HAL open science

Interest rates for insurance : models calibrations and approximations.

Sophian Mehalla

▶ To cite this version:

Sophian Mehalla. Interest rates for insurance : models calibrations and approximations.. Optimization and Control [math.OC]. École des Ponts ParisTech, 2021. English. NNT : 2021ENPC0022 . tel-03541696

HAL Id: tel-03541696

https://pastel.hal.science/tel-03541696

Submitted on 24 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

Taux d'intérêt pour l'assurance : approximations et calibrages de modèles

Mathématiques et STIC (MSTIC) ED 532

Spécialité : Mathématiques appliquées

Thèse préparée au sein du CERMICS, Ecole des Ponts ParisTech

Thèse soutenue le 18/10/2021, par
Sophian MEHALLA

Composition du jury :

Aurélien, ALFONSI Professeur, Ecole des Ponts ParisTech	*Président*
Christa, CUCHIERO Professeur, Universität Wien	*Rapporteur*
Peter, TANKOV Professeur, ENSAE	*Rapporteur*
Zorana, GRBAC Professeur, Université Paris-Diderot	*Examinateur*
Caroline, HILLAIRET Professeur, ENSAE	*Examinateur*
Sergio, PULIDO Professeur, ENSIIE	*Examinateur*
Bernard, LAPEYRE Professeur, Ecole des Ponts ParisTech	*Directeur de thèse*
Alexandre, BOUMEZOUED Docteur, Milliman Paris	*Examinateur*
Pierre-Edouard, ARROUY Actuaire, Milliman Paris	*Invité*

École des Ponts
ParisTech

École doctorale : MATHÉMATIQUES ET SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

Thèse de doctorat

Spécialité : Mathématiques Appliquées

présentée par

Sophian MEHALLA

Taux d'intérêt pour l'assurance : approximations et calibrages de modèles

---

Thèse dirigée par Bernard LAPEYRE
et préparée au CERMICS, École des Ponts ParisTech

Soutenue le 18 octobre 2021 devant le jury composé de :

Examinateurs	**Mme Caroline HILLAIRET**	ENSAE
	Mme Zorana GRBAC	Université Paris-Diderot
	M. Aurélien ALFONSI	École des Ponts ParisTech
	M. Sergio PULIDO	ENSIIE
Rapporteurs	**Mme Christa CUCHIERO**	University of Vienna
	M. Peter TANKOV	ENSAE
Directeur de Thèse	**M. Bernard LAPEYRE**	École des Ponts ParisTech
Co-encadrants	**M. Alexandre BOUMEZOUED**	Milliman Paris
	M. Pierre-Edouard ARROUY	Milliman Paris

Á la mémoire de mes grands-mères, Marie-Louise et Hasni.

Remerciements

S'il fallait l'identifier précisément, l'origine de la succession des événements qui m'a mené à la réalisation d'un doctorat en mathématiques dont ce manuscrit est l'aboutissement correspond sans doute à l'instant où Bernard Lapeyre, mon directeur de thèse, m'a accordé sa confiance en acceptant de m'encadrer pour ce qui fut mon premier stage, en 2015. Je lui suis pleinement reconnaissant pour ces six années de discussions régulières et riches durant lesquelles son enthousiasme, son intuition, l'étendue de ses connaissances et sa curiosité m'auront marqué. Je tiens à le remercier pour ses nombreux enseignements, sa patience et sa pédagogie, sa disponibilité sans commune mesure, son sens de la rigueur et de la clarté, qui ont rendu son encadrement excellent.

Je voudrais également grandement remercier Alexandre Boumezoued et Pierre-Edouard Arrouy, mes encadrants chez Milliman, pour le suivi régulier de mes travaux, pour leur disponibilité condensée notamment en les nombreuses heures dédiées aux réponses à mes questions ou à la relecture de mes travaux. Leur soutien et leur expertise font de l'équipe de Recherche et Développement de Milliman un lieu de travail très agréable.
Je tiens aussi à remercier Laurent Devineau qui m'a très tôt accordé sa confiance, alors que je n'étais encore que stagiaire. C'est grâce à lui que j'ai pu concrétiser cette ambition de réaliser un doctorat.

Toute ma reconnaissance va ensuite à Christa Cuchiero et à Peter Tankov. Je les remercie d'avoir accepté de relire attentivement mon manuscrit et de rédiger un rapport sur mon travail. Je suis très heureux qu'ils y aient trouvé un intérêt. Je remercie Aurélien Alfonsi d'avoir accepté de présider le jury de ma thèse ainsi que pour ses cours de *Master* et ses travaux de recherche qui m'ont accompagné en fil rouge tout au long de mes propres travaux. Je remercie Zorana Grbac d'avoir accepté de faire partie du jury ainsi que de m'avoir accepté en tant qu'auditeur libre dans ses cours lors de ma première année de thèse. Je remercie Sergio Pulido d'avoir accepté de faire partie du jury ainsi que de l'intérêt qu'il a pu montrer pour mon travail lors des différents exposés que j'ai réalisé. Je remercie Caroline Hillairet de me faire l'honneur de sa présence au sein du jury.

Mes travaux ont été rendu possible par le soutien, la confiance et l'accueil de l'ensemble des collaborateurs de Milliman, à commencer par Éric Serant et Jérôme Nebout dont la vision attribuant un rôle important à la Recherche & Développement me semble précieuse. Je voudrais remercier particulièrement le service *Legal*, et notamment Ulrich, pour le temps dédié à l'établissement des conditions juridiques entourant cette thèse. Un grand merci à Alexandra Rodrigues, pour sa gentillesse, systématiquement arrangeante pour toute demande. Je remercie toutes les personnes que j'ai pu côtoyer quotidiennement au sein de l'équipe R&D, qui participent à l'ambiance qui y règne et avec qui j'ai pu participer à la résolution des nombreuses questions, souvent stimulantes, qui se posent à nous. Notamment : Paul Bonnefoy-Cudraz et Hervé, avec qui j'ai eu l'opportunité de travailler à la rédaction de publications (à venir, on l'espère !), Damien, Laetitia, Julien (que je remercie particulièrement pour la relecture du premier chapitre qui suit), Amal et Yousra.

Mes remerciements vont ensuite au CERMICS pour avoir mis en place les meilleures conditions à mon accueil régulier. C'est ainsi que j'adresse un très grand merci à Isabelle Simunic, pour sa disponibilité, son aide et ses réponses à toutes les questions, cruciales et insignifiantes et grâce à qui le travail au CERMICS est largement facilité. Je remercie tous les membres du CERMICS pour la bonne ambiance en place et leur disponibilité. Je pense notamment à Jean-François Delmas qui m'a initié au calcul des probabilités en première année à l'ENSTA et grâce à qui j'ai pu intégrer le CERMICS dès 2015. Merci également aux doctorants du CERMICS, qui m'ont toujours réservé le meilleur accueil, au premier rang desquels je salue William et Adel. Je remercie ensuite Vincent Lefieux de m'avoir confié la charge d'un groupe de travaux dirigés dans son cours de statistiques à l'Ecole des Ponts.

Moins rigoureusement, mais plus justement, ce doctorat trouve en réalité ses origines il y a bien plus longtemps qu'un horizon de quelques années. Je pense ainsi notamment au soutien, à la présence et aux encouragements depuis toujours de ma famille et évidemment de mes parents, Simone et Hacène. Merci à ma mère pour sa gentillesse, son dévouement et les heures passées à faire les devoirs, certainement loin d'être étrangères à la rédaction de ce manuscrit. Merci à mon père de m'avoir transmis sa curiosité scientifique et historique, son sens de la droiture et son exigence. Ces quelques lignes constituent un bien maigre remerciement à l'égard de tout ce qu'ils ont fait pour moi depuis ma naissance. Je remercie aussi mes sœurs, Monia et Chérine, pour leur soutien et leur humour, malgré les approximations nombreuses dont peut souffrir ce dernier.

J'ai une pensée pour mes professeurs, de l'école primaire à l'école d'ingénieur ainsi que pour mes encadrants lors de mes différentes expériences professionnelles ou sportives (Benjamin Guedou, Aymeric Bettinelli, Jean-Baptiste Garnier, Fabien Ryckeboer notamment), qui m'ont transmis leur curiosité, leurs connaissances et leur exigence. Je remercie également tous mes amis, croisés sur les bancs de l'école, dans le vestiaire d'une association sportive ou ailleurs, avec qui j'ai le plaisir de partager discussions et divertissements. Je pense, *notamment*, à : Agathe, Alex B., Arthur, Cassandra, Elias Tounzal, Félix le shooteur retraité, Flavien, Gontran Lance (rien), Hélène, Jérémy (du braquet), Jordan, Kevin, Mathilde, Nicolas, Paul, Raphael, Sam et Xavoonet l'homme standard.

Enfin, je remercie Clémence de sa présence à mes côtés, de son soutien indéfectible à mon égard et de la force de caractère dont elle fait preuve.

Résumé

Cette thèse est motivée par l'établissement de méthodes de calibrage efficaces tant du point de vue du temps de calcul qu'en termes de précision dans la réplication des données observées dans un contexte assurantiel. En effet, les polices d'*assurance-vie* lient assureurs et assurés à moyen et long termes par des rémunérations dépendantes du rendement des investissements réalisés par l'assureur et potentiellement garanties pour les souscripteurs en contrepartie du versement d'une prime initiale par ceux-ci. Ces contrats contiennent également un certain nombre d'options que les assurés peuvent déclencher. La valorisation adéquate des mécanismes proposés dans les produits d'assurance-vie repose sur des méthodes de simulations aléatoires permettant d'explorer au mieux l'ensemble des comportements possibles des assurés sous divers environnements économiques. La réglementation européenne requiert que les trajectoires simulées le soient en univers *Risque-Neutre*. De plus, afin d'assurer la cohérence avec les conditions actuelles, les modèles utilisés pour générer ces trajectoires se doivent d'être calibrés aux prix de dérivés observés sur les marchés financiers à date: c'est la *market consistency*. Le Chapitre 1 présente le contexte opérationnel dans lequel s'incrit cette thèse avec une attention particulière accordée aux exigences réglementaires.

Historiquement utilisés par les fournisseurs de scénarios économiques aux assureurs, le modèle de marché LIBOR (*LIBOR Market Model*, ou LMM) et ses variantes dédiés à la modélisation des taux d'intérêt sont devenus populaires parmi les assureurs, notamment en raison de leurs capacités à satisfaire les exigences en termes de *market consistency*. En pratique, l'obtention de formules analytiques représentant les prix des marchés à répliquer est nécessaire pour rendre ces modèles opérationnels. Á ce titre, un certain nombre d'approximations est nécessaire dont la principale est la technique dite de figement («freezing») qui consiste à supprimer une part de l'aléa simplifiant ainsi le modèle. Nous discutons de la validité de cette approximation dans différents cadres et de sa possible amélioration Chapitre 2. Nous montrons que le modèle ainsi approché conserve une richesse suffisante pour permettre une réplication suffisamment précise des données de marchés dans les conditions économiques actuelles.

Nous étudions ensuite plus précisément les modèles LMM intégrant un facteur de volatilité stochastique. Un modèle très populaire parmi les assureurs assimile cette volatilité à un processus Cox-Ingersoll-Ross (CIR). Nous proposons une modélisation alternative dans laquelle des méthodes numériques efficaces fondées sur des expansions polynomiales de type Gram-Charlier peuvent être implémentées avec la garantie d'une précision suffisante. Le Chapitre 3 est l'occasion d'une revue historique de ces expansions polynomiales. Le modèle proposé et la méthode de calibrage associée sont ensuite étudiés Chapitre 4 : il repose sur l'utilisation d'un processus Jacobi afin de représenter la volatilité stochastique et dont le caractère borné est crucial pour la convergence de l'expansion Gram-Charlier. Ce processus peut être vu comme une extension du CIR au sens où sa variance instantanée est quadratique (et non affine dans le cas général). Il appartient à la classe des processus polynomiaux dont les moments peuvent être calculés par formules fermées. Nous montrons que le processus Jacobi converge vers le processus CIR au sens fort et établissons des vitesses de convergences utiles pour mesurer la distance entre ces deux modèles.

L'optimisation numérique sur laquelle repose le calibrage peut être accélérée grâce à l'information locale que comporte le gradient de la fonction représentant les prix dans les modèle étudiés. Nous établissons la formulation exacte du gradient dans le modèle d'intérêt (avec volatilité CIR) et discutons de son impact sur l'accélération du calibrage des modèles Chapitre 5. Une discussion sur l'accélération du calibrage dans le cadre du modèle polynomial (avec volatilité Jacobi) est également menée.

Mots-clefs : Modélisation, Assurance, Taux d'intérêt, Valorisation de *swaptions*, Modèle de marché *LIBOR* avec coefficient de déplacement, Volatilité stochastique, Processus affines, Processus polynomiaux, Processus de Jacobi, Expansions Gram-Charlier, Optimisation.

Abstract

This thesis is motivated by the establishment of efficient calibration methods from both computational time and accuracy points of view in insurance framework. Indeed, life insurance policies bind insurers and policyholders on through medium and long terms payments that depend on the returns of the investments realized by the company. Policies may also include some guaranteed payments or a number of options that their contract holder can activate. To purchase such policies, customer must pay an initial insurance premium. The convenient *pricing* of the mechanisms included in those life insurance contracts is based on simulation methods to best explore the possible behaviours of the policyholders under various economic environments. European regulation imposes that the generated paths should be simulated in *Risk-Neutral* universe. Moreover, to ensured the consistency with current economic conditions, models used to generated those paths are asked to be calibrated to market prices of derivatives: this is the so-called *market-consistency*. In Chapter 1, we present in more details the operational framework from which this thesis originates with a special attention dedicated to regulatory requirements.

Originally used by providers of economic scenarios, the LIBOR Market Model (LMM) and its different versions dedicated to the modelling of interest rates have become quite popular among insurers, in particular due to its ability to comply with *market consistency* criteria. In practice, obtention of analytical formulas to represent market prices to be replicated is key to calibrate the models in a reasonable time. To this end, a number of approximations are usually made. The first one is the so-called *freezing* technique that amounts to remove some randomness of the model to simplify it. We discuss its validity in various pricing frameworks and its possible improvement in Chapter 2. We will see that the model approximated this way keeps enough tractability to replicate market data with good accuracy in late market conditions.

Then we take a closer look to versions of the LMM with stochastic volatility. A very popular model among insurers in this class of models represents the volatility factor by a Cox-Ingersoll-Ross (CIR) process. We propose in this thesis an alternative approach in which efficient numerical methods of derivatives pricing, based on Gram-Charlier polynomial expansions, can be implemented with high precision. On this occasion, we provide in Chapter 3 an historical review of these polynomial expansions techniques. The proposed model and its calibration method is then studied in Chapter 4: it is based on the representation of the volatility factor by a Jacobi process; its bounded property allows to ensure the convergence of the mentioned polynomial expansions. This process can be viewed as an extension of the CIR one in the sense that its instantaneous variance is quadratic (and non affine *a priori*). It belongs to the class of polynomial processes whose moments can be calculated by closed-form formulas. We show that the Jacobi process strongly converges towards the CIR one and derive some rates of convergence useful when assessing the distance between the two models.

The numerical optimization on which the calibration is based can be speeded up thanks to the local information carried out by the gradient of pricing function in the studied models. We establish the analytical formulation of the gradient of the price in the standard modelling framework of interest (with CIR volatility process) and discuss its impact on computation time in Chapter 5. We will also assess the impact of the gradient in the polynomial model we proposed (with Jacobi volatility).

Keywords: Modelling, Insurance, Interest rates, *Swaptions* pricing, LIBOR Market Model with Displaced Diffusion, Stochastic volatility, Affine processes, Polynomial processes, Jacobi process, Gram-Charlier expansions, Optimization.

List of publications

- [AMLB20] P.-E. Arrouy, A. Boumezoued, B. Lapeyre and S. Mehalla, Jacobi Stochastic Volatility factor for the Libor Market Model, *Under revision for Finance and Stochastics*.
- [AAB+20] H. Andres, P.-E. Arrouy, P. Bonnefoy, A. Boumezoued, and S. Mehalla, Fast calibration of the LIBOR Market Model with Stochastic Volatility based on analytical gradient, *Submitted*.
- P.-E. Arrouy, A. Boumezoued, and S. Mehalla, On the *freezing* of the LIBOR Market Model, *Working paper*.

Contents

Introduction (en Français) **13**
0.1 Les GSE et Solvabilité II . . . 14
0.2 Contributions de la thèse . . . 15
0.2.1 Analyse de la technique de figement (freezing) . . . 15
0.2.2 Volatilité Jacobi et expansions polynomiales . . . 16
0.2.3 Calcul du gradient analytique pour le calibrage . . . 19

1 Introduction to Economic Scenario Generators and Solvency II **20**
1.1 Motivations of the thesis in a nutshell . . . 20
1.2 ESGs: definitions and usages . . . 21
1.3 Introduction to Solvency II regulation . . . 23
1.3.1 General considerations . . . 24
1.3.2 Pillar I: Quantitative Requirements . . . 24
1.3.3 Pillar II: ORSA . . . 29
1.4 Economic scenario generators step by step . . . 30
1.4.1 Building the regulatory initial yield curve . . . 30
1.4.2 Choosing the models . . . 42
1.4.3 Calibration on market data . . . 43
1.4.4 Simulations and validations . . . 49
1.5 Interest rates modelling . . . 50
1.5.1 Some preliminaries . . . 50
1.5.2 Short rates models . . . 56
1.5.3 Market models . . . 56
1.5.4 Replication of the smile . . . 61

2 *Freezing* approximations for the calibration of the Displaced Diffusion LIBOR Market Model **72**
2.1 Introduction . . . 72
2.2 Standard freezing method . . . 72
2.2.1 Rebonato . . . 73
2.2.2 Hull & White . . . 74
2.2.3 Historical estimations . . . 75
2.2.4 Pricing error . . . 75
2.3 Proposed alternative . . . 84
2.3.1 Rebonato-like . . . 86
2.3.2 Hull & White-like . . . 86
2.3.3 Numerical results . . . 86

3 On the Jacobi process and polynomial expansions **88**
3.1 Jacobi process: an overview . . . 89
3.1.1 Introduction . . . 89
3.1.2 Definition and existence . . . 90
3.1.3 Distributional properties . . . 92
3.1.4 Study of the moments . . . 95
3.1.5 Link with Cox-Ingersoll-Ross process . . . 101
3.1.6 Integrated Jacobi process . . . 104
3.1.7 Simulation of the Jacobi process . . . 105
3.1.8 Extensions . . . 108
3.1.9 Time inhomogeneous polynomial diffusions . . . 113
3.2 The problem of moments . . . 118
3.3 Series expansion and orthogonal polynomials: a statistical point of view . . . 121
3.3.1 History and general idea . . . 122
3.3.2 Gram-Charlier type A expansions . . . 125
3.3.3 Edgeworth series . . . 135
3.4 Density approximations in Hilbert space . . . 139

4 Jacobi Stochastic Volatility factor for the Libor Market Model **142**
4.1 Introduction . . . 142
4.2 Density approximation for stochastic volatility models . . . 144
4.3 Swaption pricing with Gram-Charlier expansion . . . 146
4.3.1 Jacobi process in the DDSVLMM . . . 146
4.3.2 Gram-Charlier expansion . . . 147
4.4 Rates of convergence . . . 152
4.4.1 Bounding error . . . 152
4.4.2 Truncation error . . . 174
4.4.3 Expression of Fourier coefficients . . . 176
4.4.4 Truncation error . . . 177
4.5 Numerical analysis . . . 180
4.5.1 Matrix exponential computation . . . 180
4.5.2 Parametrization of the DDSVLMM and its Jacobi version . . . 180
4.5.3 Why bounding the volatility factor? . . . 181
4.5.4 Convergence illustrations . . . 183
4.5.5 Pricing with Gram-Charlier series . . . 189

5 Fast calibration of the LIBOR Market Model with Stochastic Volatility based on analytical gradient **205**
5.1 Introduction . . . 205
5.2 Analytical gradient calculation and optimization routines . . . 206
5.2.1 Swaptions pricing . . . 206
5.2.2 Analytic characteristic function gradient . . . 207
5.2.3 Formulation of the calibration . . . 208
5.2.4 Calibration using gradient-based algorithms . . . 210
5.3 Calibration results . . . 212
5.3.1 Methods accuracy . . . 212
5.3.2 Time efficiency . . . 215
5.4 Extension of gradient-based approach to polynomial model . . . 216
5.4.1 Numerical results . . . 218

Appendices **222**

A Short rates models in insurance and pricing of interest rates derivatives **223**
A.1 Short rates models in insurance . . . 223
A.1.1 Hull-White model . . . 223
A.1.2 G2++ . . . 227
A.1.3 Two-factor Hull-White model . . . 228
A.1.4 CIR2++ . . . 229
A.2 Pricing of interest rates derivatives . . . 230
A.2.1 Pricing measures and hedging . . . 230
A.2.2 Change of *numéraire* . . . 233

B Detailed results on Monte-Carlo *frozen* volatilities **235**
B.1 Bachelier environment . . . 235
B.2 Black environment . . . 236

C About density function of Itô's integrals **238**

D Regularity of the price function in Heston model **241**

E Gradient vector of swaptions prices in DD-SV-LMM and optimization routines **243**
E.1 Partial derivatives of characteristic function . . . 243
E.1.1 Partial derivative of Ψ with respect to θ . . . 243
E.1.2 Partial derivative of Ψ with respect to κ . . . 243
E.1.3 Partial derivative of Ψ with respect to ϵ . . . 244
E.1.4 Partial derivative of Ψ with respect to ρ . . . 245
E.1.5 Partial derivatives of Ψ with respect to a, b, c and d . . . 246
E.2 On the Levenberg-Marquardt algorithm . . . 247
E.2.1 Standard Levenberg-Marquardt algorithm . . . 247
E.2.2 Extended Levenberg-Marquardt algorithm handling bound constraints . . 248

Introduction (en Français)

Les polices d'*assurance-vie* lient assureurs et assurés à moyen et long termes par des rémunérations garanties en contrepartie du versement d'une prime initiale par les souscripteurs. Ces contrats contiennent également un certain nombre d'options que les assurés peuvent déclencher jusqu'au terme du contrat. La valorisation adéquate des mécanismes proposés dans les produits d'assurance-vie repose sur des méthodes de simulations aléatoires permettant d'explorer au mieux l'activation des garanties et l'ensemble des comportements possibles des assurés sous divers environnements économiques. Les Générateurs de Scénarios Economiques (GSE) sont des outils numériques, devenus incontournables pour les compagnies d'assurance ou réassurance au cours de la décennie écoulée, dédiés à la génération de simulations aléatoires des principaux facteurs de risques financiers permettant de rendre compte de l'état d'une économie. Sur ces trajectoires simulées, les assureurs/réassureurs effectuent un certain nombre de traitements/calculs motivés par différents objectifs dont les principaux sont :

- Calcul des provisions techniques : ceux-ci sont définis comme étant la «meilleure estimation» des engagements de l'assureur augmentée d'une marge conservatrice («marge pour risque»). En vue de calculer la prime d'assurance versée par les souscripteurs de certains contrats complexes, comportant notamment un certain nombre d'options ou de garanties financières, qui est la contrepartie du versement d'une rente par l'assureur, la projection des contrats en question sous différentes conditions économiques est réalisée à l'aide des simulations fournies par les GSE.

- Calcul de montants de capitaux réglementaires : dans le cadre de la réglementation européenne *Solvabilité II* initiée suite à la crise financière de 2008, les compagnies d'assurance ou de réassurance se doivent d'immobiliser une certaine quantité de capitaux afin d'être en mesure de faire face à des mouvements de marché extrêmes. Le calcul de ce montant est réalisé en valorisant financièrement le passif de la compagnie d'assurance. Pour ce faire, le comportement dudit passif est projeté, depuis son état en date de calcul, sous les différentes conditions de marché simulées par le GSE ; les situations les plus adverses permettent de calculer le montant de capital à immobiliser.

- Rééquilibrage de portefeuilles d'actifs et stratégie de couverture : afin de garantir le versement de rentes garanties dans les polices d'assurances, les assureurs/réassureurs adossent leurs passifs à des produits financiers – possiblement dérivés, constituant ainsi une part de leur actif. Le choix de ces produits financiers, de leurs pondérations ou encore de leurs durées de détention peuvent être déterminés à l'aide des trajectoires simulées par les GSE.

