≤ maxIters) do
114         //for n = 1 to maxIters do
115             // if (BestValue ≥ VTR && n ≤ maxIters) then
116                 this.SingleIteration n
117                 n <- n + 1
118     done
119     // else 0
120     //printfn "Final globalBestMember: %A" globalBestMember
121     //printfn "Final BestValue: %f" BestValue
122     //printfn "Iteration %d of %d" (n - 1) maxIters
123     (globalBestMember, BestValue, n-1, F, CR)
124
125 member this.SingleIteration z =
126
127     // printfn "iteration %d" z
128     //printfn "Actual Best Member: %A and Objective Function Value %f \n"
129         globalBestMember BestValue
130
131     shuffle ind
132     shuffle indPop
133     let a1 = Array.copy indPop
134     let rt = rem [|for i in 0..NP-1 -> i + ind.[0]|] NP
135     let a2 = [|for i in 0..NP-1 -> a1.[rt.[i]]|]
136     let rt = rem [|for i in 0..NP-1 -> i + ind.[1]|] NP
137     let a3 = [|for i in 0..NP-1 -> a2.[rt.[i]]|]
138     let rt = rem [|for i in 0..NP-1 -> i + ind.[2]|] NP
139     let a4 = [|for i in 0..NP-1 -> a3.[rt.[i]]|]
140     let rt = rem [|for i in 0..NP-1 -> i + ind.[3]|] NP
141     let a5 = [|for i in 0..NP-1 -> a4.[rt.[i]]|]
142
143     let pm1 = [|for i in 0..NP-1 -> pop.[a1.[i]]|] // population shuffling
144         for various strategies
145     let pm2 = [|for i in 0..NP-1 -> pop.[a2.[i]]|]
146     let pm3 = [|for i in 0..NP-1 -> pop.[a3.[i]]|]
147     let pm4 = [|for i in 0..NP-1 -> pop.[a4.[i]]|]
148     let pm5 = [|for i in 0..NP-1 -> pop.[a5.[i]]|]
149
150     if unitRand() < tau1 then F <- F1 + unitRand()*Fu
151     if unitRand() < tau2 then CR <- unitRand()
152
153     for npart = 0 to NP-1 do
154         bm.[npart] <- globalBestMember // population filled with the best
155             member of the last iteration
156     for dim = 0 to D-1 do
157         let r1 = unitRand()

```

```

154     if r1 < CR then mui.[npart].[dim] <- 1.
155     else mui.[npart].[dim] <- 0.
156     if mui.[npart].[dim] < 0.5 then mpo.[npart].[dim] <- 1.
157     else mpo.[npart].[dim] <- 0.
158     let uie =
159         match strategy with
160         | 1 -> bm.[npart].[dim] + F*(pm1.[npart].[dim] -
161             pm2.[npart].[dim])
162         | 2 -> pm3.[npart].[dim] + F*(pm1.[npart].[dim] -
163             pm2.[npart].[dim])
164         | 3 -> pop.[npart].[dim] + F*(bm.[npart].[dim] -
165             pop.[npart].[dim]) + F*(pm1.[npart].[dim] -
166             pm2.[npart].[dim])
167         | 4 -> bm.[npart].[dim] + F*(pm1.[npart].[dim] -
168             pm2.[npart].[dim] + pm3.[npart].[dim] - pm4.[npart].[dim])
169         | 5 -> pm5.[npart].[dim] + F*(pm1.[npart].[dim] -
170             pm2.[npart].[dim] + pm3.[npart].[dim] - pm4.[npart].[dim])
171         | _ -> failwith "error"
172
173     ui.[npart].[dim] <- pop.[npart].[dim]*mpo.[npart].[dim] +
174         uie*mui.[npart].[dim]
175
176     let cc = CC.MyCon(ui.[npart],this.cc) // constraint
177
178     if (objF ui.[npart]) < (objF pop.[npart])&& cc then
179         pop.[npart] <- Array.copy ui.[npart]
180
181     if (objF pop.[npart]) < BestValue then
182
183         BestValue <- (objF pop.[npart])
184         globalBestMember <- Array.copy pop.[npart]
185         //printfn " Best Member is updated %f at iteration %d "
186             globalBestMember.[0] z

```

The following example shows how simple is the use of our DE algorithm in finding the zeros of a polynomial. It takes less than a second for DE to find 8 roots.