- Gestion des risques : le profil de risque de la compagnie peut être évalué de manière prospective grâce aux simulations fournies par les GSE. En effet, des perturbations dans les conditions économiques actuelles – servant de point de départ aux simulations des GSE - peuvent être réalisées afin de simuler une économie *stressée* ; des anticipations des

niveaux des quantités financières (niveaux d'inflation ou de taux d'intérêt par exemple) peuvent être intégrées aux GSE et donc répercutées sur les trajectoires simulées.

Davantage d'informations sont données sur l'utilisation pratique des GSE au Chapitre 1.

0.1 Les GSE et Solvabilité II

Les procédures dédiées aux calculs réglementaires sous Solvabilité II constituent un cas d'utilisation particulier mais central quant à la conception des GSE. La législation établit une procédure précise pour le calcul des fonds propres réglementaires qui repose notamment sur la construction d'une courbe de taux «sans risque» conçue comme un ajustement d'une courbe de marché. Celle-ci est structurante pour les trajectoires simulées car détermine le niveau moyen des taux d'intérêts simulés par les GSE. Les comportements des agents, reflétés notamment par leurs investissements, au sein d'une économie résultent d'abord du niveau des taux d'intérêts (nominaux et réels), et il est naturel d'anticiper que les montants réglementaires associés au risque financier dépendent au premier ordre des niveaux de taux. Cette courbe sans risque est fournie mensuellement par le régulateur européen du marché de l'assurance: l'Autorité européenne des assurances et des pensions professionnelles ou *European Insurance and Occupational Pensions Authority (EIOPA)*. Le Chapitre 1 décrit brièvement les exigences associées à Solvabilité II avec une description détaillée de la construction de la courbe sans risque — dite «courbe EIOPA».

Les modèles composant les GSE sont choisis par les assureurs selon les risques financiers qu'ils souhaitent prendre en compte lors de la projection de leur passif. Sur le marché européen, ils peuvent inclure jusqu'à six facteurs de risques à modéliser : taux d'intérêt –nominaux et réels–, indices boursiers, indices immobiliers, défauts des contreparties ou taux de change (pour les GSE modélisant plusieurs économies, encore assez peu répandus à date). Les polices d'assurances étant des contrats comportant un certain nombre d'options activables par les assurés ainsi que des garanties de prestations, les trajectoires simulées par les GSE doivent nécessairement être stochastiques afin de rendre compte de la variété des comportements possibles des assurés sous différentes trajectoires économiques. De plus, les polices d'assurances étant généralement des contrats de long termes, la modélisation des taux d'intérêt est au centre des préoccupations des assureurs.

La réglementation Solvabilité II demande aux assureurs de valoriser le passif de leur compagnie dans l'hypothèse où celui-ci peut être assimilé à un produit échangeable sur les marchés financiers. Les calculs sont alors réalisés sous l'hypothèse d'Absence d'Opportunité d'Arbitrage, c'est-à-dire en univers Risque-Neutre. De plus, et afin que la valorisation obtenue soit en cohérence avec l'environnement financier à la date de sa réalisation, les modèles se doivent de répliquer des prix de produits dérivés observés sur les marchés. Cette exigence est connue sous la dénomination de «*market consistency*». Il est alors requis de chaque modèle une réplication d'un nombre important de prix (possiblement plus de 300) de dérivés dont les caractéristiques variées nécessitent des modèles spécifiques.

Dans ce contexte, les modèles issus de l'industrie financière (banque, gestionnaire d'actifs, etc.) ont naturellement été utilisés par les assureurs. Bien que ceux-ci ne soient pas initialement développés pour les calculs assurantiels, leur intégration progressive au sein des différents outils complexes des compagnies et les ressources qui y ont été dédiées les rendent à présent inévitables et massivement utilisés. Parmi ces modèles, et en cohérence avec les commentaires précédents, ceux dédiés à la modélisation des taux d'intérêt occupent l'essentiel des ressources opérationnelles. Notamment, ce sont les modèles dont il est généralement requis la reproduction du plus grand nombre de prix de dérivés. Afin de satisfaire les exigences en *market consistency*, des modèles spécifiques peuvent être choisis – notamment, les modèles dits «à volatilité

stochastique» – qui permettent une réplication précise des données de marché en contrepartie d'un temps de calcul conséquent. Les modèles les plus populaires (au sein des assureurs) appartenant à cette catégorie sont des variantes du modèle de Brace, Gatarek et Musiela – aussi couramment dénommé *LIBOR Market Model (LMM)* – qui permettent de générer des taux négatifs. Cette dernière capacité est devenue indispensable dans les conditions économiques actuelles. Intégrant un facteur de volatilité stochastique modélisé par un processus de Cox-Ingersoll-Ross (CIR) et un coefficient de déplacement permettant de générer des taux négatifs, le modèle qui s'est peu à peu distingué parmi les modèles de référence est désigné comme étant le *Displaced Diffusion with Stochastic Volatility LIBOR Market Model (DDSVLMM)*.

Le «calibrage» des modèles désigne le processus au cours duquel les paramètres permettant la meilleure réplication des données de marché sont identifiés. Il repose sur deux étapes : l'établissement de formules analytiques pour les prix induits par le modèle (qui sont donc fonction des paramètres de ce dernier) lorsque cela est possible ce qui est primordial dans l'objectif de réaliser un calibrage en un temps raisonnable puis l'optimisation numérique au cours de laquelle la distance entre ces prix modèles et les prix de marché est minimisée. Il s'avère que la formulation des prix obtenus dans le DDSVLMM est relativement complexe à mettre en oeuvre numériquement ce qui induit un temps de calibrage lui aussi important. Cette thèse est motivée par l'établissement de méthodes de calibrages alternatives permettant de réduire le temps de calcul dédié tout en contrôlant la *market consistency*. Pour ce faire, nous discutons d'abord des approximations standard de *freezing* de modèles (Chapitre 2). Nous proposons ensuite une modélisation alternative qui intègre le processus Jacobi en lieu et place du CIR et établissons des méthodes de calibrage reposant sur des expansions polynomiales (Chapitre 3) dont nous étudions l'efficacité et la précision (Chapitre 4). Enfin, nous développons un calcul analytique du gradient de la fonction de prix pour la réduction des temps de calibrage (Chapitre 5).

0.2 Contributions de la thèse

0.2.1 Analyse de la technique de figement (freezing)

L'obtention de formules «fermées» est donc essentielle pour calibrer les modèles. Les différentes variantes du LIBOR Market Model reposent toutes sur l'approximation initiale qui consiste à faire l'hypothèse d'une faible variabilité de certaines quantités aléatoires au cours du temps. Celles-ci peuvent alors être remplacées par leurs valeurs initiales : c'est la méthode dite de «figement» (*freezing*). Les calculs de prix à mener sont alors considérablement simplifiés.

Contributions de la thèse

Nous analysons dans le Chapitre 2 les différentes approches possibles pour ce figement et discutons de sa validité dans les conditions économiques actuelles. En particulier, nous vérifions que les hypothèses sur lesquelles elles reposent sont fondées et n'induisent pas une rigidité trop importante du modèle. Des études similaires ont été menées au milieu des années 2000 sur la pertinence de ce type d'approximation dans le cadre des modèles de taux considérés. Les taux d'intérêt étaient alors positifs et l'utilisation de la formule de Black était donc la norme. Nous proposons ici de mettre à jour ces études grâce à des données récentes et de l'étendre au cas de l'utilisation de la formule de Bachelier et ce pour des périodes de temps plus longues.

Plus concrètement, la trajectoire $(S_t^{m,n})_{t\leq T_m}$ du taux *swap* de maturité T_m et de ténor T_n est modélisée par la dynamique générique suivante :

$$\mathrm{d}S_t^{m,n} = \sum_{k=m}^{n-1} \omega_k\big(t, \big(F_j(t)\big)_{j=m,\dots,k}\big) F_k(t)\mathrm{d}W_t,\ t \leq T_m, \tag{1}$$

où $(W_t)_{t\geq 0}$ est un mouvement Brownien sous la probabilité $\mathbb{P}^S$, $(F_k(t))_{t\leq T_k, k=m,\dots,n-1}$ sont les trajectoires des taux *forward* de maturités respectives $T_m, \dots, T_{n-1}$ et dont la valeur du taux swap dépend et $\big(\omega_k\big(t, (F_j)_{j=m,\dots,k}\big)\big)_{t\leq T_m, k=m,\dots,n-1}$ sont des quantités stochastiques fonctions du temps ainsi que des taux *forward*. En raison de la complexité de l'expression qui lie ces coefficients ω_k aux taux *forward*, la dynamique (1) est elle-même difficilement exploitable pour le calcul des prix d'options d'achat sur swap (*swaptions*) proportionnels à la quantité $\mathbb{E}^S\big[(S_{T_m}^{m,n} - K)_+\big]$ (où K désigne le prix d'exercice – aussi appelé *strike* – de l'option). L'idée de la technique de *figement* est d'approcher le modèle (1) en supprimant l'aléa des fonctions ω_k de sorte à ce que le calcul de l'espérance précédente puisse être réalisé par formulation explicite. Toujours en écrivant de manière générique, on remplacera le modèle (1) par le suivant :

$$\mathrm{d}S_t^{m,n} = \sum_{k=m}^{n-1} \tilde{\omega}_k\big(t\big) F_k(t) \mathrm{d}W_t, \ t \leq T_m, \tag{2}$$

où l'on a introduit $\tilde{\omega}_k(t) = \omega_k\big(t, (F_j(0))_{k=m,\dots,k}\big)$. Pour définir ces nouveaux coefficients déterministes $\tilde{\omega}_k$, deux approches sont répandues : nous les discutons et comparons leurs efficacités dans le Chapitre 2. Sous ces approximations, le calcul de $\mathbb{E}^S\big[(S_{T_m}^{m,n} - K)_+\big]$ peut alors être réalisé efficacement et le modèle (2) peut alors être calibré en un temps acceptable. Deux questions se posent alors auxquelles nous nous efforçons de répondre : l'approximation (2) du modèle de (1) est-elle toujours suffisamment riche pour permettre une réplication précise des données de marchés ? Les prix induits par le modèle approché (2) sont-ils suffisamment proches des prix induits par le modèle exact (1) ?

0.2.2 Volatilité Jacobi et expansions polynomiales

Le modèle le plus communément utilisé par les assureurs pour répliquer de l'optionalité est un modèle à volatilité stochastique semblable au modèle de Heston pour la modélisation des indices actions. Il offre une représentation semi-analytique des prix au moyen d'une transformation de Fourier de la fonction caractéristique associée au modèle en exploitant les propriétés des processus affines. Bien que permettant une réplication précise des données conforme aux critères de *market consistency*, le calibrage de ce modèle est relativement lent.

Le calcul des prix dans le modèle que nous proposons stipule que le facteur de volatilité est représenté par un processus Jacobi permettant ainsi l'utilisation des expansions polynomiales de type Gram-Charlier pour le calcul des prix induits par le modèle. Ces méthodes d'expansions assez anciennes ont été originellement introduites pour approcher des densités de probabilités inconnues. Une revue des propriétés du processus Jacobi et du développement historique des expansions polynomiales de densités est l'objet du Chapitre 3. Afin d'assurer la précision des calculs menés, la convergence de ces expansions est nécessaire. Le caractère borné du processus de Jacobi permet de s'en assurer théoriquement, contrairement au modèle de référence construit avec le processus CIR. Une de nos préoccupations sera alors de mesurer la «distance» entre le modèle proposé et le modèle de référence. Les travaux présentés dans le Chapitre 4 de cette thèse constituent une version augmentée de l'article [AMLB20] (soumis et en cours de révision).

Contributions de la thèse

Essentiellement, nous proposons d'utiliser le processus de Jacobi comme une approximation du processus Cox-Ingersoll-Ross pour la modélisation de la volatilité stochastique. La distance entre ces deux processus peut être évaluée selon différentes métriques. La distance entre les taux *swap* associés à chacune de ces volatilité peut alors être déduite ainsi que l'impact du modèle sur la valorisation des contrats dérivés.

En équations, la version du DDSVLMM utilisée pour les taux swap est similaire à (1) dans lequel un facteur de volatilité stochastique a été ajouté. En appliquant la technique de figement précédemment décrite, le modèle proposé s'écrit comme suit :

$$\begin{cases} \mathrm{d}S_t^{m,n} &= \sqrt{V_t}\sum_{k=m}^{n-1}\tilde{\omega}_k(t)F_k(t)\mathrm{d}W_t, \\ \mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}B_t, \end{cases} \quad t \leq T_m, \tag{3}$$

où $(B_t)_{t\geq 0}$ est un mouvement Brownien dont la loi jointe avec W est caractérisée par le coefficient ρ tel que $\mathrm{d}\langle W_\cdot, B_\cdot\rangle_t = \rho\mathrm{d}t$ et κ, θ, ϵ sont les paramètres définissant la dynamique de Cox-Ingersoll-Ross qui représente le processus de volatilité V. Sous cette évolution stochastique, les prix d'options d'achat sur $S_{T_m}^{m,n}$ peuvent être calculés grâce à la transformée de Fourier de la fonction caractéristique de $S_{T_m}^{m,n}$. Celle-ci s'exprime au moyen d'équations de Riccati conséquences de la propriété affine du modèle (3). Cependant, le calcul de ces transformées est opéré par des méthodes de quadrature relativement lourdes numériquement.

C'est ainsi que nous proposons d'adopter le processus de Jacobi V^J pour modéliser la volatilité stochastique. Celui-ci est solution de l'équation différentielle stochastique suivante :

$$\mathrm{d}V_t^J = \kappa(\theta - V_t^J)\mathrm{d}t + \epsilon\sqrt{Q(V_t^J)}\mathrm{d}B_t, \ t \leq T_m, \tag{4}$$

où l'on a introduit la fonction polynomiale $Q(x) = \frac{(x-v_{min})(v_{max}-x)}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2}$ avec $0 \leq v_{min} < v_{max} \leq +\infty$. Celle-ci permet de garantir le caractère borné du processus de Jacobi dans le temps : $\mathbb{P}(\forall t \geq 0 : V_t^J \in [v_{min}, v_{max}]) = 1$ dès lors que $V_0^J \in [v_{min}, v_{max}]$. Une propriété remarquable du processus Jacobi est la suivante :

$$\lim_{(v_{min},v_{max})\to(0,+\infty)} V_t^J = V_t,$$

où la convergence est à comprendre au sens fort. Nous établissons dans cette thèse des vitesses de convergences fortes et faibles, notamment dans le théorème suivant.

Théorème. *Soit $T > 0$. Il existe des constantes positives $\tilde{C}$ et C' telles que*

$$\sup_{0\leq t\leq T} \mathbb{E}^S\left[\left|V_t^J - V_t\right|\right] \leq \tilde{C}\Big/\log\left(v_{max}/v_0\right)$$

et

$$\mathbb{E}^S\Big[\sup_{0\leq t\leq T}\left|V_t^J - V_t\right|\Big] \leq C'\Big/\sqrt{\log\left(v_{max}/v_0\right)}.$$

Notons à présent $S^{m,n,J}$ le taux *swap* décrit par une dynamique similaire à (3) dans laquelle V a été remplacé par V^J. À v_{min} et v_{max} fixés, la modélisation du taux *swap* associé à la dynamique (4) peut être vue comme une approximation de la modélisation de référence (3) puisque des vitesses de convergences de $S^{m,n,J}$ vers $S^{m,n}$ peuvent à leur tour être établies.
Nous l'avons évoqué précemment, travailler avec V^J permet de calculer les prix de *swaptions* plus efficacement grâce aux expansions polynomiales de type Gram-Charlier tout en garantissant la convergence des expansions dans ce cadre. Ces prix s'écrivent alors comme des combinaisons linéaires des moments du taux *swap* :

$$\mathbb{E}^S\left[(S_{T_m}^{m,n,J} - K)_+\right] = c_0 + c_1\mathbb{E}^S\left[S_{T_m}^{m,n,J}\right] + c_2\mathbb{E}^S\left[(S_{T_m}^{m,n,J})^2\right] + c_3\mathbb{E}^S\left[(S_{T_m}^{m,n,J})^3\right] + \dots \tag{5}$$

où les coefficients $(c_i)_{i\in\mathbb{N}}$ ne dépendent pas de la loi du taux *swap* (ils dépendent notamment de K et de la fonction payoff $\varphi(x) = (x - K)_+$). Reste alors à être en mesure de calculer

efficacement la séquence de moments $\left(\mathbb{E}^S\left[\left(S_{T_m}^{m,n,J}\right)^k\right]\right)_{k\in\mathbb{N}}$, ce qui est réalisé grâce au caractère polynomial de la dynamique (4).

Pour pouvoir écrire l'identité (5), il est nécessaire d'assurer la convergence du terme de droite. Elle est conséquence du caractère borné de la volatilité : pour une borne supérieure v_{max} raisonnable, cette convergence est établie théoriquement et nous prouverons que le résultat persiste dans le cas particulier où $v_{min} = 0$. Une telle représentation des prix n'est donc *a priori* pas permise dans le modèle standard (3) pour lequel $v_{max} = +\infty$. Nous avons d'ailleurs pu démontrer que la condition classiquement imposée aux densités de variables aléatoires approchées pour assurer la convergence de l'expansion n'est pas satisfaite pour les modèles à volatilité stochastique classiques (i.e. dans lesquels la volatilité n'est pas bornée).

Théorème. *Lorsqu'un modèle à volatilité stochastique non bornée est utilisé pour représenter $S^{m,n}$, la convergence de l'expansion Gram-Charlier de la densité du taux swap n'est pas assurée.*

Apparaît alors le dilemme suivant : d'une part, et en accord avec l'avant-dernier théorème énoncé, nous souhaitons choisir le paramètre v_{max} aussi élevé que possible pour approcher précisément le DDSVLMM de référence (3) ; d'autre part, l'identité (5) ne s'écrit que pour des v_{max} suffisamment faibles.

Cela étant dit, la question de l'erreur commise en bornant la volatilité – en remplaçant V par V^J, dit rapidement – sur le calcul des prix se pose. Nous prouverons le résultat suivant (cf. Section 4.4.1.4 pour une formulation précise).

Théorème. *Soient $v_{min} > 0$ et $\beta > 1$. Soit f une fonction gain bornée deux fois continûment dérivable à dérivées bornées et telle que les fonctions $x \mapsto xf'(x)$ et $x \mapsto x^2 f''(x)$ sont bornées. Sous certaines conditions sur les paramètres de volatilité, il existe deux constantes positives $C_{v_{min}}$, c_1 et c_2 telles que*

$$\left|\mathbb{E}^S\left[f(S_{T_m}^{m,n})\right] - \mathbb{E}^S\left[f(S_{T_m}^{m,n,J})\right]\right| \leq C_{v_{min}}\left(\frac{c_1}{v_{max}} + c_2\right)$$

pour tout $v_{max} \geq \beta v_{min}$.

En pratique la série (5) doit être tronquée ce qui induit une seconde source d'erreur, l'erreur de troncation. Nous donnons des estimations de celle-ci. Notamment, lorsque $v_{min} > 0$ et sous certaines conditions supplémentaires sur les paramètres du modèle, le résultat suivant est démontré.

Théorème. *Soit $N \in \mathbb{N}$ l'ordre de troncation et soit ε_N l'erreur de troncation. Il existe des constantes $\tilde{q} \in (0,1)$ et $(C, C') \in \left(\mathbb{R}_+^*\right)^2$ telles que, dans le cas où $K = S_0^{m,n}$,*

$$\varepsilon_N \leq C\frac{\tilde{q}^{N+1}}{N^{1/4}},$$

et dans le cas où $K \neq S_0^{m,n}$,

$$\varepsilon_N \leq C\frac{\tilde{q}^{N+1}}{1-\tilde{q}}.$$

Les illustrations numériques fournies permettent d'appuyer les résultats théoriques et d'établir certaines conjectures. Concernant le temps de calcul du calibrage des modèles, la méthodologie proposée permet de réduire celui-ci d'environ 50% par rapport au calibrage du DDSVLMM de référence.

0.2.3 Calcul du gradient analytique pour le calibrage

Le second temps des calibrages des modèles consiste en la minimisation numérique de la distance entre prix modèles et prix observés au moyen d'un algorithme d'optimisation. La littérature relative à l'optimisation numérique est abondante, aussi une classe particulière d'algorithmes tire avantage de l'information contenue dans le gradient de la fonction cible. Nous explicitons celui-ci dans le cadre du modèle DDSVLMM référence (fondé sur l'utilisation du processus Cox-Ingersoll-Ross). Nous discutons également de l'apport de cette information pour le calibrage du modèle intégrant le processus Jacobi avec approximation du gradient.
Nous nous concentrons sur des algorithmes communément utilisés et implémentés dans les librairies classiques du langage R. Ceux-ci intègrent des algorithmes ne nécessitant aucune hypothèse de régularité de la fonction cible, des algorithmes à descente de gradient ou des algorithmes stochastiques. Ces derniers ne seront pas considérés étant donné qu'en pratique il est apprécié que la fonction liant les données à reproduire aux paramètres optimaux soit déterministe. Les travaux présentés dans ce chapitre constituent une version augmentée de l'article [AAB+20] (soumis à publication).

Contributions de la thèse
Avec le modèle (3), nous avons précisé que les prix de swaptions s'exprime en terme de transformée de Fourier de la fonction caractéristique du taux swap. Cette expression analytique peut alors être dérivée par rapport à chacun des paramètres du modèle. Si $\boldsymbol{\Theta}$ représente l'ensemble des paramètres du modèles, le gradient

$$\nabla_{\boldsymbol{\Theta}} \mathbb{E}^S \left[\left(S_{T_m}^{m,n} - K \right)_+ \right]$$

est lui aussi explicite. Intégré aux algorithmes d'optimisation, nous discutons de son apport en terme de temps de calcul et précision du calibrage par une étude comparative. Celle-ci est également réalisée dans le cadre du modèle polynomial construit avec (4). Selon les paramètrages choisis, la réduction du temps de calibrage par rapport à un calibrage standard sans utilisation du gradient peut être de près de 85%.

Chapter 1

Introduction to Economic Scenario Generators and Solvency II

We present in this chapter the operational environment in which this thesis has been drafted and the underlying applied problems that have motivated this work.

1.1 Motivations of the thesis in a nutshell

Economic scenario generators have quite lately become indispensable numerical tools for insurers and reinsurers. They are used on multiple occasions and possibly several times a year throughout heavy computational processes of various natures. Those computations are required to allow undertakings to pursue their activity that basically consists in managing risks. Among financial risks, interest rates one is essential for insurance activity due to the relative long-term commitments associated with insurance policies. Models dedicated to their modelling have reached significant complexity and are asked to replicate a relatively large number of market data. Efficient calibration methods for interest rates models are thus sought. This thesis is dedicated to the establishment of accurate and efficient calibration methods for interest rates models used in insurance context.

The calibration process designates the whole process starting with the input of market data and resulting in the derivation of associated model quantities. This process can be decomposed stepwise as: (i) extraction of market data; (ii) choice of the appropriate model to replicate them; (iii) numerical optimization procedure allowing to identify the optimal parameters defining the model; (iv) computation of model prices. In this thesis, we discuss all this point and our main theoretical contributions relate to steps (ii) and (iii). Point (ii) is discussed in Chapter 2: models used by insurers are quite complex to use as such and require to be beforehand approximated somehow. Two strategies have been identified related to item (iii). The first one consists in finding alternative prices representations. Following the work of [LDB20] who proposed to use expansions techniques to obtain quite easy to implement prices formulas leading to an important reduction of calibration time. However, some questions on the accuracy of the proposed method raised. [AFP17] studied similar expansion techniques in equity type context. We discuss this alternative in Chapter 4. The second one relates to the optimization algorithm: there are numerous algorithms allowing to minimize (maximize) some target function, and those based on the gradient of the objective function are much used. We use it in our interest rates context. It is discussed in Chapter 5.

1.2 ESGs: definitions and usages

Insurance business is an old activity that fundamentally consists in managing the different kinds of risks the economic agents or private individuals are exposed to. Insurance policies are designed to protect either individuals or estates and are at the centre of modern economies. The high technicality of economies of economically developed countries require the insurers to be able to face a large variety of risks. In particular, those economies are mainly driven by macroeconomic aggregates and financial quantities which explains that the financial risk arose as key for companies activity. Moreover, the backing of life-insurance contracts is realized thanks to financial derivatives traded on financial markets.

For risk management, strategic guidance, regulatory compliance, sensitivities computations or valuations of policies, the *Economic Scenario Generators* (ESGs) recently emerged as a must-have tool for insurers. Following [PCC+16], an ESG can be defined as «*a computer-based model of an economic environment that is used to produce simulations of the joint behaviour of financial market values and economic variables.*» An ESG is composed of several interacting models, each one being dedicated to the modelling of economic quantities that reflect different kind of risks. Current ESGs can account for interest rates risk, equity risk, credit risk, real estate risk and foreign exchange risk. Note that the very nature of insurance policies embedding optional guarantees of relative long life enlightens why movement in interest rates curve has predominant impact on the whole activity of insurers. Moreover, major part of the financial instruments used for backing life-insurance contracts comprises bonds (Sovereign or Corporate), possibly Zero-Coupon and other interest rates derivatives.

The following non-exhaustive list provides more common uses of ESGs in practice:

- Computations of the technical provisions: those are the "best estimate" of the commitments written in insurance policies to which a "risk margin" is added (see [dcpedr20] and below);
- Various regulatory computations: capital requirements (described below in Section 1.3 in more details) to pursue activity, valuation of balance sheet according to proper accounting standards (see [Par14] or [BRS12] notably);
- Assets and Liabilities Managements (ALM) (see [ACA20b] or [BKKR06] for instance);
- Risk management from forward-looking point of view;
- Pricing and hedging of some relatively complex insurance policies comprising optionality (e.g., variable annuities);
- New worldwide accounting standard named IFRS 17 (see [AZ20] who gave some insights on the topic with a focus on the building of the relevant risk-free curve);
- Market valuation works within a merger and acquisition transaction.

In practice, ESGs are incorporated in the process dedicated to one of the previous computation or externalised (and may be handled by an ESG provider). Following the motivation of the calculation that is made, insurer may want to employ either *Risk-Neutral* or *Real-World* ESGs. First ones are related to computations in which market data replication is required while second ones are associated to more prospective approaches.

Risk-Neutral ESG In most of the previously mentioned computations, replication of market data (prices of derivatives or volatilities) is required in order to ensure the consistency with current economic conditions. The translation of this notion for the regulatory requirements is named the *market consistency* (discussed in [VEKLP17] or [BMV19] and references therein). Furthermore, an important feature of insurance policies is that they embed optional guarantees. Consequently, the simulated paths obtained from ESGs should be *stochastic* to take into account

the value of this optionality coming from the variety of possible behaviours of policyholders. These two points mainly explain why, historically, insurers have been led to consider models coming from bank industry to perform the aforementioned calculations. Even though they were not designed for long-term projections and were mainly motivated by establishment of hedging strategy associated with high frequency[1] of portfolio rebalancing, they offer the ability to repliacte market data quite accurately based on well established methods and to generate stochastic paths. Notably, the literature relative to their calibration or simulation is significant. Now, such models are widely used and well integrated in the companies processes and though their limitations are more understood, the inertia of market practices make those models quite inescapable.

Real-World ESG For investment planning or risk management, forward-looking scenarios could be helpful for insurers. For such considerations, it is no longer the replication of market prices that is desired but that of some properties of empirical distributions. For instance, it can be required that the historical returns of the some (equity) index should be recovered in simulated scenarios. In addition to historical targets, some targets can be *motivated by economic forecasts*. For instance, anticipations on average level of inflation rates may be taken into account. Models for real-world scenarios are essentially composed of statistical models of time series. Calibrations methods are thus based on statistical techniques that involve large historical data set.

Real-World ESGs are also central for regulatory computations in Solvency II. Indeed when computing the *Solvency Capital Requirement* (SCR) (Pillar I, see further) or when projecting it on future date (Pillar II further), it is necessary to simulate the state of the economy under the Real-World measure and thus generate paths reproducing some stylized facts. Based on those historical paths, Risk-Neutral computations are performed in a second phase to value the liability side of the undertaking from a Risk-Neutral point of view and eventually derive the projected values of the *Best Estimates* (BEs). Real-World modelling is thus pivotal also for regulatory compliance.

ESG for regulatory compliance For solvency requirements, the Solvency II legislation proposed to value the balance sheet of the insurance (and reinsurance) companies from a *financial* perspective. Relevant ESG for such computations are thus *Risk-Neutral*. To be used for such regulatory calculations, ESGs should have few properties that can be found in [Eio15] in accordance with [Par14] and summarized in [dcpedr20] for the Euro zone or in [PCC+16] for broader point of view. Note that the following requirements are not specific to regulatory computations as they may be also wanted for other applications. To respect these requirements, ESGs users must set up monitoring measures; some are quite common, other can be specific to a company. Be that as it may, companies have to report and justify to the authorities the methodological choices and underlying assumptions made when setting out their ESG.

ESGs are asked to replicate the regulatory risk-free yield curve at date when the computation is performed. This is motivated by fact the whole methodology to value the balance sheet is based on risk-free discounts of cash-flows. The derivation of the regulatory risk-free term structure is not straightforward as market rates illustrate equilibrium on financial market resulting from considerations of financial agents whose concerns differ from that of insurance companies. The methodology followed by the european regulator, the European Insurance and Occupational Pensions Authority (EIOPA), to build the regulatory curve is described in details below in Section 1.4.1. The ability of replicating the initial term-structure motivates the choice of the interest rates model.

[1] Compared to time periods involved in insurance.

Models are also asked to satisfy the Non-Arbitrage Opportunity (NAO) assumption. Arbitrage opportunities are considered as rare events on market and thus not desirable in ESGs that aim to simulate economies. In this perspective, ESGs first have to accurately replicate market prices: this is the so-called *market consistency*. According to [uni09], «*the calculation of technical provisions shall make use of and be consistent with information provided by the financial markets and generally available data on underwriting risks (market consistency)*». This notion and its consequences are discussed in [VEKLP17]. ESGs have to be accurately calibrated to market prices and thus rely on appropriate calibration methods. Secondly, once calibrated, models have to be simulated consistantly with the NAO assumption. It requires accurate simulation methods. On simulated paths, some *martingale tests* are performed: empirical means are computed and compared to theoretical expectations of some martingales quantities. The test pass when the two are close enough (a tolerance threshold is given based on statistical sampling error).
More generally, several metric are usually employed to assess the accuracy of the *market consistency* criterion, at different stage of the uses of ESGs: distance between market and model prices obtained as outputs of the calibration, distance between market and Monte-Carlo prices, and distance between model and Monte-Carlo prices.

An important aspect of ESGs is their ability to jointly simulate economic drivers. The correlation between simulated paths is of interest for insurers. Correlations targets should beforehand be estimated in a model-free approach. Once the models are chosen, those correlations are translated in terms of correlations between risk drivers (Brownian motions). It is then checked that empirical and target correlations are close enough.

All those tests are though as ensuring the quality of the simulated paths that will be used for computing the Solvency Capital Requirements (SCR; see below). Note that some of these are *Monte-Carlo* tests, in the sense that they are based on empirical estimations justified by the law of large numbers and Central Limit Theorem. It is thus required to simulate a sufficient number of paths to ensure the convergence of empirical quantities and thus an accurate computation of the SCR. [dcpedr20] discussed the fact that a number of 1000 paths is a minimum[2]. If some tests still fail, some «reprocessing» may be performed either on input data or on simulated paths (cap, floor, removal of atypical data, etc.) but it should remained exceptional and have to be fully justified. Regarding the data used to calibrate the model, they must be chosen accordingly with the risk profile of the company as much as possible. It is quite straightforward for companies to chose which indices must be replicated. However, for interest rates risk, it is sometimes hard to determine what data should be used (we discuss it below). All in all, a large number (several hundreds) of market data should be replicated explaining why calibration is challenging: not only it has to be precise but also computationally efficient. It is all the more important as those regulatory computations are performed several times a year.

1.3 Introduction to Solvency II regulation

ESGs have multiple uses. Some computations listed above are specific to each insurers but the regulatory ones that are quite standardized and prescribed by the regulators. Present section is dedicated to the description of he legislation and its consequences on ESGs designs.

[2]This number of scenarios may seem as too small to ensure accurate explorations of the possible state of the world. However, due to complexity of the models used by insurers to compute the sequence of cash-flows associated with the insurance policies they have issued, operational constraints impose that the number of paths should be at most a few thousands.

1.3.1 General considerations

Motivated by the financial crisis of 2008, the Solvency II regulation came into effect on 2016, January 1st : it introduces new prudential governing rules of insurance and reinsurance undertakings and harmonizes them at a European level. While former legislation Solvency I proposed to set management rules thanks to an accounting valuation of the insurance companies, Solvency II is based on an *economic valuation* of the balance sheets. The risk profile of each company is now determined taking into account the economic environment through a more granular analysis of their portfolios. Transparency rules are also harmonized and strengthened. Measures contained in the Solvency II regulation are generally gathered in three pillars relative to:

1. Capital requirements: computations of regulatory capital - *Solvency Capital Requirement* (SCR) and *Minimal Capital Requirement* (MCR) - adapted to the risk profile of the company;

2. Risk management: related to governance and asset-liabilities management (ALM) ; implementation of measures to evaluate their consistency through time in a forward-looking approach via the *Own Risk And Solvency Assessment* (ORSA);

3. Disclosure rules: described requirements in terms of communication and reporting.

The studies driven in this thesis mainly relate to pillar 1 (and marginally to pillar 2).

1.3.2 Pillar I: Quantitative Requirements

Solvency II requires insurers and reinsurers to immobilize some of their available cash (also named «Own Funds» in accounting terms) in accordance with their risk profile in order for the company to be able to face, with a probability of 99.5%, a major crisis that could lead the insurer to the incapacity of honouring its financial commitments. To do so, the EIOPA (European regulator) proposed in [Dir09] to value the undertakings from an economic point of view: «*assets shall be valued at the amount for which they could be exchanged between knowledgeable willing parties in an arm's length transaction; liabilities shall be valued at the amount for which they could be transferred, or settled, between knowledgeable willing parties in an arm's length transaction.*». The aim of Solvency II is thus to better estimate the amount of cash the insurer will get if he is led to sell a significant part of its assets or to transfer some of its liabilities during a limited period of time to increase its available liquidity by estimating the exchange value of the company's balance sheet.

In Solvency II, the value of the assets registered in the balance sheet is their *market value*. This quantity is either directly taken from the quoted value on financial market when they are deep, liquid and transparent (*mark-to-market* approach) or computed using mathematical models (coming from bank industry) otherwise (*mark-to-model* approach). Mark-to-model approach requires the prior calibration to market data of some mathematical models to take into account late market conditions. On the liability side, we can distinguish two main types of debts: the one due to policyholder, named the *Technical Provisions* (TP) and the one due to shareholder, named the *Net Asset Value* (NAV).

The Technical Provisions Technical provisions (TPs) represent the economic value of the commitments of the insurer. Due to the variety of risks faced by insurers and their imbrications, this quantity is hard to determine. Risks associated to contracts that may be perfectly replicated by a portfolio composed of financial instruments could be valued using the value of this portfolio and in that case, TPs equal the market value of the replicating portfolio. However, most of

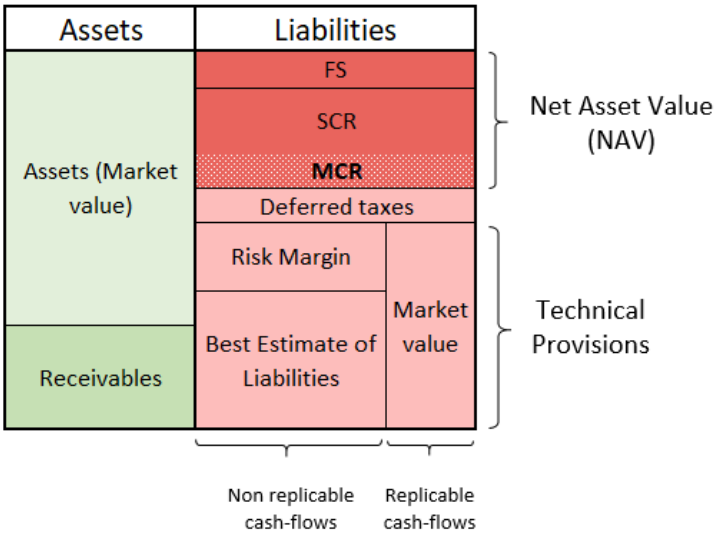


Figure 1.1: Balance sheet under Solvency II legislation.

the insurance contracts can not be perfectly replicated and the assessment of the associated economic value of the liabilities is a complex task. Such liabilities are valued using a *Best Estimate* (BE) approach (see the dedicated paragraph below). On top of that, a *Risk Margin* (RM) is added to the BE as a prudential stock accounting for the non-replicability of some insurance risks:

$$\text{Technical Provisions} = \text{Best Estimate} + \text{Risk Margin}.$$

The Best Estimate: it is defined in [Dir09] as «*the probability-weighted average of future cash-flows, taking into account of the time value of money (expected present value of future cash-flows), using the relevant risk-free interest rate term structure. The calculation of the best estimate shall be based upon up-to-date and credible information and realistic assumptions and be performed using adequate, applicable and relevant actuarial and statistical methods*». Mentioned cash-flows originate from all kinds of risks faced by insurers. They can be split in two categories: financial risks that can be replicated and non-financial ones that are non-hedgeable. In formulas, the BE (also sometimes referred to as BEL standing for *Best Estimate of Liabilities*) expresses as

$$\text{BE} = \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[\sum_{n\geq 1} D(0,n)\text{CF}_n\right] \tag{1.1}$$

where $(D(0,n))_{n\geq 1}$ are discounting factors defined in (1.20) associated to risk-free term structure, $\mathbb{Q}$ is the standard Risk-Neutral probability measure associated to financial risks, $\mathbb{P}$ is the historical probability measure used for other risks (behaviour of policyholders, mortality, longevity, etc.) and CF_n is the cash-flow delivered at time $n \geq 1$ for the n-th period. In practice, the BE is estimated through a Monte-Carlo approach using simulations of future cash-flows:

$$\text{BE} = \lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L}\sum_{n=1}^{N} D^{(l)}(0,n)\text{CF}_n^{(l)}.$$

For a given simulation l, the sequence $(\text{CF}_n^{(l)})_{1\leq n\leq N}$ comprises cash-in and and out-flows so that they can be decomposed as $\text{CF}_n^{(l)} = \text{CF}_n^{(l),\text{out}} - \text{CF}_n^{(l),\text{in}}$ at any time $n \geq 1$. Practitioners

use *Assets Liability Management* (ALM) models to compute sequences of cash-flows: those are complex models that take as inputs *simulated paths of the state of the world* and appreciate the interactions between assets and liabilities of the company through the optional guarantees carried out in the insurance policies issued by the company. The paths on which the computation of cash-flows is based are either built from historical data (mortality or longevity tables) or generated by mathematical models gathered in ESGs. Among them, those dedicated to market risk are coming from bank industry: they are calibrated to market data so that the BE depends on late economic condition. Note also that the computation of the BE allows to integrate national politics scheme as those models integrate operative accounting rules in each country.

Interest rates play a particular role in the computation of Best Estimate. It can be directly seen in the definition (1.1) as the sequence of discount factors only depends on interest rates. The joint distribution of interest rates and other risk factors that intervene in the computation of cash-flows is key to compute the BE. Moreover, among the risk factors themselves the interest rates are prominent as noted in [BGK17] and [Eio14] since 72.1% of the assets portfolio are composed of bonds (31.6% of sovereign and 40.5% of corporate) for the representative portfolio of the Euro zone and 72.3% (29.1% sovereign and 43.2% of corporate) for France[3].

Risk Margin: the RM adds up to the BE to take into account the non replicability of most of insurer's liabilities; «*[it] shall be such as to ensure that the value of the technical provisions is equivalent to the amount that insurance and reinsurance undertakings would be expected to require in order to take over and meet the insurance and reinsurance obligations.*» following [Dir09]. Alternatively, it can be interpreted as the amount the shareholders will have to invest in the company during the years to come to allow the company to pursue its activity in respect of the legislation. Following [uni09] or [CN+14], «*risk margin shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance and reinsurance obligations over the lifetime thereof.*».

In formulas,

$$\mathrm{RM} = r_{CoC} \sum_{t \geq 1} P(0,t)\mathrm{SCR}_t \tag{1.2}$$

where r_{CoC} is the Cost-of-Capital rate, SCR_t is the Solvency Capital Requirement at time t whose definition will be detailed below and $P(0,t)$ is the Zero-Coupon bond price introduced in Paragraph 1.5.1.2 and valued using risk-free rate curve. Note that computation of SCR_t following its description below is only permit for $t = 1$ since for $t > 1$, the exhaustive process would be too heavy computationally speaking. In practice, a number of approximations are made to obtain value of risk margin. The rate r_{CoC} stands for the cost a company would endure for holding an eligible amount of own funds: it is determined by EIOPA and currently set to 6% per year in the statutes.

The Net Asset Value

The Net Asset Value (NAV) represents the available wealth of the company, at the evaluation date, that has been accumulated by the company since the start of its exercise. The NAV corresponds to the shareholder's equity. It can be defined as being the difference between the market value of the assets and the Technical Provisions:

$$\mathrm{NAV} = \text{Market Value of Assets} - (\mathrm{BE} + \mathrm{RM}). \tag{1.3}$$

The NAV comprises the regulatory capital requirements –described below– which is the

[3]Data can be found in Excel file following www.eiopa.europa.eu/content/eiopa-updates-representative-portfolios-calculate-volatility-adjustments-solvency-ii-risk_en.

amount the insurance/reinsurance undertaking must hold during the lifetime of their contracts to exercise their activity, and a Free Surplus (FS) which is the theoretical amount of cash the shareholders could (theoretically) retrieve without impacting the business of the company.

Capital Requirements

The Solvency II legislation requires the insurer to hold a certain quantity of cash during the lifetime of the insurer's commitment. This quantity is the *Solvency Capital Ratio* (SCR). It is part of the available cash of the company. The whole amount of available cash forms what is named the *Own Funds* (OF) of the company and is classified according to the degree of availability of the cash (keep in mind that the final aim of the legislation is to avoid lack of liquidity). The part of Own Funds that is the «more available» (roughly speaking), named EOF (Eligible Own Funds) is used to compute the solvency ratio EOF/SCR. This ratio is an economic indicator useful to deem the solvency of the undertaking: when it is greater than one, the company is considered as being solvable at one year, i.e. being able to fulfil its commitments in the year to come in 99.5% of states of the world. The *Minimum Capital Requirement* (MCR) is the minimum value the NAV can take without intervention of the supervisory authorities and is such that MCR $\leq$ SCR.

Practically, a company endures an economic ruin when its NAV becomes non-positive. The SCR is defined as being the minimum amount the insurance/reinsurance must hold at the date of evaluation so that the probability for the company to endure an economic ruin during the next year is smaller than a 0.5% threshold:

$$\mathrm{SCR} = \inf\{x \in \mathbb{R} : \mathbb{P}(\mathrm{NAV}_1 \leq 0 | \mathrm{NAV}_0 = x) \leq 0.005\}, \tag{1.4}$$

where NAV_0 is the value of the NAV at evaluation date and NAV_1 is the random variable representing the available cash one year later. As the NAV depends on the value of the SCR (see Figure 1.1), the definition (1.4) is implicit. In practice, the SCR is generally computed as a quantile on the one-year loss distribution through a Value-at-Risk (VaR). The one-year loss evaluated at time $t = 0$ is the random quantity defined by $L = \mathrm{NAV}_0 - D(0,1)\mathrm{NAV}_1$. The SCR may alternatively be defined as the 99.5%-quantile of L:

$$\begin{aligned}
\mathrm{SCR} &= \inf_{x\in\mathbb{R}} \Big\{ \mathbb{P}\big(L \leq x\big) \geq 0.995 \Big\} \\
&= \inf_{x\in\mathbb{R}} \Big\{ \mathbb{P}\big(D(0,1)\mathrm{NAV}_1 + (x - \mathrm{NAV}_0) \leq 0\big) \leq 0.005 \Big\} \\
&= \mathrm{NAV}_0 + \inf_{x\in\mathbb{R}} \Big\{ \mathbb{P}\big(D(0,1)\mathrm{NAV}_1 + x \leq 0\big) \leq 0.005 \Big\} \\
&= \mathrm{NAV}_0 + \mathrm{VaR}_{0.5\%}\big(D(0,1)\mathrm{NAV}_1\big) \\
&= \mathrm{NAV}_0 - q_{0.5\%}\big(D(0,1)\mathrm{NAV}_1\big).
\end{aligned} \tag{1.5}$$

where the VaR associated to the random variable X at level $\lambda \in [0,1]$ is defined as $\mathrm{VaR}_\lambda(X) = \inf_{m\in\mathbb{R}} \Big\{ \mathbb{P}(X+m \leq 0) \leq \lambda \Big\}$ and is linked to the quantile function q via $\mathrm{VaR}_\lambda(X) = -\sup_{m'\in\mathbb{R}} \Big\{ \mathbb{P}(X < m') \leq \lambda \Big\} =: -q_\lambda(X)$.

To compute SCR under Solvency II, insurers can chose either to apply the *Standard Formula* or to develop an *Internal Modelling* approach.

The standard formula is a formula prescribed by the regulator. The underlying idea is to decompose the whole risk faced by the company into elementary risks for which the associated SCR is easier to compute. In a second step, those «sub-SCRs» are aggregated to obtain the

total SCR through what is often named the *square-root formula*[4]. The decomposition, following a *bottom-up* approach, is common to all insurers and is examined in more details below.
Alternatively, the internal modelling approach leaves free to the company the choice of the methodology to compute the quantile of the one-year available capital, in line with definition (1.4). It allows for the undertaking to better take into account the specificities of its own risk profile. However, the whole process to compute the value of the SCR with internal models must be approved by the regulator: such an approach requires significant human resources and is thus more suited for large companies.

Standard Formula
The computation of the SCR using the standard formula is built following the bottom-up aggregating approach described in Figure 1.2. Elementary risks (interest rates, mortality, equity, etc.) are gathered in different modulus: market, health, life, non-life, default. The modulus «Intangible» accounts for the risks associated to intangible assets (comprising intellectual properties, licenses, etc.).

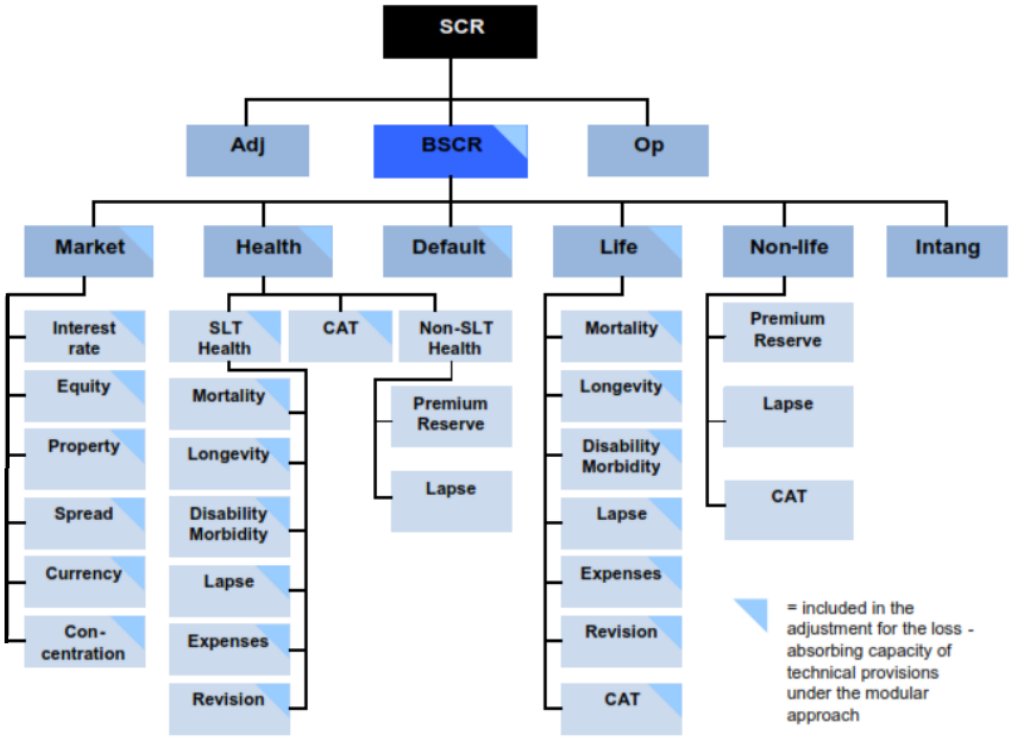


Figure 1.2: Structure of the Standard Formula in Solvency II.