```

1
2 let dime = 8
3 //let b = Array.init dime (fun _ -> floor(randDouble 2. 20.))
4
5 let b = [| 2. ; -15.; 200.; -3000. ; 1.83 ; -15.33; 177.22; -2999.875 |]
6
7 let objF (x:float[]) =
8     (Array.init (Array.length b) (fun i -> (x.[i] -
9         b.[i])*(x.[i] - b.[i])))
10     |> Array.sum
11
12 //let startArr = Array.create dime 1.
13 let startArr = [| 0. ; -18.; 180.; -2900. ; 1. ; -2. ; 88. ; -3111. |]
14 let de = new DE_SearchGlobalMin(objF, dime, 2*dime, 1e-10, -1., 1.,5000 ,2,
    startArr)

```

```
15 de.cc <- ( Array.create dime -5000.0, Array.create dime 500.0 )
16
17 let resultDE = de.DoLearning
18 printfn " solution ---> %A" resultDE
```


Bibliography

- M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel. Implementation of a restarted krylov subspace method for the evaluation of matrix functions. *Linear Algebra and its Applications*, 429:2293–2314, 2008.
- Y. Ait-Sahalia. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica*, pages 223–262, 2002.
- C. Albanese. Operator methods, abelian processes and dynamic conditioning. *Working Paper*, 2007.
- C. Albanese and O. Chen. Credit barrier models in a discrete framework. *Proceedings of the American Mathematical Society Snowbird conference*, 2003.
- C. Albanese and A. Mijatovic. Convergence rates for diffusions on continuous-time lattices. *Working Paper, Imperial College London*, pages 1–22, 2006a.
- C. Albanese and A. Mijatovic. Convergence rates for diffusions on continuous-time lattices. *Working Paper, Imperial College London*, pages 1–22, 2006b.
- C. Albanese, O. Chen, and A. Dalessandro. Dynamic credit correlation modeling. *Working Paper, Imperial College London*, pages 1–17, 2005a.
- C. Albanese, O. Chen, A. Dalessandro, and A. Vidler. Dynamic conditioning and credit correlation baskets. *Working Paper*, pages 1–31, 2005b.
- C. Albanese, H. Lo, and A. Mijatovic. Spectral methods for volatility derivatives. *Working Paper*, pages 1–40, 2009.
- A. Andersson and P. Vanini. Credit migration risk modeling. *Working Paper*, pages 1–33, 2009.
- D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics, 2004.
- D. H. Bailey and P. N. Swarztrauber. The Fractional Fourier Transform and Applications. *SIAM Review* 33, 33:389–404, 1991.
- B. S. Balakrishna. Lévy subordinator model of default dependency. *Working Paper*, pages 1–16, 2009.
- M. Baxter. Levy process dynamic modelling of single name credits and cdo tranches. *Nomura Interantional, Working Paper*, 2006.

- N. Bennani. A note on markov functional loss models. *Working Paper Dresdner Kleinwort*, pages 1–16, 2006.
- T. R. Bielecki, S. Crépey, and A. Herbertsson. *Markov Chain Models of Portfolio Credit Risk*. Working Paper, 2009.
- D. Brigo, A. Pallavicini, and R. Torresetti. Calibration of cdo tranches with the dynamical generalized-poisson loss model. *Working Paper Banca IMI*, pages 1–41, 2007a.
- D. Brigo, A. Pallavicini, and R. Torresetti. Default correlation, cluster dynamics and single names: The gpcl dynamical loss model. *International Journal of Theoretical and Applied Finance*, pages 1–35, 2007b.
- P. Carr and D. Madan. Option Valuation using the Fast Fourier Transform. *Journal of Computational Finance*, 3, 1999.
- A. Chapovsky, A. Rennie, and P. Tavares. Stochastic modeling for structured credit exotics. *Merril Lynch International, Working Paper*, 2006.
- R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman and Hall, Financial Mathematics Series, 2004.
- A. Cousin and J. P. Laurent. Hedging issues for cdos. *Chapter 17 in the book - The Definitive Guide to CDOs, edited by G. Meissner. Risk Books*, pages 461–480, 2008.