For each elementary risk (interest rates, equity, natural disasters, etc.), an *economic capital* is computed: it represents the sensitivity of the balance sheet (illustrated in Figure 1.1) with respect to a marginal variation in the risk factor associated to the considered elementary risk. Note that as defined in Solvency II, those marginal variations are coherent with historical shocks of each risk factor occurring once every 200 years (99.5% of probability of occurrence under historical measure probability). For instance, the economic capital associated to interest rates is a valuation of the risk resulting from a sudden movement in the yield curve. Economic capital is defined as being the difference between the central value of the NAV - obtained using

[4]The aggregation formula is obtained by assuming the ellipticity of the joint distribution of elementary losses and that of joint distribution of sub-SCRs.

current state of risk factors - and the shocked value of the NAV - obtained when shocking the associated risk factor:

$$\mathrm{EC}_x = \mathrm{NAV}_x^{\mathrm{Central}} - \mathrm{NAV}_x^{\mathrm{Shocked}}$$

for $x \in \{\text{interest rates}, \text{equity}, \text{mortality}, \cdots\}$. Sub-SCRs associated to each modulus are defined by a first aggregation formula

$$\mathrm{SCR}_m = \sqrt{\sum_{(i,j)\in R_m^2} \rho_{i,j}^m \mathrm{EC}_i \mathrm{EC}_j},$$

where R_m is the set of elementary risks in the modulus m and $\rho_{i,j}^m$ is the correlation between elementary risks within modulus m. The correlation matrix $(\rho_{i,j}^m)_{(i,j)\in R_m^2}$ is given by the regulator. Secondly, the aggregation of the sub-SCRs is done in a similar fashion to get the so-called Base SCR (BSCR):

$$\mathrm{BSCR} = \sqrt{\sum_{(m,n)\in M^2} \rho_{m,n}^M \mathrm{SCR}_m \mathrm{SCR}_n},$$

where M is the set of all risks modulus and $(\rho_{m,n}^M)_{(m,n)\in M^2}$ is the correlation matrix between risks modulus, also published by the regulator. The BSCR defined this way does not take into account or the operational risks (human errors, system breakdown, etc.) nor the ability for the insurer to absorb a part of the loss by differing some taxes or by reducing the technical provisions. The operational SCR, denoted by SCR_{op} , and an adjustment are added to the BSCR so that both phenomena are actually integrated in the final SCR:

$$\mathrm{SCR} = \mathrm{BSCR} + \mathrm{SCR}_{op} + \mathrm{Adj}. \tag{1.6}$$

Internal Model

Before going further we give a few insights on the *Internal Modelling* approach. The previous standard formula is prescribed by the regulator, and thus does not take into account the specificities of the insurers portfolios. To compute their solvency requirements, undertakings can alternatively establish their own methodology. Each step of the proposed methodology should be submitted and justified to the regulatory authority, which makes its validation a very long process and requires important human and operational resources.

The so-called 'nested simulations' approach that underlies Internal Models (although not widespread in practice) consists in (see [BRS12]) first simulating under Real-World measure the considered risk drivers over 1 year. At the end of it, calibrations of Risk-Neutral are performed on those projected economic environments. Simulations under Risk-Neutral measure can then be achieved over the lifetime of the liabilities of the company (several decades). The obtained Risk-Neutral paths are used to derive the distribution of the NAV and compute the SCR as previously stated. Again, this process is really cumbersome from computational point of view and a number of approximations is necessary. Other forms of Internal Models can be found that rely on the calibration of a response function based on a sample of points underlying the full distribution, see [FLCM16]. Either way, intensive recalibrations of Risk-Neutral models are generally considered.

1.3.3 Pillar II: ORSA

The second pillar of the Solvency II directive aims at assessing, in a continuous and prospective way, the adequacy between followed strategic directions and the risk profile of the company. An *Overall Solvency Need* is assessed upon which *Enterprise Risk Management* direction rely on. Its

practical implementation mainly consists in projecting the SCR through time up to a relevant time horizon. The procedure depicted above to assess the one-year SCR is computationally heavy and its generalization to a multi-period framework is thus an issue.

Following [VD12], the solvency constraint on a multi-period setting can write

$$\mathbb{P}\left(\bigcap_{t=1}^{T}(\mathrm{NAV}_t \geq 0)\right) \geq \eta \Leftrightarrow \mathbb{P}\Big(\min_{t\in[\![1,T]\!]} \mathrm{NAV}_t \geq 0\Big) \geq \eta$$

where NAV_t denotes the time-t value of the NAV and η is a given threshold in $]0,1[$. The insurer needs to simulate its balance sheet on multiple dates up to a given time horizon. The Net Asset Value is defined in (1.3) as being the difference between market value of the assets and the Best Estimate. The time-t computation of the Net Asset Value requires thus to compute the time-t market value of the assets and the time-t value of the BE:

$$\mathrm{BE}_t = \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[\sum_{s\geq t} D(t,s)\mathrm{CF}_s \big| \mathcal{F}_t\right].$$

The projection of the BE in the future is a complex task. Recall that market value of assets is either directly taken from financial market or derived using mathematical models. For both approach, the market value of the balance sheet thus depends on economic conditions: directly in the first case and through the calibration procedure in the second one. Furthermore, the BE is a function of the interest rates term structure at date of evaluation. Computing value of BE at time $t \geq 1$ requires thus prior simulations of economic market conditions at this date. Those *primary simulations* should be realistic and consistent with historic trends: they are performed under the *Real-World measure* (sometimes also called *Historical measure*, even though they do not rigorously coincide). Once «historical» scenarios have been generated, the financial valuation of the balance sheet can be performed for each state of the world thanks to *Risk-Neutral models*. These are calibrated on primary paths so that calibration is repeatedly executed. Hence the need of efficient calibration methods to lower the computational cost of the whole process. However, this procedure can not be followed to the letter as the number of scenarios to ensure a satisfactory estimation of the BE through time would be too massive. In practice, the projection of the ORSA is completed based on proper approximations made around deterministic projections of the balance sheet. A recent work on the derivation of efficient projection method through time of the BE based on multi-level Monte-Carlo techniques is made in [ACA20a].

Note that, theoretically, the methodology presented here for computations in ORSA, to project the BE over 1 year, is very similar to the one presented above for internal models. However in practice, computations done for ORSA compliance are very different to that related to internal models because of the aforementioned necessary approximations.

1.4 Economic scenario generators step by step

1.4.1 Building the regulatory initial yield curve

We have seen in the preceding paragraphs that the interest rates and, in particular, the risk-free term structure at each valuation date is of importance for the insurance to discount the future cash-flows. The *risk-free* yield curve used for Solvency II compliance is published monthly by the regulator and can be found following: `www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en`. We now discuss how this *risk-free* curve is built.

The risk-free curve is built from instruments quoted on markets that carry various types of market risks (credit, liquidity, etc.). To (ideally) obtain a risk-free curve adapted to insurance specificities, some adjustments should be performed on these market quantities. Before detailing those adjustments, we first discuss what kind of instruments are allowed to be extracted to build the regulatory curve in Solvency II.

1.4.1.1 Choice of the market yield curve

Ideally, those rates should:

1. not carry credit risk;
2. be realistic, meaning that one can actually earn this rate using a «common» investment strategy and without taking tremendous risks;
3. be determined in a reliable and robust way –in particular, the methodology to obtain the rates has to be transparent, justified and permanent over time;
4. be quoted on *Deep, Liquid* and *Transparent* (DLT) markets –more details below;
5. be free of any technical bias. Examples of such bias are given by the International Actuarial Association (IAA) in p. 46 of [Mea09]. Distortions of market prices that can arise at some particular dates (for instance at the end of the year, some unusual market movements can be observed due to upcoming end of the financial year) or due to political decisions are mentioned. Indeed government bond prices can be affected by governments decisions to issue govies or to promulgate legislation that enhance national financial institutions to buy national governments bonds.

This leads to consider largest market rates: swaps or bonds market (as specified in Article 44 of [Reg14]). However swap rates bear credit risk that should be suppressed to fulfil the first aforementioned requirement; conversely, governments bonds (also named govies) can suffer from technical bias as discussed above. In the Euro zone, there exists no institutional bond[5] and it is swap rate[6] whose floating leg is the 6-month EURIBOR which is chosen as basic instrument to derive the relevant risk-free curve. EURIBOR is the rate at which banks in the Euro zone grant loans between themselves (IBOR) for a given duration. This duration for the most exchanged rates goes from 1 day (1D) to 1 year (1Y): we will denote EURIBOR-6M the 6-month EURIBOR. The extracted curve is composed of a finite number of points associated to different maturities.

Fourth requirement needs the selected rates to be traded on DLT markets. A market is said to be *Deep, Liquid and Transparent* or to pass the DLT assessment if the market participants (i) can trade a substantial notional without significantly impacting the quoted prices, (ii) have immediate access to information on quoted prices and ongoing transactions and (iii) the two previous conditions are permanent. The chosen market curve should be analysed from this point of view: do all the maturities in this curve satisfy the DLT assessment? The criterion set by the regulator for a market to be considered as DLT in the Euro zone is the following: at least 10 daily trades per month and 50 000 000 EUR of notional daily traded per month (see [Tec19]). The last DLT revision was performed on end of the year 2020 and the results were the following for the Euro zone: maturities from 1Y to 12Y are all considered as DLT as well as 15Y and 20Y (see [Tec]).

[5] Sometimes anticipated to be named «Eurobond».

[6] See dedicated Section 1.5.1.4 for a precise definition of swap rates.

1.4.1.2 The Credit Risk Adjustment

The basic risk-free interest rate curve is derived from quoted swap rates: consequently, they carry credit risk and thus do not fulfil the first requirement. The *Credit Risk Adjustment* (CRA) aims at removing the credit risk from the yield curve. CRA is a time-independent coefficient defined to be one half of the average spread over one year between the floating leg of the chosen quoted swaps and the *Overnight Indexed Swap* (OIS) of the same tenor. For instance, the risk-free initial term structure for Swiss is derived using swaps whose floating leg refer to the six month IBOR based on the Swiss franc (CHF). Consequently, the 6-month Swiss franc OIS rate is used to compute the CRA. There is an exception to this rule: for the Euro zone, the floating leg of the basic swap rate relates to the EURIBOR-6M, as discussed in the previous paragraph but the CRA is computed using is the spread between the 3-month EURIBOR (EURIBOR-3M) and the Euro OIS rate –named EONIA for Euro OverNight Index Average. OIS rates are IBOR with maturity 1D: it is the rate at which banks grant loans for 1 day between themselves. Those are the rates with the shortest maturity that can be found on markets; consequently, they are generally assumed to be free of credit risk[7]. Furthermore, the CRA is imposed to lie in between 10 basis points (bps ; $1bp = 0.01\%$) and 35 bps. It is applied as a parallel downward shift on all the considered market swap curve. It is computed once and for all, there is no schedule for periodic update. The value of the CRA currently (legislation in effect since January 1st , 2016) prevailing in the Euro zone is of 10bps = 0.1%.

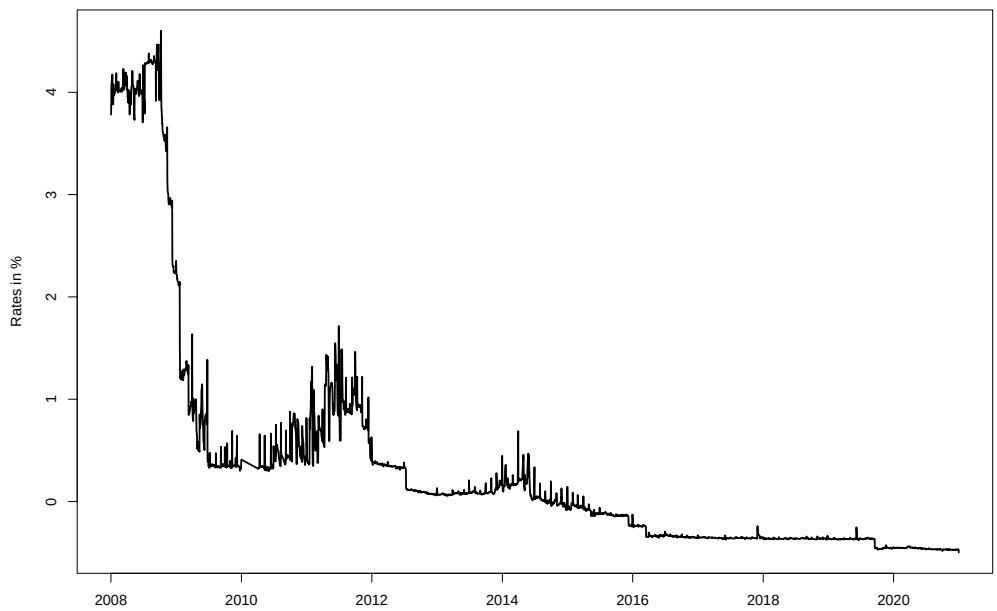


Figure 1.3: Eonia rates between 01/02/2008 and 12/31/2020. Sources: www.emmi-benchmarks.eu.

In formulas, if $(s_{t_n}^{Mkt})_{n=1,\dots,N}$ denotes the extracted swap rates, the CRA is uniformly applied

[7]This is not rigorously true: every transaction is submitted to credit risk however the shortness of the maturity leads to neglect the credit risk.

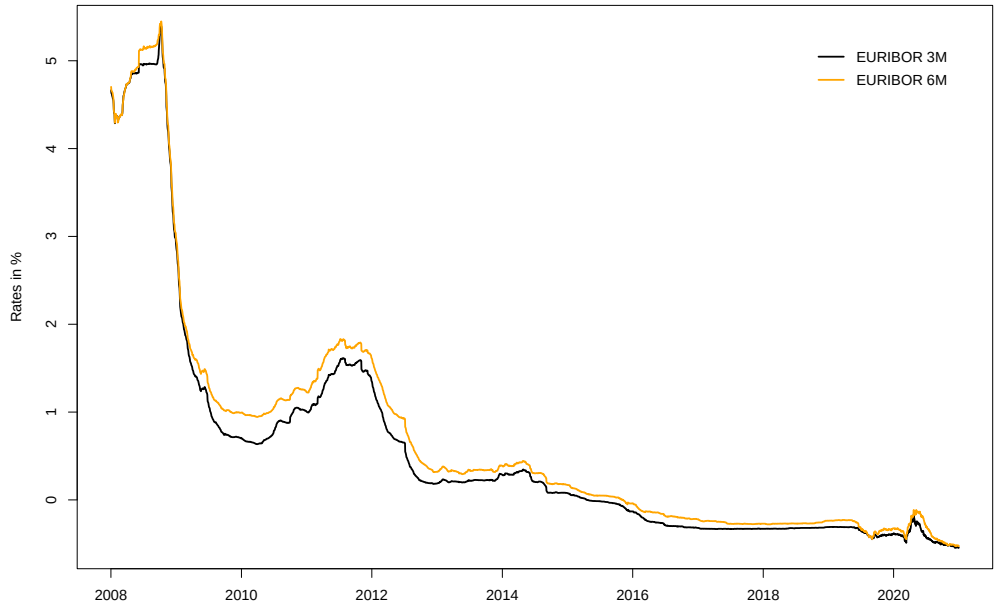


Figure 1.4: 6-month (orange) and 3-month (black) Euribor rates between 01/02/2008 and 12/31/2020. Source: www.emmi-benchmarks.eu.

to obtain the new term-structure $(s_{t_n})_{n=1,\dots,N}$ that will be used in the following:

$$s_{t_n}^{\text{CRA}} = s_{t_n}^{\text{Mkt}} - \text{CRA}, \; n = 1, \dots, N. \tag{1.7}$$

1.4.1.3 The Volatility Adjustment

Once the credit risk is removed, Solvency II aims at removing the other types of risks carried by the market rates. Market quantities (rates or prices) result from continuous rebalancing of portfolios of financial industries, in particular market participants operating on swap rates derivatives. However, those financial industries and participants are exposed to the liquidity risk. To illustrate how liquidity risk can arise, let us consider a market participant P who wants to sell some of its asset(s). In some cases, market players can consider that the assets proposed by P are worthless. Then P may not find a counterpart he could deal with[8]. Furthermore, he may have other engagements with other counterpart(s) and may not have enough liquidity to honour them as long as he has not sell the mentioned assets.
Under Solvency II, it is considered that insurers are not fully exposed to this risk since they keep their assets up to their maturing dates (if they are not compelled to sell it prematurely to honour their commitments). Insurance policies engage insurers on several years: to respect their commitments towards the policyholders, companies will invest in financial assets and hold them up to the expiry of the contract. By introducing the *Volatility Adjustment* (VA), European legislator aims at adjusting the market yield curve so that liquidity risk is suppressed. The VA is still discussed as it is not applied by all insurance legislators (e.g., the Swiss regulator, FINMA, does not take into account this adjustment). Indeed, some events can force insurers to sell their assets to meet their short-term commitments despite mean- and long-term liabilities.

[8]This does not necessarily means that the asset price is close to zero: such event occurs

For instance, if an insurer promises to a policyholder a return rate of 2.9% over 10 years and if he can buy *and keep* a bond paying a coupon rate of 3% over 10 years, then the insurer can hedge himself and earn the spread. This is the ideal situation that is barely met in practice since most of insurance policies comprise a buyout clause meaning that the life duration of insurance contract is unknown. During a crisis, an insurer can thus be led to sell its 3%-paying bonds. Optionality embedded in issued contracts are generally such that companies can not perfectly hedge their positions at the date when contracts are issued: this is why some argue that insurers actually are exposed to liquidity risk.

In the remaining of this section dedicated to the computation of the VA, risk-free rates will refer to the rates from which the CRA has been removed. The VA is computed (i) at currency area level (e.g., Euro zone) and (ii) and a national level. For each point of view, a reference portfolio representing the insurance/reinsurance market of the considered economic area (currency zone or country) is used. The reference portfolios are composed of govies (bonds issued by countries) and corporate bonds (corporates, loans, securities, etc.). The spread between the return rate of the assets composing the reference portfolio and the risk-free rate can be decomposed into two components: the credit risk (default risk) and an illiquidity premium. The latter defines the VA. In other words, the VA includes a *spread* component, which is the difference between return rate of the reference portfolio and the risk-free rate and a *Risk Correction* –also named *fundamental spread*– that accounts for the credit/downgrade risk part carried by the *spread* component.

VA computation

VA computation relies on representative portfolios at national and currency levels. They provide a global overview of the composition of the portfolios of insurers (i) located in a given country (and thus submitted to a common national legislation) and (ii) dealing with the same currency. The composition of these portfolios is published[9] by the regulator and is based on data gathered on insurance markets via regulatory reporting (that is the subject of the third pillar of the Solvency 2 regulation). Those portfolios are split into two parts: bonds (govies and corporates) and other types of financial assets. We provide in Table 1.1 the composition in bonds of some representative portfolios at national and currencies levels. It should be understood in the following sense: the representative part of portfolios of insurers dealing with the Great Britain Pound that is related to bonds is composed of 1% of French bonds, 1% of German bonds and 98% of British bonds; that of insurers located in Hungary comprises 1% of Bulgarian bonds, 1% of Italian and 98% of Hungarian bonds.

Representative portfolio accounting for the currency area (resp. country) of interest is composed of a part $P^{\text{gov,crrncy}}$ (resp. $P^{\text{gov,cntry}}$) of govies and a part $P^{\text{corp,crrncy}}$ (resp. $P^{\text{corp,cntry}}$) of corporate bonds/loans. For $x \in \{\text{gov}, \text{corp}\}$ and $y \in \{\text{crrncy}, \text{cntry}\}$, spread associated to $P^{x,y}$, denoted $S^{x,y}$, is defined as the difference between two distinct *Internal Effective Rates* (IER) –see below for more explanations of the IERs and their computations:

$$S^{x,y} = \text{IER}_{P^{x,y}} - \text{IER}_{\text{risk-free}},$$

where $\text{IER}_{P^{x,y}}$ stands for interest rate that can be earned from cash-flows in the reference portfolio $P^{x,y}$ and $\text{IER}_{\text{risk-free}}$ is the risk-free rate associated to the same portfolio. The Risk Correction (RC) of $P^{x,y}$ is computed as

$$RC^{x,y} = \text{IER}_{P^{x,y}} - \text{IER}_{\text{corrected}}$$

[9] www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures-0_en.

Currency		National	
EUR	4% Austria, 9% Belgium, 22% Italy, 1% Finland, 32% France, 9% Spain, 14% Germany, 2% Ireland, 1% Poland, 1% Luxembourg, 3% Netherlands, 1% Portugal, 1% Slovakia.	France	3% Austria, 5% Belgium, 72% France, 3% Germany, 1% Ireland, 7% Italy, 1% Netherlands, 1% Portugal, 6% Spain, 1% US.
GBP	1% France, 1% Germany, 98% UK.	UK	1% Belgium, 3% France, 4% Germany, 1% Netherlands, 77% UK, 10% US, 1% Japan, 1% Australia, 2% Canada.
HUF	100% Hungary.	Hungary	1% Bulgaria, 98% Hungary, 1% Italy.

Table 1.1: Weights of the classes of assets composing the reference portfolios.

where $\mathrm{IER}_{\mathrm{corrected}}$ is the corrected interest rate generated by the cash-flows of $P^{x,y}$. Computation of the two VA components is done as follows: for $y \in \{\mathrm{crrncy}, \mathrm{cntry}\}$

$$\mathrm{VA}_y = 0.65 \times (S_y - RC_y), \tag{1.8}$$

with

$$\begin{cases} S_y &= w_y^{\mathrm{gov}} \max\left(S^{\mathrm{gov},y}, 0\right) + w_y^{\mathrm{corp}} \max\left(S^{\mathrm{corp},y}, 0\right), \\ RC_y &= w_y^{\mathrm{gov}} \max\left(RC^{\mathrm{gov},y}, 0\right) + w_y^{\mathrm{corp}} \max\left(RC^{\mathrm{corp},y}, 0\right), \end{cases}$$

where w_y^{gov} (resp. w_y^{corp}) denotes the portion of govies (resp. corporate bonds) in the considered reference portfolio. The coefficient 65% in the definition of the VA (1.8) is statutory (see [Tec19]). Representative portfolios are mainly composed of bonds but not uniquely as they comprise other types of financial assets which explains that $w_y^{\mathrm{gov}} + w_y^{\mathrm{corp}} < 1$. These weights are published by the regulator; their values on December 2020 are given in Table 1.2 to fix the ideas.

Euro zone (Currency)		France (Country)	
$w_{\mathrm{crrncy}}^{\mathrm{gov}}$	31.6%	$w_{\mathrm{cntry}}^{\mathrm{gov}}$	29.1%
$w_{\mathrm{crrncy}}^{\mathrm{corp}}$	40.5%	$w_{\mathrm{cntry}}^{\mathrm{corp}}$	43.2%

Table 1.2: Weights of the classes of assets composing the reference portfolios.

Finally, the total VA coefficient is given by:

$$\mathrm{VA} = \begin{cases} 0.65\left(S_{\mathrm{crrncy}}^{RC} + \max(S_{\mathrm{cntry}}^{RC} - 2S_{\mathrm{crrncy}}^{RC}, 0)\right), \text{ if } S_{\mathrm{crrncy}}^{RC} \geq 1\% \\ 0.65 S_{\mathrm{crncy}}^{RC}, \text{ if } S_{\mathrm{crrncy}}^{RC} < 1\%, \end{cases}$$

where $S_y^{RC} = S_y - RC_y$ for $y \in \{\mathrm{crrncy}, \mathrm{cntry}\}$.

We now describe how the Internal Effective Rates are computed. We assume that the cashflows delivered by the bonds can be approximated by a single cash-flow. This assumption is acceptable when return rates of each bond are relatively small and flat as justified thereafter.

Namely, we assume that having one bond of duration T_D delivering K cash-flows $(c_k)_{k=1,\dots,K}$ is equivalent to have one Zero-Coupon (ZC) bond of maturity T_D whose final payment is $\sum_{k \leq K} c_k$. On the one hand, consider the spot value of a bond delivering $(c_k)_{k=1,\dots,K}$ cash-flows on dates $(t_k)_{k=1,\dots,K}$ discounted with the rate r (time-independent since it has been assumed to

be flat) which satisfies, by definition,

$$V_0 = \sum_{k \leq K} \frac{c_k}{(1+r)^{t_k}}.$$

On the other hand, consider the spot value of the ZC bond delivering $\sum_{k \leq K} c_k$ at time $T_D = \sum_{k \leq K} t_k \frac{\frac{c_k}{(1+r)^{t_k}}}{\sum_{j \leq K} \frac{c_j}{(1+r)^{t_j}}}$ –that is the duration of the above considered bond– which is given by

$$\tilde{V}_0 = \sum_{k=1}^{K} \frac{c_k}{(1+r)^{T_D}}.$$

The difference of value writes thus as $V_0 - \tilde{V}_0 = \sum_{k=1}^{K} c_k \left(\frac{1}{(1+r)^{t_k}} - \frac{1}{(1+r)^{T_D}} \right) \approx r \sum_{k=1}^{K} c_k \left(T_D - t_k \right)$. Assuming rates are small ($r << 1$), $e^{-rt_k} \approx 1$ and the formula giving the duration yields that

$$V_0 - \tilde{V}_0 \approx 0.$$

The reference portfolio composed of multiple bonds of different maturities is thus assimilated to a portfolio composed of ZC bonds each delivering one unique cash-flow at different dates. Denote by V_0^{pf} the value of the reference portfolio (either composed of govies or corporate bonds) and by $\omega_n = P(0, T_n)/V_0^{pf}$ the portion of T_n-ZC bonds in the equivalent portfolio. The associated cash-flow at time T_n is equal to its proportion in the portfolio capitalized with the return rates associated with the bonds composing the reference portfolio: denote by $\mathrm{CF}_1, \ldots, \mathrm{CF}_n$ the cash-flows delivered on dates $T_1, \ldots, T_n$ capitalized with rates $(y_i)_{i=1,\ldots,n}$ so that $\mathrm{CF}_i = \omega_i (1+y_i)^{t_i}$, for $i = 1, \ldots, n$. The IER of the equivalent portfolio is defined to be the unique rate satisfying

$$V_0^{pf} = \sum_{i=1}^{n} \frac{\mathrm{CF}_i}{(1+\mathrm{IER})^{t_i}}.$$

By inverting this relationship, one is able to derive the value of the different IERs. They depend on the way to capitalize the cash-flows:

i. if the i-th cash-flow is valued using the return rate of the T_i-bond, i.e. $y_i = r_i$, the output IER is that of the reference portfolio: $\mathrm{IER}_{P^{x,y}}$ is obtained;
ii. if the i-th cash-flow is valued using the risk-free rate, i.e. $y_i \equiv r$, the $\mathrm{IER}_{\text{risk-free}}$ is obtained;
iii. if the i-th cash-flow is valued using the return rate of the T_i-bond corrected of the default/downgrade rate, i.e. $y_i = r_i^c$, the $\mathrm{IER}_{\text{Corrected}}$ is obtained; to subtract the default/-downgrade risk in the return rate, the following procedure is applied:
    - if dealing with govies portfolio, the corrected rate is $r_i^c = r_i - 0.30 \times \max(\mathrm{LTAS}_{gov}, 0)$ where LTAS_{gov} is the Long-Term Average Spread over the risk-free rate of assets of the same duration, same credit quality and asset class;
    - if dealing with corporate portfolio, the corrected rate is $r_i^c = r_i - \max(0.35 \times \mathrm{LTAS}_{corp}, \mathrm{PD}_i + \mathrm{CoD}_i)$ where LTAS_{corp} is the Long-Term Average Spread over the risk-free rate of assets of the same duration, same credit quality and asset class; PD_i is the credit spread corresponding to the probability of default on the i-th risky assets and CoD_i is the cost of downgrade that is the expected loss subsequent to a downgrading of the i-th risky asset (when considering a buy and replace strategy).

The proportions 30% and 35% applied to the LTA spread are fixed by the regulator.

We conclude this section on the VA computation by saying a word on the computation of the LTA spread, PD and CoD appearing in previous third bullet point.

- The Long Term Average spread is an arithmetic mean of daily quoted spreads over the last 30 years. When dealing with govies, the missing data are replaced by the flat average obtained with available data; when dealing with corporate bonds, a linear interpolation (with step time equals a 10th of a year) is performed between available maturities and extrapolated flat before the first available maturity.
- PD_i and CoD_i are rates that should be added to risk-free rate to take into account fact that corporate bonds can either default or downgrade (upgrades or stays are not considered as risks). They are determined using historical one-year transition matrix and risk-free term structure. They are defined to be implicit rate so that the discount factor over the whole period of time projection (typically, 30 years) equals the risk-free discount factor over the same period from which the successive discounted costs of downgrade/default has been subtracted. These costs are defined to be the difference between the market values of the risky bonds before and after the downgrade/default event while considering a buy and replace strategy. Both probabilities of defaults and cost of downgrades are published by the regulator for each class of assets.

The value of VA in December 2020 is of 7 bps. Previously, we have discussed how to obtain the curve composed of swap rates integrating s^{CRA} –see (1.7). Assume for a moment we have a continuous swap rates curve. Those are translated into ZC rates, denoted $(r_t^{\text{CRA}})_{t\leq T}$, to which the VA is applied as a parallel upward shift:

$$r_t^{\text{VA}} = r_t^{\text{CRA}} + \text{VA},\ t \leq T.$$

For some regulatory computations (notably, the Risk-Margin), a curve without VA is also supplied by the regulator:

$$r_t = r_t^{\text{CRA}},\ t \leq T.$$

The interested reader can refer to the EIOPA's documentation [Tec19]. The following section is dedicated to the derivation of a continuous ZC rates curve.

1.4.1.4 Interpolation/extrapolation of market data

The purpose of this section is to describe how discrete curves are transformed into continuous term structures through an interpolation/extrapolation method. This interpolation/extrapolation process should be performed consistently with the regulatory constraints listed in the beginning of Section 1.4.1.1. The term structure obtained as output of this process is the *regulatory curve*. In practice, EIOPA publishes[10] on a monthly basis the regulatory curve.

Ultimate Forward Rate (UFR) and Last Liquid Point (LLP)

Maturities encountered in insurance policies are much longer than those implied in inter banks contracts: the extracted market curve has to be extrapolated to very long-term maturities that are not quoted on the market (usually, up to 120 or 150 years). In addition, most of the used interest rates models generates continuous yield curves. To consolidate calibration of the interest rates model, a replication of the initial term structure using a high number of maturities would be helpful; hence the necessity of interpolating market curve.

The interpolation/extrapolation process requires a prior assessment of firstly a point (i.e. a maturity) disassociating the interpolation from the extrapolation and secondly a value toward

[10] Also following www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en.

which the *extrapolated one-year forward curve* should converge. These quantities are respectively named the *Last Liquid Point* (LLP) and the *Ultimate Forward Rate* (UFR). We stress that the UFR concerns the asymptotic value of *one-year forward rates* whereas extracted market data are *a priori* not forward rates. The UFR is thus applied to forward rates deduced from market data and not directly on them.

Legislation defines a precise criterion for the determination of the LLP which is the largest maturity among those satisfying the DLT assessment and beyond which the market can no longer be considered as DLT. It is the so-called *residual volume criterion* and states that the cumulated volume of bonds with maturities greater than or equal to the relevant maturity should be smaller than a given portion of the total volume of traded bonds. The legislation currently sets this threshold to 6% (see [Tec19]). A similar residual volume criterion is applied to traded swaps for the Euro zone only and results in a LLP set to 20 years. For other currencies, the LLP is taken being equal to the last quoted maturity satisfying the DLT assessment.

The *Ultimate Forward Rate* plays a crucial role in the legislation as it determines the level of the second part of the regulatory curve. The UFR is defined as being the sum of two terms: the expected real rates that is the same for all currencies, and the expected inflation rate based on the inflation target of central banks and is thus currency-dependent. First term is defined as a basic average of past real rates observed since 1961 across Belgium, Germany, France, Italy, Netherlands, UK and US. The second term takes values in $\{1\%, 2\%, 3\%, 4\%\}$. The inflation target for the Euro zone is set to 2%[11] by the European Central Bank (ECB). Written in the end of 2000 decade, Solvency II legislation initially set the UFR equals to 4.20% for Euro (also valid for British pound and US dollar whereas a value of 3.2% is set for Japanese yen). Through the 2010 decade, interest rates markets have considerably vary (downward) and UFR value has been repeatedly updated: in 2017, new theoretical UFR is obtained to be equal to 3.65%; in 2020, new assessment provides the value of 3.55%. However, following the legislation requirement, annual change in UFR are limited to 15 bps: on January 1st 2020, the effective UFR value was of 3.75%; as of January 1st 2021, the effective UFR value is of 3.60%.

Before moving onto the details of the extrapolation method, two quantities still must be chosen: the Convergence Point (CP) and the tolerance threshold. Namely, the convergence point is the date at which the UFR is reached within the prescribed tolerance threshold. The convergence point is set to $\max(\text{LLP}+40, 60)$ years, so that the convergence period is $\max(60-\text{LLP}, 40)$ and the tolerance is of 1 bp. The parameter monitoring the convergence speed in the extrapolation method (denoted by α in the regulator documentation) is chosen to be the lowest value allowing to reach the tolerance level of the UFR by the convergence point.

Extrapolation/interpolation: the Smith-Wilson approach

The Smith-Wilson methodology enables to interpolate a given discrete curve while taking into account some long-term constraints in the extrapolation part. In particular, it allows to ensure a convergence to a given target. In addition, the smoothness of the obtained curve can be parametrized. Definition (1.27) of the instantaneous forward rate at time 0 of maturity $u \geq 0$ and denoted $f(0, u)$ directly implies that spot ZC price of maturity $T \geq 0$ can be written as

$$P(0, T) = \exp\left(-\int_0^T f(0, u)\mathrm{d}u\right) := P(T).$$

[11] See www.ecb.europa.eu/mopo/html/index.en.html.

We define the *yield intensity function* $T \mapsto y(T)$ as being the flat rate over the period $[0, T]$ giving the same value to the spot bond:

$$P(0,T) = \exp\left(-Ty(T)\right) \Longleftrightarrow y(T) = \frac{1}{T}\int_0^T f(0,u)\mathrm{d}u.$$

The Smith-Wilson approach, as described in [SW01] and adapted to the present context, proposes a model for the discount factor function $T \mapsto P(0,T)$ from which the yield intensity function and forward rates are deduced. To describe the Smith-Wilson method in a more general context, let us denote by $T \mapsto P(T)$ the present value of a basket composed of N fixed income instruments whose values are known initially; so is the value of the basket. Smith and Wilson proposed to model the present value of the basket as

$$P(T) = e^{-\omega T} + \sum_{j=1}^{N} \zeta_j \left(\sum_{k=1}^{J} c_{j,k} W_\alpha(T, u_k) \right), T \geq 0, \tag{1.9}$$

where the $(u_j)_{j=1,\ldots,N}$ are the tenors of the N involved financial instruments, $c_{j,k}$ is k-th cashflow of the j-th financial instrument, $\omega = \log(1 + UFR)$ is such that the first term in the preceding equation corresponds to a simply discounting with the UFR, $(u,v) \mapsto W_\alpha(u,v)$ is the Wilson function defined by $W(u,v) = e^{-\omega(u+v)}\Big(\alpha \min(u,v) - \exp\big(-\alpha \max(u,v)\big) \times \sinh\big(\alpha \min(u,v)\big)\Big)$, α is the parameter monitoring the convergence speed as discussed in the end of the preceding paragraph and $(\zeta_j)_{j=1,\ldots,N}$ are coefficients to be determined during calibration procedure. The Wilson function is thus at the core of the method: we illustrate it in Figure 1.5 for different values of α. Observe that it is a symmetric function: $W_\alpha(u,v) = W_\alpha(v,u)$. As mentioned, the parameter α controls the speed of convergence of the interpolated/extrapolated curve that is built. The implicit assumption is that the Wilson function is a converging function, which is also observed (and could be theoretically proved).

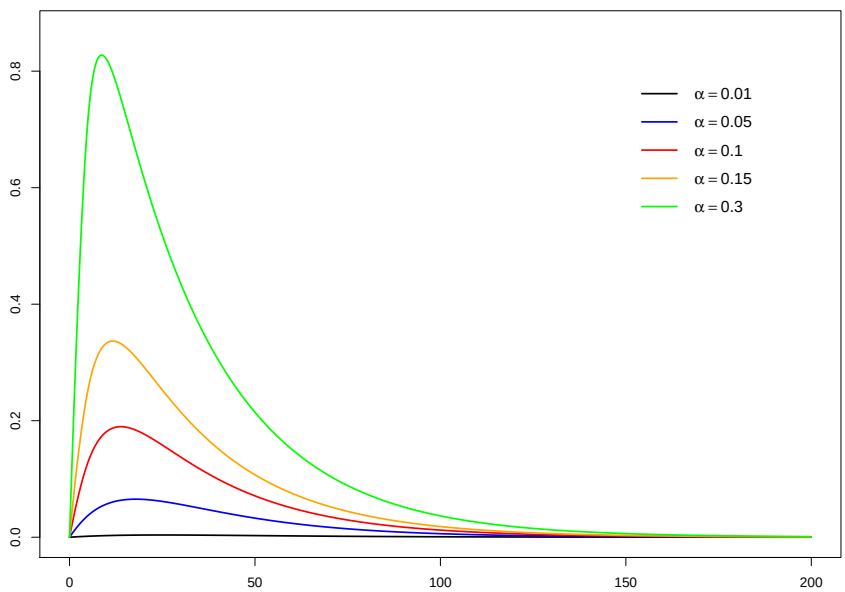


Figure 1.5: Wilson function $v \mapsto W_\alpha(u,v)$ for different values of α where $u \equiv 5$ has been set.

Let us now move onto the details of the Smith-Wilson method. Let us denote by $\boldsymbol{\zeta} = {}^t(\zeta_1, \dots, \zeta_N)$ the parameter vector and $\boldsymbol{u} = {}^t(u_1, \dots, u_N)$ the vector composed of the tenors of selected interest rates derivatives (where ${}^t\boldsymbol{x}$ represents the transpose vector of $\boldsymbol{x}$). Let $\boldsymbol{C}$ be the $N \times J$ matrix of the cash-flows delivered by the considered financial instruments: the j-th row of this matrix is composed of the N cash-flows associated to the j-th financial instrument. Let $\boldsymbol{d} = {}^t(e^{-\omega u_1}, \dots, e^{-\omega u_N})$ and defined by $\boldsymbol{D}^\star$ the unique diagonal matrix of size N such that $\boldsymbol{D}^\star \cdot \mathbf{1} = \boldsymbol{d}$ where $\mathbf{1}$ is the vector whose components are all equal to 1. Let us denote by $\boldsymbol{p}^{\text{Mkt}} = {}^t(p_1, \dots, p_N)$ the vector of observed market prices of the N financial instruments composing the basket and $\boldsymbol{p}^{\text{Mod}} = {}^t(P(u_1), \dots, P(u_N))$ the vector composed of model prices coming from (1.9). The purpose of the calibration is to select the set of parameters $\boldsymbol{\zeta}$ ensuring that $\boldsymbol{p}^{\text{Mkt}}$ and $\boldsymbol{p}^{\text{Mod}}$ are close enough. We will denote by $\boldsymbol{W}_\alpha(v, u) = {}^t\big(W_\alpha(v, u_1), \dots, W_\alpha(v, u_N)\big)$ and by $\boldsymbol{W}_\alpha = \boldsymbol{W}_\alpha(u, u)$ the matrix of coefficients $\boldsymbol{W}_\alpha(u, u)_{i,j} = \boldsymbol{W}_\alpha(ui, uj)$. We define the matrix $\boldsymbol{H}_\alpha$ as the representation matrix of the function H_α: $\big(\boldsymbol{H}_\alpha\big)_{i,j} = \alpha \min(u_i, u_j) - \exp\big(-\alpha \max(u_i, u_j)\big) \times \sinh\big(\alpha \min(u_i, u_j)\big) =: H_\alpha(u_i, u_j)$; observe that $\boldsymbol{W}_\alpha = \boldsymbol{D}^\star \boldsymbol{H}_\alpha \boldsymbol{D}^\star$.

Rewriting the parametric shape (1.9) with the introduced quantities provides

$$P(t) = e^{-\omega t} + \boldsymbol{W}_\alpha(t, \boldsymbol{u})\boldsymbol{C}\boldsymbol{\zeta} = e^{-\omega t}\big(1 + \boldsymbol{H}_\alpha(t, \boldsymbol{u})\boldsymbol{D}^\star \boldsymbol{C}\boldsymbol{\zeta}\big). \tag{1.10}$$

This relationship has to hold true for all $t \in u_1, \dots, u_N$ so that we can write in matrix form:

$$\boldsymbol{p}_\alpha^{\text{Mod}} = \boldsymbol{d} + \boldsymbol{W}_\alpha \boldsymbol{C}\boldsymbol{\zeta}. \tag{1.11}$$

Furthermore, the market value of a fixed income instruments delivering cash payments equals the sum of its discounted cash-flows. Thus the i-th market price $\boldsymbol{p}_i^{\text{Mkt}}$ should be such that

$$\boldsymbol{p}_i^{\text{Mkt}} = \sum_{j=1}^{J} c_{i,j} P(u_j),$$

where P is the theoretical parametrized present value shape of Smith-Wilson (1.9). It can be written in matrix form as ${}^t\boldsymbol{C}\boldsymbol{p}_\alpha^{\text{Mod}} = \boldsymbol{p}^{\text{Mkt}}$. With (1.11), it yields ${}^t\boldsymbol{C}\boldsymbol{p}_\alpha^{\text{Mod}} = {}^t\boldsymbol{C}\boldsymbol{d} + {}^t\boldsymbol{C}\boldsymbol{W}_\alpha\boldsymbol{C}\boldsymbol{\zeta} = {}^t\boldsymbol{C}\boldsymbol{d} + {}^t\boldsymbol{Q}\boldsymbol{H}_\alpha\boldsymbol{Q}\boldsymbol{\zeta}$ where we have introduced $\boldsymbol{Q} = \boldsymbol{D}^\star\boldsymbol{C}$. Since we have imposed ${}^t\boldsymbol{C}\boldsymbol{p}_\alpha^{\text{Mod}} = \boldsymbol{p}^{\text{Mkt}}$, we can identify the optimal parameter vector as

$$\boldsymbol{\zeta}_\alpha^* = \Big({}^t\boldsymbol{Q}\boldsymbol{H}_\alpha\boldsymbol{Q}\Big)^{-1}(\boldsymbol{p}^{\text{Mkt}} - {}^t\boldsymbol{C}\boldsymbol{d}).$$

The optimal $\boldsymbol{\zeta}_\alpha^*$ depends on the value of the convergence speed parameter α. So far, its value has not been determined: α is considered as a meta-parameter of the model. To calibrate it we plug the optimal vector $\boldsymbol{\zeta}_\alpha^*$ into (1.10) for any time $t > 0$. The instantaneous forward rate implied by the optimal vector of parameters writes

$$f(0, t) = -\frac{\partial \log P(T)}{\partial T}\Big|_{T=t} = \omega - \frac{\boldsymbol{G}_\alpha(t, \boldsymbol{u})\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^*}{1 + \boldsymbol{H}_\alpha(t, \boldsymbol{u})\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^*},$$

where $\boldsymbol{G}_\alpha$ is the matrix representation of the function G_α defined as the derivative of H_α:

$$G_\alpha(u, v) = \frac{\mathrm{d}H_\alpha(u, v)}{\mathrm{d}v} = \begin{cases} \alpha\big(1 - e^{-\alpha u}\cosh(\alpha v)\big), & v \le u, \\ \alpha e^{-\alpha v}\sinh(\alpha u), & u \le v. \end{cases}$$

Following the same notation as above we have, for $v \ge \max(u_1, \dots, u_N)$, $H_\alpha(\boldsymbol{u}, v) = \alpha\boldsymbol{u} - e^{-\alpha v}\sinh(\alpha\boldsymbol{u})$ and $G_\alpha(\boldsymbol{u}, v) = \alpha e^{-\alpha v}\sinh(\alpha\boldsymbol{u})$ where the hyperbolic sinus function is applied component-wise

to the vector $\alpha \boldsymbol{u}$. It follows that $f(0,t) = \omega + \frac{\alpha}{1-\kappa e^{\alpha t}}$ where $\kappa = \frac{1+\alpha^t \boldsymbol{u}\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^*}{\sinh(\alpha^t \boldsymbol{u})\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^*}$ only depends on α and ω. The convergence at point $U \geq \max(u_1, \ldots, u_N)$ can be measured by the function gap as

$$g(U) := |f(0,U) - \omega| = \frac{\alpha}{|1 - \kappa e^{\alpha U}|}.$$

For the asymptotic value of the extrapolated shape to be within 1 bp interval around the target as of the convergence point CP, α should be such that

$$g(\mathrm{CP}) \leq 0.0001,$$

where CP is the convergence point. The determination problem for α can be tricky as the behvaior of gap function g depends on the inputs, as observed in ([LL16]). Indeed, $|1-\kappa e^{\alpha U}| = 0 \Leftrightarrow 1 + \Big(\alpha^t \boldsymbol{u} - e^{-\alpha U}\sinh(\alpha^t \boldsymbol{u})\Big)\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^* = 0$. Furthermore, (1.10) can alternatively be written as

$$P(U) = e^{-\omega U}\Big(1 + (\alpha^t \boldsymbol{u} - e^{-\alpha U}\sinh(\alpha^t \boldsymbol{u}))\boldsymbol{Q}\boldsymbol{\zeta}_\alpha^*\Big).$$

Thus the gap function has singularities if and only if model prices reach zero which can happen depending on the input prices.

In practice, to determine α one solves the following minimization problem:

$$\min \alpha$$
$$\text{under constraints (i) } \alpha \geq 0.05 \text{ (ii) } g(\mathrm{CP}) \leq K,$$

for some upper boundary $K = 0.0001$ set in the legislation.

Let us now sum up how the VA is integrated in practice for building the Euro curve (this approach may be different for other currencies): the discrete curve $(r_{t_n})_{n=1,\ldots,N}$ (where t_N is the maturity of the LLP –see preceding paragraph for details on the Last Liquid Point) integrating the CRA coefficient (see Section 1.4.1.2), is interpolated/extrapolated to get a smooth curve of ZC rates $(r_t)_{t\geq 0}$ without VA applying the Smith-Wilson method. On this obtained curve, the VA is applied as a parallel upward shift on maturities before that of the LLP (i.e. t_N) to get

$$r_t^{\mathrm{VA}} = r_t + \mathrm{VA},\ t \leq t_N,$$

and the Smith-Wilson extrapolation method is applied to that upward shifted curve to get $(r_t^{\mathrm{VA}})_{t\geq 0}$ also converging to the UFR. While the VA is a non time-dependent coefficient, a continuous term structure of it is sometimes defined as the spread between these two curves:

$$\mathrm{VA}_t = r_t^{\mathrm{VA}} - r_t,\ t \leq 0.$$

Consequently, $\mathrm{VA}_t = \mathrm{VA}$ for $t \leq t_N$ and $\lim_{t\to+\infty} \mathrm{VA}_t = 0$ since both curve are extrapolated towards the UFR.

In Figure 1.6, regulatory curves with and without VA on December 2020 are plotted: the small value of the VA is such that the two curves almost coincides. In Figure 1.6, the market curve is compared to the EIOPA one (without VA) for maturities between 1 to 50. The impact of the UFR is significant beyond 20 years. In particular, the market curve remains under 0.0% for all term structure. Solvency II aims at better estimating the risk profile of undertaking by taking into account the economic environment. However in view of Figure 1.7, the question of the appropriateness of the financial value of the balance sheet derived using regulatory curve rises.

Finally, it is important mentioning that discussions are currently led by EIOPA to update

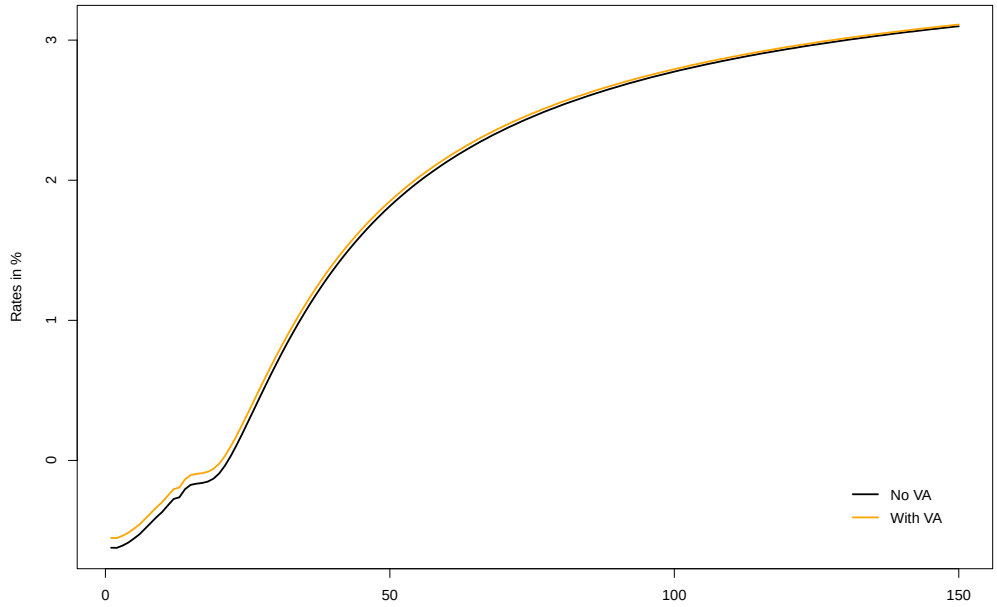


Figure 1.6: Regulatory risk-free rates curves as of December 2020. UFR = 3.60%, convergence parameter in Smith-Wilson method $\alpha = 0.136091$, LLP = 20, convergence period is 40 years, VA = 7 bps and CRA = 10 bps.

methodologies notably regarding the setting of the VA and the interpolation/extrapolation method. Regarding the definition of the VA issues have been identified in the way the representative portfolios are aggregated. Suggestions have been brought in order to rectify it and *"to reflect undertakings' specificities in the VA"*. Several issues are also discussed regarding the interpolation/extrapolation method, notably the underestimation of technical provisions it induces and an alternative one has been proposed. Details can be found following www.eiopa.europa.eu/sites/default/files/solvency_ii/eiopa-bos-20-749-opinion-2020-review-solvency-ii.pdf and www.eiopa.europa.eu/content/consultation-paper-opinion-2020-review-of-solvency-ii_en.