- A. Dalessandro. Correlated continuous time markov chains in multidimensional vector spaces with application to derivatives pricing. *Working Paper*, pages 1–37, 2010a.
- A. Dalessandro. Credit migration risk and point in time credit dynamics: A new perspective for credit risk management. *Working Paper*, pages 1–41, 2010b.
- A. Dalessandro. All your hedges in one matrix. *Submitted for publication*, pages 1–32, 2010c.
- A. Dalessandro and M. Gilli. Nonparametric estimation of credit migration matrices. *4th Conference on Computational Management Science*, pages 1–40, 2007.
- G. Digraziano and L.C.G. Rogers. A dynamic approach to the modelling of correlation credit derivatives using markov chains. *University of Cambridge, Working Paper*, 2005.
- D. Duffie and N. Garleanu. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57:1–46, 2001.
- D. Duffie and K.J. Singleton. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, 2003.
- Darrell Duffie, Leandro Saita, and Ke Wang. Multi-period corporate default prediction with stochastic covariates. *Working Paper*, pages 1–43, 2007.
- A. Eckner. Risk premia in structured credit derivatives. *Working Paper*, pages 1–48, 2007.
- Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, 1999.

- E. Errais, K. Giesecke, and L. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1:642–665, 2010.
- S.T. Ethier and T.G. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley, 1985.
- S. Fligewski, H. Frydman, and W. Liang. Modelling the effect of macroeconomic factors on corporate default and credit rating transitions. *Working Paper*, pages 1–53, 2006.
- R. A. Friesner, L. S. Tuckerman, B. C. Dornblaser, and T. V. Russo. A method for exponential propagation of large systems of stiff nonlinear differential equations. *Journal of Scientific Computing*, 4:327–354, 1989.
- E. Gallopoulos and Y. Saad. Efficient solution of parabolic equations by polynomial approximation methods. 13:1236–1264, 1992.
- G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London, 1996.
- P.R. Halmos. *Finite-Dimensional Vector Spaces*. Princeton University Press, 1947.
- S. Heston. A closed-form solution of option with stochastic volatility. *The Review of Financial Studies*, 6:327–343, 1993.
- J. Hull and A. White. Valuing credit default swaps I: No counterparty default risk. *Journal of Derivatives*, 8(1):29–40, Fall 2000.
- J. Hull and A. White. Dynamic models of portfolio credit risk: A simplified approach. *Working Paper University of Toronto*, pages 1–41, 2007.
- J. Hull and A. White. Dynamic models of portfolio credit risk: a simplified approach. *Journal of Derivatives*, 15(4):9–28, Summer 2008.
- R. B. Israel, J. S. Rosenthal, and J. Z. Wei. Finding generators for markov chains via empirical transition matrices, with application to credit ratings. *Mathematical Finance*, 11:245–265, 2001.
- R. A. Jarrow, D. Lando, and S. M. Turnbull. A markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10:481–523, 1997.
- M. Joshi and A. Stacey. Intensity gamma: A new approach to pricing portfolio credit derivatives. *Working Paper, Royal Bank of Scotland*, pages 1–22, 2005.
- I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer Verlag, 1991.
- S. J. Koopman, A. L., and A. Monteiro. The multi-state latent factor intensity model for credit rating transitions. *Journals of Econometrics*, pages 1–45, 2005.
- E. Kreyszig. *Introductory Functional Analysis with Application*. Wiley Classic Library, 1978.
- D. Lando. *Credit Risk Modeling: Theory and Applications*. Princeton Series in Finance, 2004.

- D. Lando and A. Mortensen. On the pricing of step-up bonds in the european telecom sector. *Journal of Credit Risk*, 1:71–110, 2005.
- D.X. Li. On default correlation: A copula function approach. *The RiskMetrics Group, Working Paper*, 99-07:1–28, 2000.
- M. L. Liou. A novel method of evaluating transient response. *Proceedings of the IEEE*, 54(1), 1966.
- H. Lo. *Markov chains and the pricing of derivatives*. PhD Thesis, Department of Mathematics, Imperial College London, 2009.
- E. Lukacs. *Characteristic Functions*. Charles Griffin & Company Limited london, 1958.