Remark 1. *To overcome the illiquidity of large maturities rates derivatives and thus the lack of information provided by financial markets, building of yield curve comprising very large (even infinite) maturities has been proposed in [EKMH14] based on Ramsey rule.*

1.4.2 Choosing the models

The choice of the models to be integrated in ESGs depend on multiple criterion. First, the insurers select models only for the risks that are aimed at being modelled; since interest rates and equity risks are major for the companies, they are systematically modelled. Regarding other kind of risks, there is no imperative for the companies to model all of them (real interest rates, real estates, inflation, credit and foreign exchange). Secondly, for the considered risks, the choice of the model depends on the (i) stylized facts observed on historical series that are wanted to be reproduced by the model in Real-World ESGs or (ii) the market data to replicate for Risk-Neutral ESGs.

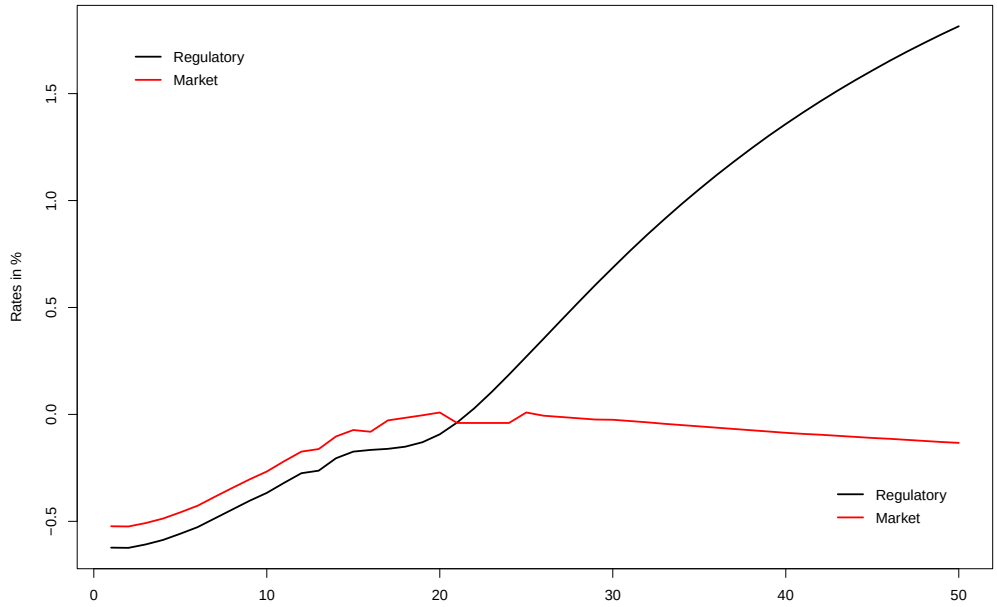


Figure 1.7: Regulatory (without VA) and market term structure (composed of ZC rates) as of December 31st, 2020. Maturities 31 to 39 and 41 to 49 are not quoted and have been built using a linear interpolation method.

We give some examples related to the first case: interest rates model have lately been required to be able to generate negative rates; they are asked to generate non-perfect correlation between rates of different maturities; models for rates or indices can be asked to reproduce empirical standard deviations. In a forward-looking approach, some models may be designed in order to reach preassigned targets: for instance, real rates models may be thought to reach expected inflation levels. Moreover, the ease of use, the efficiency of the calibration procedure, the work necessary for updating the model or the interpretability of the model parameters are the main concerns driving the choice of the undertakings for their models.

In second case, the accuracy with which the market data are replicated, the computational time dedicated to the calibration process are central when choosing Risk-Neutral models. For instance, some models are not able to replicate Away-From-the-Money (AFM) derivatives prices. The efficiency of the calibration process mainly relies on the ability to derive closed-form expressions of prices of derivatives in the considered model. The ease of use, the efficiency of the calibration procedure, the work necessary for updating the model or the interpretability of the model parameters are still of importance in Risk-Neutral universe.

1.4.3 Calibration on market data

We focus again on nominal interest rates models. In addition to the initial term structure, one needs prices of derivatives to calibrate the interest rates models. We now discuss how these instruments can be chosen by insurers in practice though the rigorous justification of the selected products is not always possible in view of the variety of the composition of insurers' portfolios.

1.4.3.1 Calibration instruments

We start by discussing what kind of interest rates derivatives should be considered to calibrate the associated models.

Types of financial derivatives
The final choice of the calibration instruments is left to the discretion of the undertaking. It turns out that a wide majority of insurers (if not all) chose to replicate swaptions prices. Several reasons justify this choice:

- swaptions are simple and liquid options on interest rates;
- swaption market offers a wide range of instruments (long maturities, various strikes): there are more quoted swaptions than quoted caps/floors;
- swaptions are deemed able to capture the joint distribution of forward rates (in particular, the terminal correlation between them should be accessible, as discussed in [BM07], Section 6.6).

This last point is motivated by fact that companies (generally) invest in bonds of different maturities when issuing one insurance contract. Insurers aims thus at capturing the correlation between forward rates of distinct expiries.

In practice, calibration procedures are *generally* based on volatilities of ATM swaptions giving the right to enter a swap whose fixed leg is EURIBOR-6M (semi-annual payments) of maturities $M \in \{1, \dots, 10, 15, 20, 25, 30\}$, tenors $T \in \{1, \dots, 10, 15, 20, 25, 30\}$ and OTM swaptions of same maturities, tenor $T = 10$ and relative strikes $k \in \pm\{0.25\%, 0.5\%, 1\%, 1.5\%, 2\%\}$.

Choice of maturities and strikes
Ideally, insurers choose the maturities/tenors and strikes to be replicated consistently with the characteristics of their liabilities (duration, optionality levels, ...). However, contrary to financial industries, there is no clear hedging strategy for insurers. First because insurers are exposed to non hedgeable risks, as previously discussed. Secondly, one issue concerns the nature of the involved products: *insurance policies* –found in the liability side– are of different nature from that of *financial derivatives* that are selected to calibrate the models. An analogy between insurance policies and financial instruments can be made to motivate the selection. We retake the example found in [BC16] of a Euro contract with a minimum guaranteed rate and a profit sharing clause. Let us denote by r_g the continuously compounded minimum guaranteed rate, by r_{share} the profit sharing rate, by r_{fees} the rate associated to fees, r_d is the effectively delivered rate to the policyholder, r is the risk-free short rate (with or without VA, no matter here) and $\mathrm{MP}(t)$ is the mathematical provision of the policyholder: it is the time-t capitalization of the contract valued with the delivered rate. To ensure the delivery of coupons to the policyholder, the insurer invests policyholder's deposit in a financial asset whose time-t value is denoted by $A(t)$ associated with a return rate $r_{ret}(t) = \ln\left(\frac{A(t)}{A(t-1)}\right)$. The successive return rates are assumed to be independent. The delivered rate is the maximum rate between the guaranteed rate and the net return rate of the asset: $e^{r_d(t)} - 1 = \max\Big(e^{r_g} - 1, r_{share}(e^{r_{ret}(t)} - 1) - (e^{r_{fees}} - 1)\Big)$. Differently written, we obtain that

$$e^{r_d(t)} = e^{r_g} + r_{share} \max\left(e^{r_{ret}(t)} - \frac{r_{share} + e^{r_{fees}} + e^{r_g} - 2}{r_{share}}, 0\right).$$

By definition, the mathematical provision satisfies $\mathrm{MP}(t) = \mathrm{MP}(0)\prod_{i:t_i\leq t} e^{r_d(t_i)}$ so that the time-t market value of the liability associated to the considered contract writes:

$$\begin{aligned}
L_t &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^t r_v \mathrm{d}v}\mathrm{MP}(t)\right] \\
&= \mathrm{MP}(0)\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^t r_v \mathrm{d}v}\prod_{i:t_i\leq t}\left\{e^{r_g} + r_{share}\max\left(e^{r_{ret}(t_i)} - \frac{r_{share}+e^{r_{fees}}+e^{r_g}-2}{PB}, 0\right)\right\}\right] \\
&= \mathrm{MP}(0)\prod_{i:t_i\leq t}\mathbb{E}^{\mathbb{Q}}\left[e^{r_g-\int_0^t r_v \mathrm{d}v} + r_{share}e^{-\int_0^t r_s \mathrm{d}s}\max\left(e^{r_{ret}(t_i)} - \frac{r_{share}+e^{r_{fees}}+e^{r_g}-2}{r_{share}}, 0\right)\right] \\
&= \mathrm{MP}(0)\prod_{i:t_i\leq t}\left\{\mathbb{E}^{\mathbb{Q}}\left[e^{r_g-\int_0^t r_v \mathrm{d}v}\right] + r_{share}\mathrm{Call}(t_{i-1}, t_i, \frac{r_{share}+e^{r_{fees}}+e^{r_g}-2}{r_{share}})\right\},
\end{aligned}$$

where $\mathrm{Call}(t, s, k)$ is a call option price at time t of maturity s, of strike k and whose underlying takes the value $e^{r_{ret}(s)}$ at maturity date. In that simple case, the insurer is led to replicate call options of maturities $(t_i)_i$ that are the dates at which the revaluations of the contract are done. In practice, it may be not possible to do such an analogy with all insurance contracts issued by the company and the determination of the considered financial securities is done on expert judgement. Moreover, though such an analogy is possible, a precise link between optionality of the assets (that may be similar to optionality of various financial instruments) and derivatives to be used for calibration is not easy to establish.

Alternative strategy consists in attributing weights to each swaption included in the calibration process: a null weight means that the associated swaption is not considered while other swaptions could be overweighted to be accurately replicated. We mention two methodologies to derive the mentioned weights: (i) first method computes the impact of a movement in swaptions volatilities in actuarial quantities -notably, on the NAV; it amounts to compute the *vega* of the NAV; (ii) second one consists in regressing the cash-flows delivered in the ALM model with the cash-flows delivered by financial instruments and, from regression coefficients, deduce weights to monitor the cash-flows replication error. Simpler approach that consists in attributing a null weight to undesired swaptions and a weight equals to 1 to others is also sometimes implemented.

1.4.3.2 Pricing conventions

In the two paragraphs that follow, we pre-empt some notions defined in more detail in Section 1.5. In particular, notations introduced further in Sections 1.5.1.4 and 1.5.1.5 are reused here. We will see in those sections that there exists a unique rate making the value of a swap contract fair at each time prior to the beginning of the exchanges definind the swap contract: it is the swap rate. Depending on the value of this swap rate at the maturity of the swaption, its owner will exercise its option or not. Namely, price of a swaption writes as a call option with underlying the swap rate. We justify it below: from the discounted payoff of the swaption in (1.32), we deduce the price of a payer swaption of maturity T_m, tenor $T_n - T_m$ and strike K, at

time $t \leq T_m$ as

$$\begin{aligned}
PS(t) &= \mathbb{E}^* \left[D(t,T_m) \Big(\sum_{k=m+1}^{n} P(T_m,T_i)(T_i - T_{i-1}) \left(F(T_m,T_{i-1},T_i) - K \right) \Big)_+ \Big| \mathcal{F}_t \right] \\
&= \mathbb{E}^* \left[\frac{B(t)}{B(T_m)} \Big(P(T_m,T_m) - P(T_m,T_n) - K \sum_{k=m+1}^{n} P(T_m,T_i)(T_i - T_{i-1}) \Big)_+ \Big| \mathcal{F}_t \right] \\
\iff \frac{PS(t)}{B(t)} &= \mathbb{E}^* \left[\frac{1}{B(T_m)} B^S(T_m) \big(S(T_m,T_m,T_n) - K \big)_+ \Big| \mathcal{F}_t \right]
\end{aligned}$$

where B represents a risk free asset (see Paragraph 1.5.1.1) and the expectation is defined with Risk-Neutral measure. It is convenient to work under the probability measure $\mathbb{P}^S$ often named *forward swap measure* following the terminology of [Jam97] and associated to the *numéraire*

$$B^S(t) = \sum_{k=m+1}^{n} P(t,T_i)(T_i - T_{i-1}) \tag{1.12}$$

called the *annuity* of the swap. According to the change of *numéraire* pricing equation (A.18), we finally get that

$$PS(t) = B^S(t) \times \mathbb{E}^S \left[\big(S(T_m,T_m,T_n) - K \big)_+ \Big| \mathcal{F}_t \right]. \tag{1.13}$$

On markets, swaptions are quoted in terms of *implied volatilities*. Swaptions prices can then be recovered by injecting the extracted implied volatility into a standard pricing formula. There are two pricing conventions, well known by practitioners: *Bachelier* and *Black* formulas we present in more details.

Bachelier convention

The Bachelier (or normal) pricing formula is based on the famous equation introduced in [Bac00]. The swaption price is obtained by assuming the swap rate follows a normal type dynamics under the forward swap measure. Namely, for $t \geq 0$, the swap rate is assumed to evolve following

$$\mathrm{d}S_t^{m,n} = \sigma_{m,n}^N \mathrm{d}W_t^S, \tag{1.14}$$

where $(W_t^S)_{t\geq 0}$ is a Brownian motion under $\mathbb{P}^S$ and where $\sigma_{m,n}^{\mathrm{N}} > 0$ is the volatility of the swap rate. Under this dynamics, the swaption price at time 0 writes:

$$PS^{\mathrm{Bach}}(0,K,\sigma_{m,n}^N) = B^S(0)\sigma_{m,n}^N \sqrt{T_m} \Big(\xi d \phi(\xi d) + \varphi(\xi d) \Big) \tag{1.15}$$

where $\xi = 1$ in case of a payer swaption ($\xi = -1$ for receiver), $d = (S_0^{m,n} - K)/(\sigma_{m,n}^N \sqrt{T_m})$, ϕ is the cumulative distribution function of the standard normal distribution and φ is its density function (in particular, $\phi' = \varphi$). The implied volatility parameter $\hat{\sigma}_{m,n}^{\mathrm{N}}$ (observed on markets) is defined as the (unique) parameter so that

$$PS^{\mathrm{Mkt}} = PS^{\mathrm{Bach}}(0,K,\hat{\sigma}_{m,n}^{\mathrm{N}}).$$

Black convention

The Black pricing formula follows from the well-known Black-Scholes formula introduced in

different context (essentially for equity modelling, see [Bla76] or [Mar78]). The swaption price is obtained by assuming the swap rate follows a log-normal type dynamics under the forward swap measure. Namely, for $t \leq 0$, the swap rate is assumed to evolve following

$$\mathrm{d}S_t^{m,n} = \sigma_{m,n}^{LN} S_t^{m,n} \mathrm{d}W_t^S, \tag{1.16}$$

where $(W_t^S)_{t\geq 0}$ is a Brownian motion under $\mathbb{P}^S$ and where $\sigma_{m,n}^{\mathrm{LN}} > 0$ is the volatility of the log-swap rate. Under this dynamics, the swaption price at time 0 writes:

$$PS^{\mathrm{Black}}(0, K, \sigma_{m,n}^{LN}) = B^S(0)\Big(S_0^{m,n}\xi\phi(\xi d_1) - K\xi\phi(\xi d_2)\Big) \tag{1.17}$$

where $\xi = 1$ in case of a payer swaption ($\xi = -1$ otherwise), $d_1 = \big(\ln(S_0^{m,n}/K) + T_m(\sigma_{m,n}^{LN})^2/2\big)/(\sigma_{m,n}^{LN}\sqrt{T_m})$ and $d_2 = d_1 - \sigma_{m,n}^{LN}\sqrt{T_m}$. The implied volatility parameter $\hat{\sigma}_{m,n}^{\mathrm{LN}}$ is defined as the (unique) parameter so that

$$PS^{\mathrm{Mkt}} = PS^{\mathrm{Black}}(0, K, \hat{\sigma}_{m,n}^{\mathrm{LN}}).$$

The choice between these two conventions is mainly determined by the level of interest rates. In the current low –possibly negative– rates environment, the use of the Black formula is not suited. To give some ideas about this point, first place ourselves in case when rates of maturities between T_m and T_n are negative; it amounts to say that the term-structure $T \mapsto P(0,T)$ is increasing over $[T_m, T_n]$. Following formula (1.15), the ATM spot price writes

$$PS^{\mathrm{Black,\ ATM}}(0) = \big(P(0,T_m) - P(0,T_n)\big)\Big(2\phi\Big(\frac{\sigma_{m,n}^{LN}\sqrt{T_m}}{2}\Big) - 1\Big)$$

and is thus a negative quantity which is not admissible since we are dealing with derivatives whose payoffs are positive. In case when rates are positive but close to zero, some instabilities already happen. In that second case, the term structure is almost flat and the swaption price is close to zero proportionally to the quantity $\big(P(0,T_m) - P(0,T_n)\big)$. The ratio $PS^{\mathrm{Black},ATM}(0)/\big(P(0,T_m) - P(0,T_n)\big)$ is then not necessarily bounded. The implied volatility for ATM swaptions in Black environment writes

$$\sigma_{m,n}^{LN} = \frac{2}{\sqrt{T_m}}\phi^{-1}\Big(\frac{1}{2}\Big(1 + \frac{PS^{\mathrm{Black,\ ATM}}(0)}{P(0,T_m) - P(0,T_n)}\Big)\Big).$$

Yet the inverse function ϕ^{-1} is not defined over the full real lines and its explosion near the boundaries of its domains explains that Black volatilities may be not quoted for short maturities when interest rates are low. On the contrary, in the Bachelier framework, one can show that

$$\sigma_{m,n}^{N} = \sqrt{\frac{2\pi}{T_m}}\frac{PS^{\mathrm{Black,\ ATM}}(0)}{B^S(0)}$$

which does not suffer from non-definiteness.

1.4.3.3 Difficulties induced by the regulation

We have discussed how Solvency II requires insurers and reinsurers to replicate market data through the market consistency criterion. Common pricing models express those market prices that have to be replicated as functions of implied volatility and *market* interest rates curve:

$$\text{Market prices} = f(\text{Term-structure}, \text{Implied volatilities}).$$

Nevertheless, legislation also imposes to use the risk-free regulatory term structure of which we have detailed its derivation in Section 1.4.1. If one wants to replicate market prices, it is thus necessary to prior adjust the implied volatilities: they are denominated *pseudo-volatilities*. Conversely, if one wants to replicate quoted implied volatilities, an adjustment on prices should be made: we talk of *pseudo-prices*.

1.4.3.3.1 Regulatory recommendations In paragraph 2 of Article 10 of [Par14] it is stated that: «*As the default valuation method insurance and reinsurance undertakings shall value assets and liabilities using the quoted market prices in active markets for the same assets or liabilities*». Article 22 in paragraph 3 of the same document claims that «(...) *where insurance and reinsurance undertakings use a model to produce projections of future financial market parameters, it shall comply with all the following requirements: (a) it generates asset prices that are consistent with asset prices observed in financial markets; (b) it assumes no arbitrage opportunity (...).*». On the basis of the comments made above, in particular about the impossibility to replicate both payer and receiver swaptions prices, it turns out that using prices as invariant quantity does not allow to account for the no arbitrage opportunity. Official texts leave the way open to both volatilities or prices as invariant when talking about «*inputs other than quoted prices that are observable for the asset or liability, including (...) implied volatility (...)*». Indeed, most of financial data provider directly quote implied volatility parameter rather than prices. Replicating implied volatility during the calibration process instead of the prices of derivatives seems to be a good way to conciliate the official guidelines.

1.4.3.3.2 Prices invariance and pseudo-volatilities When market prices are replicated, *pseudo-volatilities* parameters have to be defined. We illustrate our statements by examining the case of swaption pricing in a Bachelier framework. Swap rate is a function of ZC bond prices that are linked to yearly-compounded $(Y(0,T))_{T\leq T^*}$ rates, as we will see in relationship (1.23), through

$$P(0,T) = \frac{1}{(1+Y(0,T))^T}.$$

Annually-compounded rates depend on the short rate process $(r_t)_{t\geq 0}$ defined in (1.19).

Market swaptions prices are functions of the market curve $(r_t^{\text{Mkt}})_{t\geq 0}$. They are usually quoted in terms of relative strikes defined as $k_{m,n,K} = K - S_0^{m,n}$, where K is the absolute strike of the considered swaption and $S_0^{m,n}$ is the spot value of the swap rate. Consequently, relative strike is function of term-structure since $S_0^{m,n}$ is depending on spot prices of ZC bonds (see swap rate definition in (1.29)). Regulatory constraints ask us to replicate market swaptions prices using the regulatory curve $(r_t^{\text{Reg}})_{t\geq 0}$ obtained in Section 1.4.1. Denote by $\sigma_{m,n,K}$ the implied volatility parameter quoted on market for the $T_m \times T_n$ swaption of absolute strike K. The pseudo-volatility parameter $\hat{\sigma}_{m,n,K}$ is defined to be the parameter satisfying

$$\underbrace{PS^{\text{Mkt}}}_{\text{Market data}} = \underbrace{PS^{\text{Bach}}\Big(0, S_0^{m,n,\text{Reg}} + k_{m,n,K}^{\text{Reg}}, \hat{\sigma}_{m,n,K}, (r_t^{\text{Reg}})_{t\geq 0}\Big)}_{\text{Bachelier prices using regulatory yield curve}}.$$

In a classical use, one would input $(r_t^{\text{Mkt}})_{t\geq 0}$, $S_0^{m,n,\text{Mkt}}$, $k_{m,n,K}^{\text{Mkt}}$ and the observed implied volatility $\sigma_{m,n,K}$ in the right hand side above. For ATM derivatives, $k = 0$ and $PS^{\text{Bach}}\Big(0, S_0^{m,n,\text{Mkt}} + k_{m,n,K}^{\text{Mkt}}, \sigma_{m,n,K}, (r^{\text{Mkt}_t})_{t\geq 0}\Big) = \frac{B^{S,\text{Mkt}}(0)\sigma_{m,n,K}\sqrt{T_m}}{\sqrt{2\pi}}$ where we recall that $B^{S,\text{Mkt}}(0) = \sum_{j=m}^{n-1} \Delta T_j$ $P^{\text{Mkt}}(0, T_{j+1})$ the annuity induced by market curve. Denoting by $B^{S,\text{Reg}}(0)$ is the annuity

induced by regulatory curve, the price invariance yields

$$\hat{\sigma}_{m,n,K} = \frac{B^{S,\mathrm{Mkt}}(0)}{B^{S,\mathrm{Reg}}(0)}\sigma_{m,n,K}.$$

For AFM options same comments apply even though no analytical formulation of the pseudo-volatilities exist and a numerical procedure is required.

Limitations The well known Call-Put parity (see for instance [Sto69]) writes in present rates context as a link between payer and receiver swaption prices. In equation, it writes

$$\begin{aligned} PS(0) - RS(0) &= B^S(0)\Big(\mathbb{E}^S\big[(S^{m,n}_{T_m} - K)_+\big] - \mathbb{E}^S\big[(K - S^{m,n}_{T_m})_+\big]\Big) \\ &= B^S(0)\big(S^{m,n}_0 - K\big) \end{aligned} \tag{1.18}$$

after the price formula (1.13). Recall that both quantities $B^S(0)$ and $S^{m,n}_0$ depend on the initial term-structure curve. When using the regulatory curve to calibrate the interest rates model, the previous identity (1.18) becomes then $PS^{\mathrm{Reg}}(0) - RS^{\mathrm{Reg}}(0) = B^{S,\mathrm{Reg}}(0)(S^{m,n,\mathrm{Reg}}_0 - K)$. However quoted prices are such that $PS^{\mathrm{Mkt}}(0) - RS^{\mathrm{Mkt}}(0) = B^{S,\mathrm{Mkt}}(0)(S^{m,n,\mathrm{Mkt}}_0 - K)$. Consequently, the use of regulatory curve when considering prices as invariant induces some inconsistencies with market data as in particular it does not allow to reproduce both payer and receiver swaption prices.

1.4.3.3.3 Volatilities invariance and pseudo-prices Alternatively, one can replicate implied volatilities and define *pseudo-prices*. Namely, the market implied volatility is plugged into pricing formulas; depending on the choice of the interest rates term structure (market or regulatory), one recovers market prices or obtained different *pseudo-prices*. In formulas, using the Bachelier pricing convention, one has

$$\begin{aligned} PS^{\mathrm{Mkt}} &= PS^{\mathrm{Bach}}\Big(0, S^{m,n,\mathrm{Mkt}}_0 + k^{\mathrm{Mkt}}_{m,n,K}, \sigma_{m,n,K}, (r^{\mathrm{Mkt}}_t)_{t\geq 0}\Big) \\ PS^{\mathrm{Pseudo}} &= PS^{\mathrm{Bach}}\Big(0, S^{m,n,\mathrm{Reg}}_0 + k^{\mathrm{Reg}}_{m,n,K}, \sigma_{m,n,K}, (r^{\mathrm{Reg}}_t)_{t\geq 0}\Big). \end{aligned}$$

The pseudo-price is obtained by the Bachelier (1.15) or Black (1.17) pricing formula depending on the choice of the user. Observe that pseudo-prices allow to reproduce both payer and receiver swaption prices in regulatory environment through (for instance) the Call-Put parity since $PS^{\mathrm{Pseudo}} - RS^{\mathrm{Pseudo}} = B^{S,\mathrm{Reg}}(0)\big(S^{m,n,\mathrm{Reg}}_0 - K\big)$.

1.4.4 Simulations and validations

Once calibrated, models are simulated to produce the economic scenarios paths. The number of simulations should be sufficient to ensure the convergence of the Monte-Carlo computations made on these paths. Numerical approximations for simulations are standard: most common are discretization schemes (such as the famous Euler-Maruyama one, see for instance [KP92] on the topic) and weak simulations methods. The choice is made for each model depending on which method offers the best trade-off between the computational time (related to the simplicity of numerical methods) and accuracy of the numerical method.

The paths should be generated with high accuracy. In particular, empirical means of martingale quantities should be close to their initial values: this constitutes the *martingale tests* that should pass to validate the simulated paths. Moreover, the empirical correlation structure

between different financial drivers should match the input one. Those tests are required by the regulator and to pass them, some models can be simulated with a step time of one year without loss of precision but for some complex models (such as stochastic volatility dynamics) a smaller time increments is necessary.

That being said, and as already mentioned, generated paths should allow to appreciate all possible activations of options in different economic conditions. As for the calibration process, the simulation one should be made consistently with the risk profile of the company (see www.eiopa.europa.eu/content/guidelines-valuation-technical-provisions_en and notably guideline 55).

1.5 Interest rates modelling

We propose to give in this section an overview of the main models chosen by insurers in their ESGs to handle the interest rates risk along with their main interesting properties from an insurer point of view. We focus on most popular market models and briefly discuss short rates models for which more details are given in Appendix A with some technical regarding the pricing of derivatives of interest rates derivatives. The interested reader can refer to the authoritative and complementary books of Damiano Brigo and Fabio Mercurio [BM07] and Tomas Björk [Bjö09] for a fuller picture of the interest rates modelling. The following listing is non exhaustive and partly inspired from some parts of [BM07].

1.5.1 Some preliminaries

Beforehand, we introduce some definitions and notations on interest rates and their derivatives as the terminology employed in interest rates modelling is rich.

1.5.1.1 Bank account

We first introduce the money-market account (or bank account) that is the *risk-free* asset: we denote by $B(t)$ the time-t value of the money-market account. We assume that there exists a process $(r_t)_{t\geq 0}$ such that the process B evolves according to the following dynamics

$$\mathrm{d}B(t) = r_t B(t)\mathrm{d}t, \tag{1.19}$$

starting from $B(0) = 1$. Equivalently

$$B(t) = \exp\left(\int_0^t r_s \mathrm{d}s\right).$$

The value of $B(t)$ is assumed to be known at time t. Over the infinitesimal time interval $[t, t+\delta]$, a first order approximation yields that the value of the money-market account changes as $B(t+\delta) = (1 + r_t\delta)B(t)$. The process r, generally named *short rate*, is interpreted as the instantaneous interest rates delivered by the bank account. As being associated with the risk-free asset, r is alternatively referred to as the *risk-free rate*.

The risk-free asset accounts for the time value of money and establish an equivalence between amounts paid at different times. From the definition of the bank account value, we deduce that investing 1 unit of currency at time $t = 0$ delivers an amount $B(T)$ at time T. Similarly, investing $1/B(T)$ at initial date –in a deterministic case, for simplicity– provides 1 unit of currency at final date. More generally, for $t < T$, the *discount factor* $D(t, T)$ provides the time-t amount of

currency that is equivalent to a payment of 1 at time T. Arbitrage-free reasoning yields that

$$D(t,T) := \frac{B(t)}{B(T)}. \tag{1.20}$$

The money account is key in financial modelling as it allows to compare amounts of cash delivered at different dates. In this thesis if $(X_t)_{t\geq 0}$ represents some financial process, its discounted version is defined as $\tilde{X}_t = X_t/B(t)$. In particular case when $(X_t)_{t\geq 0}$ reduces to a time independent quantity, multiply it by the discount factor deflines its *discounted* equivalent : if X denotes some financial quantity delivered at time T, it can be convenient to consider its equivalent amount at time $t \leq T$ that is the discounted cash-flow $\tilde{X} := D(t,T)X$.

1.5.1.2 Zero-Coupon bonds and spot interest rates

The fundamental interest rate derivative to introduce –already mentioned several times in this thesis– is the *Zero-Coupon bond* (denoted ZC bond for short). Let us consider the following contract: at a future fixed date T, named the *maturity*, the contract delivers 1 unit of currency to the policyholder, with no intermediate payment. Time $t = 0$ corresponds to the date when the ZC bond is issued. The time-t value of this contract is denoted by $P(t,T)$ and is the amount of money to pay at time t to get 1 at time T. It is straight to get that $P(T,T) = 1$. As a price of a contract, it is natural to require P to be non-negative, $P(t,T) \geq 0$ for all (t,T), and to be known at any time t, $P(t,T)$ is $\mathcal{F}_t$-measurable.

Zero-Coupon bond is a contract linking payments of cash-flows at different dates. It is somehow linked with the discount factor discussed earlier. Recall $D(t,T)$ is an equivalent amount to a future amount of money, and thus depends on the path of the short rate process between t and T while P is known at time t. If the risk-free rate is deterministic, $P(t,T) = D(t,T)$ for any pair (t,T). However, in case when $(r_t)_{t\geq 0}$ is stochastic, $D(t,T)$ is also a stochastic quantity at any time while $P(t,T)$ is deterministic from the time t and thus the two quantities differ.

Continuously-compounded spot interest rate

Place ourselves at time t and assume that an investor holds an amount of money $P(t,T)$. Is there a constant rate at which he could *continuously* invest his money so that he recovers 1 unit of currency at time T? This rate exists and is the *continuously-compounded spot interest rate*, denoted by $R(t,T)$ and defined as

$$R(t,T) := -\frac{1}{(T-t)} \ln P(t,T) \Longleftrightarrow P(t,T) = e^{-(T-t)R(t,T)}. \tag{1.21}$$

In the particular deterministic case when the short rate process is constant, the continuously-compounded interest rate equals the short rate.

Simply-compounded spot interest rate

Now, rather than continuous accruing of invested money up to the maturity time T, we consider accruing proportional to the time-to-maturity $T - t$. Starting from an amount $P(t,T)$, we are looking for a constant rate prevailing over the time period $[t,T]$ at which one could invest its money to obtain 1 unit at maturity date T. This rate is the *simply-compounded spot interest rate*, denoted by $L(t,T)$ and defined as

$$L(t,T) := \frac{1 - P(t,T)}{(T-t)P(t,T)} \Longleftrightarrow P(t,T) = \frac{1}{1 + (T-t)L(t,T)}. \tag{1.22}$$

The notation $L(t,T)$ for simply-compounded rates comes from the LIBOR market on which simply-compounded rates are quoted.

Annually-compounded spot interest rate
In this paragraph, an annual reinvestment strategy is considered. Starting from an amount $P(t,T)$, we are looking for a constant rate over the time period $[t,T]$ at which one could invest its money once a year to obtain 1 unit at ending date T. This rate is the annually-compounded spot interest, denoted by $Y(t,T)$ and defined as

$$Y(t,T) := \frac{1}{P(t,T)^{1/(T-t)}} - 1 \Longleftrightarrow P(t,T) = \frac{1}{(1+Y(t,T))^{(T-t)}}. \tag{1.23}$$

Note that Y is often named *ZC rate*.

Observe that, in our definition of the ZC bond, it is implicitly assumed that the final payment of 1 unit of currency is guaranteed[12]. The spot interest rates introduced just above are all defined using the contract price $P(t,T)$. When defining them over infinitesimal time intervals, the three compounding convention are equivalent and amounts to consider the short rate as in (1.19).

Proposition 1.1. *For any time t,*

$$\lim_{T\to t^+} R(t,T) = \lim_{T\to t^+} L(t,T) = \lim_{T\to t^+} Y(t,T) = r_t.$$

Proof. Place ourselves over the infinitesimal time interval $[t,t+\delta]$ and assume that short rate process is such that $r_{t+\delta}$ is $\mathcal{F}_t$-measurable. An amount of money $B(t+\delta)P(t,t+\delta)$ invested at time t allows to obtain $B(t+\delta)$ units of currency at time $t+\delta$. Furthermore, investing $B(t)$ at time t delivers also $B(t+\delta)$ in $t+\delta$. Arbitrage-free assumption implies that $P(t,t+\delta) = B(t)/B(t+\delta)$. A first order approximation gives then $P(t,t+\delta) = 1 - r_t\delta$. Observe that all compounding convention (1.21)-(1.22)-(1.23) coincide at first order: $P(t,t+\delta) = 1 - \delta R(t,t+\delta) = 1 - \delta L(t,t+\delta) = 1 - \delta Y(t,t+\delta)$ (for the third convention, it is additionally used that $\ln(1+x) \simeq x$ for $|x| << 1$). Hence the result. □

Term-structure of interest rates
The term-structure of interest rates, or the yield curve (or zero-coupon curve), is the graph of the function

$$T \mapsto \begin{cases} L(t,T),\ t < T \leq t+1, \\ Y(t,T),\ T > t+1, \end{cases}$$

depicting the dependence of interest rates quoted at time t with respect to their maturities. This graph at time $t=0$ is the initial yield curve. Alternatively, depending on the context, one may be interested in the yield curve

$$T \mapsto R(t,T).$$

Finally, the zero-coupon bond curve is the graph, at any time t, of the function

$$T \mapsto P(t,T),\ T > t.$$

[12]The risk-free asset introduced in Section 1.5.1.1 is theoretical as in practice none asset is totally free of counterpart risk.

1.5.1.3 Forward rates

Forward rates are interest rates determined today for an investment over a future time consistently with known market information. They are simply-compounded interest rates as in (1.22) but seen at time prior to the period over which they prevail. It can be understood using the second fundamental interest rates derivative: the *Forward Rate Agreement* (FRA for short). The FRA is an exchange contract associated to a future period $[T, S]$ in which two counterparts A and B agree on a nominal value N and a fixed rate K. The working of the contract is the following: at expiry date $S > T$, B pays to A the agreed rate K for the period $[T, S]$, that is an amount of $N \times (S - T) \times K$; in return, A pays to B a sum of $N \times (T - S) \times L(T, S)$, depending on the floating rate $L(T, S)$ that is unknown prior to T. The spot rate $L(T, S)$ has been introduced in (1.22): it is known at time T, is of maturity S and depends on the market condition between T and S. At a time $t \leq T$, we are looking for the fixed rate K^* on which A and B should agree to make this contract fair; namely, the value of this contract should be the same for A and B. Since the positions of A and B are opposite, the rate we are looking for K^* is making the contract value zero at time t.

For A, the variation of wealth writes at time S, using the definition of the simply-compounded rate (1.22) as

$$(S - T)N\,(K - L(T, S)) = N\left((S - T)K - \frac{1}{P(T, S)} + 1\right)$$

and vice versa for B's wealth variation. We want to convey this quantity to its equivalent in time t. Since holding $P(T, S)$ at time T is equivalent to have 1 in S, holding 1 at time T is equivalent to have $1/P(T, S)$ in S. Moreover, 1 unit of currency at time T is equivalent to $P(t, T)$ at time t. Eventually, having $1/P(T, S)$ in S is equivalent to have $P(t, T)$ at time t. Furthermore, it is clear that the amount $(S-T)K+1$ at time S is equivalent to $P(t, S)(S-T)K+P(t, S)$ in time t. Thus, rewriting the profit and loss of A at time $t \leq T$ yields the value of the FRA contract:

$$\mathrm{FRA}(t, T, S, N, K) := N\xi\,[(S - T)KP(t, S) - P(t, T) + P(t, S)] \tag{1.24}$$

where $\xi = 1$. For B, the profit and loss writes similarly with $\xi = -1$. The FRA is said *receiver* from A's point of view and *payer* from B's. There exists a unique rate making this contract fair (i.e. that does not depend on ξ) and known at time t: it is named *simply-compounded forward rate* –or *forward rate* for short, denoted by $F(t; T, S)$ and defined for $t \leq T$ by

$$F(t, T, S) := \frac{1}{S - T}\left(\frac{P(t, T)}{P(t, S)} - 1\right). \tag{1.25}$$

Similarly to spot interest rates, we can define an *instantaneous forward rate*. These can be defined as a limiting case of the definition (1.25) when period $[T, S]$ becomes infinitesimal. Indeed, observe first that for any time $t \leq T$,

$$\lim_{S \to T^+} F(t, T, S) = \lim_{S \to T^+} \frac{1}{P(t, S)} \frac{P(t, T) - P(t, S)}{S - T} = -\frac{\partial \ln P(t, T)}{\partial T}, \tag{1.26}$$

which allows to define the *instantaneous forward rate* as

$$f(t, T) := \lim_{S \to T^+} F(t; T, S) = -\frac{\partial \ln P(t, T)}{\partial T}. \tag{1.27}$$

Consequently, the ZC bond price writes in term of instantaneous forward rate as

$$P(t,T) = \exp\left(-\int_t^T f(t,u)\mathrm{d}u\right).$$

Remark 2. *In practice, instantaneous forward rates are little used. Observe that definition of the instantaneous forward rates requires to assume the map $T \mapsto P(t,T)$ to be differentiable.*

1.5.1.4 Forward swap rates

We have just seen that FRA is a contract on a single future investment. It is naturally generalized in a multiple successive investments contract, called the *Interest Rates Swap* (IRS). This contract allows the two counterparts for exchanging a fixed leg against a floating one on a settled schedule starting in the future. Namely, at each date $T_i \in \boldsymbol{\mathcal{T}} := \{T_m, \dots, T_{n-1}\}$, the payer of the fixed leg disburses

$$N(T_{i+1} - T_i)K,$$

for the period $[T_i, T_{i+1}]$ while the floating leg pays out[13]

$$N(T_{i+1} - T_i)L(T_i, T_{i+1}),$$

where $L(T_i, T_{i+1})$ is the spot rate introduced in (1.22). Observe that the amount paid by the floating leg at time T_i for the period $[T_i, T_{i+1}]$ is known at time T_i, $m \leq i \leq n-1$. The discounted payoff of an IRS writes

$$\sum_{i=m}^{n-1} ND(T_i, T_{i+1})(T_{i+1} - T_i)\xi \times \Big(K - L(T_i, T_{i+1})\Big)$$

where $\xi = 1$ for a *receiver* IRS and $\xi = -1$ for a *payer* IRS[14]. Observe that owning a Receiver IRS is like having a portfolio of FRA according to the preceding discussion. The valuation formula (1.24) provides the value of the mentioned portfolio as

$$\begin{aligned}\mathrm{IRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K, N) &= \sum_{i=m}^{n-1} \xi \mathrm{FRA}(t, T_i, T_{i+1}, N, K) \\ &= N\xi \left(\sum_{i=m}^{n-1} K(T_{i+1} - T_i)P(t, T_{i+1}) + P(t, T_n) - P(t, T_m)\right).\end{aligned} \tag{1.28}$$

The value of the IRS for the Payer counterpart satisfies $\mathrm{PIRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K, N) = -\mathrm{RIRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K, N)$ where RIRS is the price of a receiver swap. Again, there exists a unique fixed rate K^* known at time $t < T_m$ making the Interest Rates Swap fair at time t that is $\mathrm{PIRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K^*, N) = \mathrm{RIRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K^*, N) = 0$. This rate is named *forward swap rate* and is defined by

$$S(t, T_m, T_n, \boldsymbol{\mathcal{T}}) := \frac{P(t, T_m) - P(t, T_n)}{\sum_{i=m}^{n-1}(T_{i+1} - T_i)P(t, T_{i+1})}. \tag{1.29}$$

This rate is often simply named *swap rate* even though one should keep in mind that it is a rate prevailing over a future period of time. If there is no ambiguity, we shall equally denote this

[13] Our framework is similar to that of [WZ06] but a little different from that of [BM07] in which authors consider that payments of the fixed and the floating legs do not coincide in time.

[14] Similary to FRA, the convention is that an IRS is said to be a *Payer IRS* if the fixed leg is paid and the floating leg is received; oppositely, the swap contract is a *Receiver IRS*.

swap rate by $S_t^{m,n}$ for brevity. A swap contract is said to be issued *at par* when its spot price coincides with the underlying nominal, i.e. $S_0^{m,n} = N$.

It should be stressed that the swap rate process is a function of forward rates implied during the period over which the swap rate prevails. Dividing both the numerator and the denominator in (1.29) by $P(t, T_m)$ and using (1.25), we get

$$\frac{P(t,T_i)}{P(t,T_m)} = \prod_{k=m}^{i-1} \frac{P(t,T_{k+1})}{P(t,T_k)} = \prod_{k=m}^{i-1} \frac{1}{1+(T_{k+1}-T_k)F(t,T_k,T_{k+1})},$$

which finally leads to

$$S(t,T_m,T_n,\boldsymbol{\mathcal{T}}) = \frac{1-\prod_{k=m}^{n-1}\left[1+(T_{k+1}-T_k)F(t,T_k,T_{k+1})\right]^{-1}}{\sum_{i=m}^{n-1}(T_{i+1}-T_i)\prod_{k=m}^{i}\left[1+(T_{k+1}-T_k)F(t,T_k,T_{k+1})\right]^{-1}}. \tag{1.30}$$

By defining weights $\alpha_i(t) = \frac{(T_{i+1}-T_i)P(t,T_{i+1})}{\sum_{k=m}^{n-1}(T_{k+1}-T_k)P(t,T_{k+1})}$ that are stochastic quantities, the swap rate expresses alternatively

$$S(t,T_m,T_n,\boldsymbol{\mathcal{T}}) = \sum_{k=m}^{n-1} \alpha_k(t) F(t,T_k,T_{k+1}). \tag{1.31}$$

In this representation, note that the weights α_k are also dependent on the forward rates.

1.5.1.5 Caplet/floorlet and swaption

We now turn to the description of derivatives on previously introduced rates. The *caplet* is a contract *agreed at a date t prior to T* that delivers in T the positive difference between the rate prevailing over the period $[T, S]$ and a fixed rate named *strike* denoted by K. In formulas, the discounted payoff of this product is

$$D(t,T)(S-T)\big(F(t,T,S)-K\big)_+, \ t \le T.$$

Caplet is a contract allowing to hedge against a rise of the interest rates. A *cap* is a sequence of caplets settled on a schedule $\{T_m, \dots, T_n\}$: the discounted price of a cap writes

$$\sum_{i=m}^{n-1} D(t,T_i)(T_{i+1}-T_i)\big(F(t,T_i,T_{i+1})-K\big)_+, \ t \le T_m.$$

A cap can be viewed as a payer IRS contract in which the successive payments are made only if they have a positive value.

Similarly, a *floorlet* is a product allowing to protect against a fall on the interest rates and works oppositely compare to caplet. Its discounted payoff writes

$$D(t,T)(S-T)\big(K-F(t,T,S)\big)_+, \ t \le T.$$

A *floor* is a sequence of floorlets whose discounted payoff is:

$$\sum_{i=m}^{n-1} D(t,T_i)(T_{i+1}-T_i)\big(K-F(t,T_i,T_{i+1})\big)_+, \ t \le T_m.$$

A floor can be viewed as a receiver IRS contract in which the successive payments are made

only if they have a positive value.

Third basic instrument we will be particularly interested in later on: the options on swap rate or more commonly, *swaptions*. These are *options* giving the right to their owner to enter into an Interest Rates Swap at a future date T_m, named the maturity of the swaption. The length of the period over which possible payments occur, $T_n - T_m$, is called the *tenor* of the swaption. The set of dates $\boldsymbol{\mathcal{T}} = \{T_m, \ldots, T_n\}$ is sometimes called the tenor structure. The fixed leg of the underlying swap is composed of payments of fixed rate K named the strike of the swaption. The time-t value of an IRS is given by (1.28) in which we use the definition (1.25):

$$\mathrm{IRS}(t, T_m, T_n, \boldsymbol{\mathcal{T}}, K, N) = N\xi \sum_{k=m}^{n-1} P(t, T_{i+1})(T_{i+1} - T_i)\left(F(t, T_i, T_{i+1}) - K\right), \, t \leq T_m.$$

At time T_m, the swaption's owner decides whether or not to enter the swap contract: the option is exercised if it has a positive value so that the discounted payoff of the swaption of maturity T_m is given by:

$$ND(t, T_m)\left(\sum_{k=m}^{n-1} P(T_m, T_{i+1})(T_{i+1} - T_i)\xi\left(F(T_m, T_i, T_{i+1}) - K\right)\right)_+, \, t \leq T_m. \tag{1.32}$$

For $\xi = 1$ (the case $\xi = -1$ is similar) we obtain thanks to (1.25) that the discounted payoff writes also

$$ND(t, T_m)\sum_{k=m}^{n-1} P(T_m, T_{i+1})(T_{i+1} - T_i) \times \Big(S(T_m, T_m, T_n, \boldsymbol{\mathcal{T}}) - K\Big)_+, \, t \leq T_m.$$

A swaption is said to be At-The-Money (ATM) if $K = S(0, T_m, T_n, \boldsymbol{\mathcal{T}})$, In-The-Money (ITM) if $K < S(0, T_m, T_n, \boldsymbol{\mathcal{T}})$ and Out-Of-Money (OTM) if $K > S(0, T_m, T_n, \boldsymbol{\mathcal{T}})$. In this thesis, we will refer to both OTM and ITM swaptions by Away-From-the-Money (AFM) options. On market, AFM swaptions are quoted in terms of *relative strikes* denoted by $k_{m,n,K}$ and defined by $k_{m,n,K} := K - S_0^{m,n}$.

Remark 1. *For simplicity, the nominal N is generally not denoted in pricing formulas. It amounts to consider $N = 1$ which does not impact generality of statements.*

1.5.2 Short rates models

Short rates models focus on the modelling of the evolution of the short rate process. They offer a good flexibility as one is free to choose the dynamics representing this time evolution. The most common short rates models met among insurers practices are the Hull & White, the G2++, the two-factor Hull & White and the CIR2++ models. They all provide an exact replication of the initial yield curve and the ability of generate non-positive rates. The Hull & White model is described by a one dimensional dynamics inducing perfect correlation on interest rates of different tenors; however, it is (thus) a quite simple Gaussian model making it highly tractable. The two-dimensional extension of it constitutes the two-factor Hull & White model and fixes the mentioned limitation; it is very similar to the G2++ model. Finally, the CIR2++ is a non Gaussian model a little harder to manipulate. More details can be found in Appendix A.

1.5.3 Market models

Market models usually referred to as LIBOR Market Model (LMM) have been introduced in [BGaM97]. They have became very popular for two main reasons: first because they focus on

the modelling of directly observable quantities on markets (hence their denomination) namely the forward rates (1.25) and swap rates (1.29); second because they model each mentioned rates through a log-normal distribution allowing the use of the practical and popular Black's formula.