- R.E. Lynch, J.R. Rice, and D.H. Thomas. Tensor product analysis of partial difference equations. *Working Paper, Communication by A.S. Householder*, pages 378–384, 1963.
- D. Madan, P. Carr, and E. C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2:79–105, 1998a.
- D. B. Madan and E. Seneta. The variance gamma (V.G.) model for share market returns. *Journal of Business*, 63(4):511–524, 1990.
- D.B. Madan, P.P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 79:79–105, 1998b.
- Markit. Markit Credit Index definition and rules ,
<http://www.markit.com/en/products/data/indices/credit-and-loan-indices/cdx/cdx.page>.
- A. D. Metz. A cyclical model of multiple-horizon credit rating transitions and default. *Moody's Investors Service, Working Paper*, pages 1–53, 2007.
- A. Mijatović and M. Pistorious. Continuously monitored barrier options under markov processes. *Working Paper Imperial College London, to appear in Mathematical Finance*, pages 1–62, 2009.
- A.R. Mitchell and D.F. Griffiths. *The Finite Difference Methods in Partial Differential Equations*. Wiley, New York, 1980.
- C. Moler and C. F Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *Society for Industrial and Applied Mathematics*, 20(4): 801–836, 1978.
- A. Mortensen. Semi-analytical valuation of basket credit derivatives in intensity-based models. *Journal of Derivatives*, 13:8–26, 2006.
- B. Nour-Omid. Applications of the lanczos algorithm. *Computational Physics Communications*, 53, 1989.
- V. Putyatin, D. Prioul, and S. Maslova. A markovian approach to modelling correlated defaults. *Risk Magazine*, pages 1–22, 2005.
- L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 1 and 2. Cambridge, 2000.

- K. Podgórski S. Kotz, T. J. Kozubowski. *The Laplace Distribution and Generalizations*. Birkhäuser, 2001.
- Y. Saad. Analysis of some Krylov subspace approximations to the matrix exponential operator. (RIACS-90-ExpTh), 1990.
- K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.
- P. Schönbucher. *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. John Wiley and Sons Ltd, 2003.
- P. Schönbucher. Portfolio losses and the term structure of loss transition rates: a new methodology for pricing portfolio credit derivatives. *ETH Zurich, Working Paper*, 2005.
- W. Schoutens. *Lévy Processes in Finance*. Wiley, 2005.
- Til Schuermann and Yusuf Jafry. Measurement and estimation of credit migration matrices. *Working Paper*, pages 1–47, 1997a.
- Til Schuermann and Yusuf Jafry. Metrics for comparing credit migration matrices. *Wharton Financial Institutions Working Paper*, pages 1–32, 1997b.
- A. N. Shiryaev. *Essentials of stochastic finance*. World Scientific Publishing Co. Inc., 1999.
- S. E. Shreve. *Stochastic Calculus for Finance II, Continuous Time Models*. Springer, 1993.
- J. Sidenius, V. Piterbarg, and L. Andersen. A new framework for dynamic credit portfolio loss modeling. *Working Paper, Bank of America*, 2005.
- R. B. Sidje. Parallel algorithms for large sparse matrix exponentials: application to numerical transient analysis of markov processes. *Ph.D. Thesis, University of Rennes I*, 1994.
- M. Sofa. Convergence d'opérations lineaires non bornees. *Review Roumaine Math. Pure Appliq.*, 12:373–389, 1967.
- R. Storn and K. Price. Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces. *Technical Report TR-95-012, ICSI*, pages 1–22, 1995.
- P. Stumpp and M. Coppola. Moody's analysis of us corporate rating triggers heightens need for increased disclosure. *Moody's Special Comment*, 2002.
- T. Suzuki. On the rate of convergence of the difference finite element approximation for parabolic equations. *Proc. Japan Acad. Ser. A Math. Sci.*, 54:327–354, 1978.
- D. Tavella and C. Randall. *Pricing Financial Instrument with the Finite Difference Method*. Wiley Financial Engineering, 2000.
- T. Wilson. Portfolio credit risk i. *Risk*, (10(9)):111–117, 1997a.
- T. Wilson. Portfolio credit risk ii. *Risk*, (10(10)):56–61, 1997b.