Before going into the specification of the LMM, we specify the dynamics of the Zero-Coupon bond prices under general condition. To do so, two key general assumptions are made:
(*A*.1) there exists a d-dimensional Brownian motion $(\boldsymbol{W}_t)_{0\leq t\leq T^\star}$ that generates the market filtration: $\mathcal{F}_t := \sigma\Big(\boldsymbol{W}_u, u \leq t\Big)$, $0 \leq t \leq T^\star$;
(*A*.2) there exists a probability measure $\mathbb{P}^\star$ (Risk-Neutral measure), equivalent to the historical one $\mathbb{P}$, under which all discounted prices processes of Zero-Coupon bonds are $\mathcal{F}_t$-martingales; that is, for all $0 \leq u \leq T^\star$, $\big(\tilde{P}(t,u)\big)_{t\leq u}$ are $\mathcal{F}_t$-martingales. Those assumptions are enough to get that ZC bonds prices express as exponential martingales. Proposition 6.1.3 in [LL11] demonstrates this result.

Proposition 1.2 ([LL11])**.** *For each maturity $T \leq T^\star$ there exists a multidimensional adapted process $(\boldsymbol{\sigma}(t,T))_{t\leq T}$ so that, under Risk-Neutral measure, the dynamics of ZC bonds price writes*

$$\frac{\mathrm{d}P(t,T)}{P(t,T)} = r_t\mathrm{d}t + \boldsymbol{\sigma}(t,T)\cdot \mathrm{d}\boldsymbol{W}_t^\star,\ t \leq T, \tag{1.33}$$

where $(r_t)_{t\leq T^\star}$ is the short risk-free rate associated to the Risk-Neutral measure $\mathbb{P}^\star$ and $(\boldsymbol{W}_t^\star)_{0\leq t\leq T^\star}$ is a Brownian motion under this probability measure.

1.5.3.1 The LIBOR Market Model

The acronym « LMM » can refer to different models. Indeed, since the original work of [BGaM97], numerous works have extend their frameworks in several ways (see for instance [AA00], [Sch00], [WZ06], [Jam97], [Jam99], [HL08]). Since the financial crisis of 2008, new extension of the LMM have emerged in order to account for the remarkable increase of the spread between LIBOR rates and OIS ones. On this topic, we refer to [GPSS15]. Lately, an extension allowing jumps and occurrence of negative rates in a multiple curve setting has been proposed in [EGG20].

The Log-normal Forward-LIBOR Model (LFM)
We first focus on the modelling of forward rates as defined in (1.25): those are functions of zero-coupon bonds and are thus observable quantities.

Let be given a time horizon $T^\star$ and a tenor structure $\boldsymbol{\mathcal{T}} = \{T_1, \dots, T_N\}$, $T_1 \leq \dots T_N \leq T^\star$ and let us denote by $\Delta T_k = T_{k+1} - T_k$. The ZC bonds dynamics (1.33) is valid for all bonds of maturity $T_k \in \boldsymbol{\mathcal{T}}$ and we denote by $\boldsymbol{\sigma}_k(t) = \boldsymbol{\sigma}(t, T_k)$. Let us denote by $F_k(t) = F(t, T_k, T_{k+1})$ the forward rate prevailing over $[T_k, T_{k+1}]$. As originally introduced, the LIBOR market model framework consists in assuming, for each $k \in [\![1, N-1]\!]$, the existence of deterministic, bounded and piecewise continuous multidimensional function of time $t \in [0, T_k] \mapsto \boldsymbol{\gamma_k}(t)$ such that

$$\mathrm{d}F_k(t) = (\dots)\mathrm{d}t + F_k(t)\boldsymbol{\gamma_k}(t)\cdot \mathrm{d}\boldsymbol{W}_t,\ t \leq T_k, \tag{1.34}$$

where $(\boldsymbol{W}_t)_{t\leq T_k}$ is a multidimensional Brownian motion. Note that the $\boldsymbol{\gamma_k}$ functions convey some information on volatility structure of forward rates but also on the correlation structure between them since, for $j \neq k$,

$$\mathrm{d}\left\langle F_j(.), F_k(.)\right\rangle_t = F_j(t)F_k(t)\boldsymbol{\gamma}_j(t)\cdot\boldsymbol{\gamma}_k(t)\mathrm{d}t.$$

According to their definition (1.25), forward rates are functions of ZC bonds and Itô's lemma can be applied with (1.33). To ensure consistency between the induced dynamics and the proposed one (1.34), no arbitrage requirement links the volatility structure of ZC bonds to that of forward rates through

$$\frac{1+\Delta T_k F_k(t)}{\Delta T_k F_k(t)}\big(\boldsymbol{\sigma}_{k+1}(t)-\boldsymbol{\sigma}_k(t)\big)=\boldsymbol{\gamma}_k(t),$$

which conversely allows to define the volatility structure of ZC bonds as

$$\boldsymbol{\sigma}_k(t)=-\sum_{j=1}^{k-1}\frac{\Delta T_j F_j(t)}{1+\Delta T_j F_j(t)}\boldsymbol{\gamma}_j(t),$$

since $\boldsymbol{\sigma}_1$ can be set to 0 as discussed in [BGaM97].

From the change of *numéraire* technique discussed in Section A.2.2, we note that dividing a price process by a *numéraire* process allows to make it a martingale under the appropriate measure. Here, it is common to study the k-th forward rate under the probability measure $\mathbb{P}^{T_{k+1}}$ associated to the numéraire $\big(P(t,T_{k+1})\big)_{t\le T_{k+1}}$ and often named T_{k+1}-*forward neutral measure*, introduced in a more general setting in (A.19). By doing so, the k-th forward rate writes

$$F_k(t)=F_k(0)\exp\left(\int_0^t \boldsymbol{\gamma}_k(s)\cdot \mathrm{d}\boldsymbol{W}_s^k-\frac{1}{2}\int_0^t \|\boldsymbol{\gamma}_k(s)\|^2\mathrm{d}s\right) \tag{1.35}$$

where $(\boldsymbol{W}_t^{k+1})_{t\le T_k}$ is a $\mathbb{P}^{T_{k+1}}$-Brownian motion. We recall that $F_k(0)=\big(P(0,T_k)-P(0,T_{k+1})\big)/\big((T_{k+1}-T_k)P(0,T_{k+1})\big)$ and is thus an observable quantity that is inherently reproduced by the LMM. Note also that the representation (1.35) is well defined since $t\mapsto\|\boldsymbol{\gamma}_k(t)\|$ has been assumed to be bounded. This formulation is particularly suited for caplet/floorlet pricing as it allows the use of Black's formula.

Proposition 1.3. *Prices of caplet and floorlet at time $t\le T_k$ given by LFM are given by*

$$\mathrm{Cpl}(t,T_k,T_{k+1},K)=P(t,T_k)\big(F_k(t)\Phi(d_1)-K\Phi(d_2)\big)$$

and

$$\mathrm{Flr}(t,T_k,T_{k+1},K)=P(t,T_k)\big(K\Phi(-d_2)-F_k(t)\Phi(-d_1)\big),$$

where we denoted by Φ the cumulative function of the standard distribution, $d_1=\big(\ln(F_k(t)/K)+v_k(t)^2/2\big)/v_k(t)$, $d_2=d_1-v_k(t)$ and $v_k(t)=\sqrt{T_k-t}\tilde{v}_k(t)$ and $\tilde{v}_k^2(t)=\int_t^{T_k}\|\boldsymbol{\gamma}_k(s)\|^2\mathrm{d}s/(T_k-t)$.

Proof. From the pricing formula (A.20), we get:

$$\begin{aligned}\mathrm{Cpl}(t,T_k,T_{k+1},K)&=\mathbb{E}^*\left[\frac{B(t)}{B(T_k)}\big(F(T_k,T_k,T_{k+1})-K\big)_+\big|\mathcal{F}_t\right]\\&=P(t,T_k)\mathbb{E}^{T_k}\left[(F_k(T_k)-K)_+\big|\mathcal{F}_t\right]\\&=P(t,T_k)\big(F_k(t)\Phi(d_1)-K\Phi(d_2)\big).\end{aligned}$$

To obtain the last identity, we use (1.35) to get that $F_k(T_k)\overset{d}{=}F_k(t)\exp\left(\sqrt{\int_t^{T_k}\|\boldsymbol{\gamma}_k(s)\|^2\mathrm{d}s}\times G-\frac{1}{2}\int_t^{T_k}\|\boldsymbol{\gamma}_k(s)\|^2\mathrm{d}s\right)$ where G is a centered/reduced Gaussian random variable. □

Remark 3. *We work here in a discrete tenor setup which is more convenient for practical uses. The problem of the existence of a solution to (1.34) in a continuous tenor setup as in [BGaM97]*

is more tricky.

The Risk-Neutral and spot-LIBOR measures in LFM

We have already defined the Risk-Neutral probability measure as being the measure associated with the bank account asset (1.19). However, the bank account is not a *natural* asset in the LIBOR market model framework with discrete preassigned tenor structure since bank account represents the value of a continuously reinvested notional at risk-free rate. In the LMM framework we work in, no such rate exist and building a strategy coinciding with such an investment is not possible in a discrete tenor structure. Note that in [BGaM97] the definition of a bank account process is possible as the authors worked with continuous tenor structure.

From the change-of-*numéraire* toolkit in Section 2.3 of [BM07], one obtains that the dynamics of the k-th forward rate under the Risk-Neutral probability measure writes

$$\mathrm{d}F_k(t) = -F_k(t)\boldsymbol{\sigma}_{\ln P(.,T_k)}(t)\cdot\boldsymbol{\gamma}_k(t)\mathrm{d}t + F_k(t)\boldsymbol{\gamma}_k(t)\cdot\mathrm{d}\boldsymbol{W}_t,\ t\le T_k, \tag{1.36}$$

where $\boldsymbol{\sigma}_{\ln P(.,T_k)}$ is the diffusion coefficient of the log-price process of T_k-ZC bonds. We introduce the index function of the next expiring forward rate in the tenor structure as $m(t) = \inf\{j \ge 1 : t \le T_j\}$ and we observe that

$$P(t,T_k) = P(t,T_{m(t)})\prod_{j=m(t)}^{k-1}\frac{1}{1+\Delta T_j F_j(t)}.$$

$\boldsymbol{\sigma}_{\ln P(.,T_k)}$ is thus the diffusion coefficient of $\ln P(t,T_{m(t)}) - \sum_{j=m(t)}^{k-1}\ln\big(1+\Delta T_j F_j(t)\big)$. Moreover, using the definition of instantaneous forward rate (1.27), we have $\ln P(t,T_k) = -\int_t^{T_k} f(t,u)\mathrm{d}u$ and thus the diffusion coefficient of log-price of ZC bond is the integrated coefficient diffusion of instantaneous forward rate.

In the following, $\boldsymbol{\sigma}_f$ is such that $\mathrm{d}f(t,T) = (\cdots)\mathrm{d}t + \boldsymbol{\sigma}_f(t,T)\cdot\mathrm{d}W_t$ and we use the notation $\boldsymbol{u}(t)\cdot\Big(\int_s^S \boldsymbol{v}(u)\mathrm{d}u\Big) = \sum_{i=1}^n \boldsymbol{u}^{(i)}(t)\int_s^S \boldsymbol{v}^{(i)}(u)\mathrm{d}u$, where n is the length of the vector functions $\boldsymbol{u}$ and $\boldsymbol{v}$ and $\boldsymbol{u}^{(k)}(t)$ is the k-th component of vector function $\boldsymbol{u}$.

Proposition 1.4 ([BM07])**.** *The drift coefficient in (1.36) writes, for all $t \le T_k$:*

$$-\boldsymbol{\sigma}_{\ln P(.,T_k)}(t)\cdot\boldsymbol{\gamma}_k(t) = \sum_{j=m(t)}^{k-1}\frac{\Delta T_j F_j(t)}{1+\Delta T_j F_j(t)}\boldsymbol{\gamma}_j(t)\cdot\boldsymbol{\gamma}_k(t) + \boldsymbol{\gamma}_k(t)\cdot\Big(\int_t^{T_{m(t)}}\boldsymbol{\sigma}_f(t,u)\mathrm{d}u\Big). \tag{1.37}$$

Proof. See Proposition 6.3.2 in [BM07]. □

This dynamics is hard to handle in practice. To circumvent this difficulty, the LMM is generally studied under an alternative probability measure that is often assimilated to the Risk-Neutral measure but that actually does not coincide *a priori*. Consider the quantity

$$B_d(t) = \frac{P\big(t,T_{m(t)}\big)}{\prod_{k=1}^{m(t)-1}P(T_k,T_{k+1})} = P\big(t,T_{m(t)}\big)\prod_{k=1}^{m(t)-1}(1+\Delta_k F_k(T_k)).$$

It corresponds to an investment of one unit of currency successively reinvested in zero-coupon bonds of successive maturities of the tenor structure (see p.219 in [BM07] for the detailed reasoning). The probability measure $\mathbb{P}^d$ associated with the numéraire B_d is usually called the *spot LIBOR measure*, and the $\mathbb{P}^d$-dynamics of forward rates are simpler.

Proposition 1.5 ([BM07]). *Under spot LIBOR measure, k-th forward rate dynamics writes*

$$\mathrm{d}F_k(t) = F_k(t) \sum_{j=m(t)}^{k} \frac{\Delta T_j F_j(t)}{1+\Delta T_j F_j(t)} \boldsymbol{\gamma}_j(t) \cdot \boldsymbol{\gamma}_k(t)\mathrm{d}t + F_k(t)\boldsymbol{\gamma}_k(t) \cdot \mathrm{d}\boldsymbol{W}_t^d, \ t \leq T_k, \tag{1.38}$$

with $\left(\boldsymbol{W}_t^d\right)_{t \leq T^\star}$ a $\mathbb{P}^d$-Brownian motion.

Proof. Proposition 6.3.3 in [BM07]. □

Observe that dynamics coming from (1.37) and (1.38) match if $\boldsymbol{\sigma}_f(t,u) = 0$ for all $u \in [t, T_{m(t)}]$. Using the link with HJM framework, i.e. fact that $\boldsymbol{\sigma}$ expresses as the opposite of the integral of $\boldsymbol{\sigma}_f$ (see Chapter 5 in [BM07]), this requirement means that at any time t the volatility of the next expiring zero-coupon bond is zero: $\boldsymbol{\sigma}(t, T_{m(t)}) = 0$. This is the condition generally required to ensure spot-LIBOR measure coincide with Risk-Neutral one (see for instance [BGaM97] or [Sch02]).

The Log-normal Forward-Swap Model (LSM)

Log-normal forward swap model focus on the modelling of the forward swap rate process. We consider a tenor structure $\boldsymbol{\mathcal{T}'} = \{T_m, \ldots, T_n\}$ and denote by $S_t^{m,n} = S(t, T_m, T_n, \boldsymbol{\mathcal{T}'})$ the swap rate defined in (1.29). We place ourselves under the forward-swap measure $\mathbb{P}^S$ introduced in (A.22) and associated to the numéraire B^S defined in (A.21). Similarly to the LFM, the modelling framework of LSM consists in assuming a log-normal dynamics for the swap rate process

$$\mathrm{d}S_t^{m,n} = S_t^{m,n}\boldsymbol{\sigma}^{m,n}(t) \cdot \mathrm{d}\boldsymbol{W}_t^{m,n}, \ t \leq T_m, \tag{1.39}$$

with $(\boldsymbol{W}_t^{m,n})_{t \leq T_m}$ a Brownian motion under $\mathbb{P}^S$. This time evolution is convenient for pricing of swaptions (payer or receiver) as the Black's formula can be employed again. We provide below simply the formula for valuing payer swaptions.

Proposition 1.6. *Price at time $t \leq T_m$ of a payer swaption writes*

$$PS(t, T_m, T_n, \boldsymbol{\mathcal{T}'}, K) = B^S(t)\big(S_t^{m,n}\Phi(d_1) - K\Phi(d_2)\big)$$

where we denoted by Φ the cumulative function of the standard distribution, $d_1 = \big(\ln(S_t^{m,n}/K) + v_{m,n}(t)^2/2\big)/v_{m,n}(t)$, $d_2 = d_1 - v_{m,n}(t)$ and $v_{m,n}(t) = \sqrt{T_m - t}\tilde{v}_{m,n}(t)$ and $\tilde{v}_{m,n}^2(t) = \int_t^{T_m} \|\boldsymbol{\sigma}^{m,n}(s)\|^2 \mathrm{d}s/(T_m - t)$.

Proof. Similarly to Proposition 1.3, we have:

$$\begin{aligned} PS(t, T_m, T_n, \boldsymbol{\mathcal{T}'}, K) &= \mathbb{E}^* \left[\frac{B(t)}{B(T)} \left(S_{T_m}^{m,n} - K\right)_+ \middle| \mathcal{F}_t \right] \\ &= B^S(t)\mathbb{E}^S \left[\left(S_{T_m}^{m,n} - K\right)_+ \middle| \mathcal{F}_t \right] \end{aligned}$$

and the result follows from similar computations to that of Black formula for caplet pricing in Proposition 1.3. □

Forward swap rate express as a function of forward rates as depicted in formulas (1.30) and (1.31). One can naturally wonder if swaption prices obtained using Proposition 1.6 or those induced by the dynamics of forward rates (1.36) are the same. In other words, are the LFM and LSM frameworks consistent? Not surprisingly, the answer is no. Indeed, deriving the dynamics

of forward rates under $\mathbb{P}^S$ leads to a complicated set of equations in which forward rates are not log-normally distributed *a priori*. Conversely, dynamics of swap rate can be derived under a particular forward neutral measure $\mathbb{P}^{T_k}$ and again there is no reason for the swap rate to be log-normal. However, this incompatibility is rather theoretical. In practice, the error induced by considering that both swap rates and forward rates are log-normal is low. This point is studied in [BDB01] and we will come on it back later on. The interested reader can refer to Section 6.8 of [BM07] for more details on the inconsistencies between the two models.

1.5.4 Replication of the smile

Solvency II legislation asks insurers models to be market consistent, that is being calibrate to market prices of derivatives. As explained in Section 1.4.3, those can comprise AFM options and in particular AFM swaptions. It is well known that original models such as the Black-Scholes models for equity modelling are not parametrized enough to be able to replicate multiple AFM options prices. The introduction of «smiled models» has been crucial to do so. The probably most popular model allowing the replication of a full *smile*[15] is the Heston model introduced in [Hes97] and which lies in the family of stochastic volatility model. It extends the Black-Scholes model [BS73] by adding a stochastic volatility factor whose time evolution is depicted by a Cox-Ingersoll-Ross process. Local stochastic volatility (LSV) models are another common class of smiled models: the dependency of the implied volatility with respect to the strike is represented through a leverage function of the time and the current value of the underlying. Those models were introduced for equity modelling, but similar developments have naturally been introduced in market models. The most common market models used by insurers are described below.

1.5.4.1 Local stochastic volatility models

1.5.4.1.1 Displaced Diffusion LMM This model, usually referred to as the DD-LMM by practitioners, lies in the family of market models as described in Section 1.5.3.1.

Forward rates To take into account non-positive forward rates observed on market data, a displacement coefficient $\delta \geq 0$, often named *shift*, is introduced. We retake the tenor structure of Section 1.5.3.1. The DD-LMM focus on the modelling of time evolution of *shifted forward rates* $\tilde{F}_k := F_k + \delta,\ k \in \{1, \dots, N\}$. Let $N_f \in \mathbb{N}$. In the DD-LMM, shifted forward rates are assumed to be represented by the dynamics (1.34): the existence of a N_f-dimensional function $t \mapsto \boldsymbol{\gamma}_k(t)$ is assumed such that the shifted forward rates evolve as

$$\mathrm{d}\tilde{F}_k(t) = (\cdots)\mathrm{d}t + \tilde{F}_k(t)\boldsymbol{\gamma}_k(t) \cdot \mathrm{d}\boldsymbol{W}_t^{k+1},\ t \leq T_k, \tag{1.40}$$

with $\tilde{F}_k(0) = F_k(0)+\delta$. The volatility structure is assumed to be represented by a humped shape with piecewise constant time dependency and parametrized as $\boldsymbol{\gamma}_k(t) \equiv \boldsymbol{\gamma}_k(T_i) = f(T_i) \times g(T_k - T_i)\boldsymbol{\beta}_{k-i+1}$ for all $t \in [T_i, T_{i+1}[$, $i \leq k-1$, where $g(t) = (a+bt)e^{-ct} + d$ with $(a,b,c,d) \in (\mathbb{R}_+)^4$ and $f(t) = \theta + (1-\theta)e^{-\kappa t}$ with $(\theta, \kappa) \in (\mathbb{R}_+)^2$ and where $\boldsymbol{\beta}_i$ is a unitary vector of size N_f for any $i \leq N$. Those vectors stand for the correlation structure between forward rates since, for

[15] We refer here to the *smile* as being the graph of the function linking the strike of an option to its price/implied volatility.

$t \in [T_i, T_{i+1}[$, $i \leq \min(j-1, k-1)$,

$$\langle F_k(.), F_j(.)\rangle_t = \sum_{p=0}^{i-1} \Delta T_p f(T_p)^2 g(T_k - T_p) g(T_j - T_p) \boldsymbol{\beta}_{k-p+1} \cdot \boldsymbol{\beta}_{j-p+1} + \big(t - T_i\big) f(T_i)^2 g(T_k - T_i) g(T_j - T_i) \boldsymbol{\beta}_{k-i+1} \cdot \boldsymbol{\beta}_{j-i+1}.$$

In practice, those vectors can be determined using a Principal Component Analysis on historical time series of log-shifted forward rates. In this work, we will consider them as meta-parameters determined once and for all and we will not detail their obtention. To price caplets/floorlets as in Proposition 1.3, one has to compute the integrated variance process. For spot prices, one will be interested in the following quantity:

$$\int_0^{T_k} \|\boldsymbol{\gamma}_k(s)\|^2 \mathrm{d}s = \sum_{i=1}^{k-1} \Delta T_i g(T_k - T_i)^2 f(T_i)^2.$$

Swap rates In the DD-LMM, *shifted swap rates* dynamics are derived using (1.31):

$$\begin{aligned} S_t^{m,n} + \delta &= \frac{1 - \prod_{j=m}^{n-1} \big(1 + \Delta T_j (\tilde{F}_j(t) - \delta)\big)^{-1}}{\sum_{j=m}^{n-1} \Delta T_j \prod_{l=m}^{j} \big(1 + \Delta T_l (\tilde{F}_l(t) - \delta)\big)^{-1}} + \delta \\ &=: \psi\big(\tilde{F}_m(t), \ldots, \tilde{F}_{n-1}(t)\big) + \delta \end{aligned}$$

where it is important to observe that the function ψ linking forward rates to the swap rates does depend on the shift coefficient explaining why this parameter is structuring in this the DD-LMM.

Under the forward swap measure $\mathbb{P}^S$, the swap rate is a martingale and the Itô's formula yields

$$\begin{aligned} \mathrm{d}\tilde{S}_t^{m,n} &= \sum_{j=m}^{n-1} \partial_j \psi(t) \mathrm{d}\tilde{F}_j(t) \\ &= \sum_{j=m}^{n-1} \big(F_j(t) + \delta\big) \partial_j \psi(t) \boldsymbol{\gamma}_j(t) \cdot \mathrm{d}\boldsymbol{W}_t^S \end{aligned} \tag{1.41}$$

where $\big(\boldsymbol{W}_t^S\big)_{t \leq T_m}$ is a Brownian motion under $\mathbb{P}^S$ and we set $\partial_j \psi(t) = \frac{\partial \psi}{\partial x_j}\big(\tilde{F}_m(t), \ldots, \tilde{F}_{n-1}(t)\big)$, $j \in \{m, \ldots, n-1\}$. Spot value is prescribed by observed data $\tilde{S}_0^{m,n} = \frac{P(0,T_m) - P(0,T_n)}{B^S(0)} + \delta$. The partial derivative $\partial_j \psi$ can be derived in closed form formula as (see for instance in [WZ06]):

$$\partial_j \psi(t) = \alpha_j(t) + \frac{\Delta T_j}{1 + \Delta T_j \big(\tilde{F}_j(t) - \delta\big)} \sum_{l=m}^{j-1} \alpha_l(t) \big(\tilde{F}_l(t) - \tilde{S}_t^{m,n}\big),$$

where the weights $(\alpha_k)_{k \in [\![m, \ldots, n-1]\!]}$ have been introduced in (1.31). The obtained dynamics (1.41) is too complex to be fully handled as it stands: in particular, computations of swaption prices require uses of Monte-Carlo methods and closed form expressions can not be obtained *a priori*. In practice, this dynamics is approximated using the *freezing* technique that will be studied in more details in the dedicated Chapter 2. As of now, illustrate this technique by what

we will call the *Hull & White approximation*. It consists in assuming that the ratios

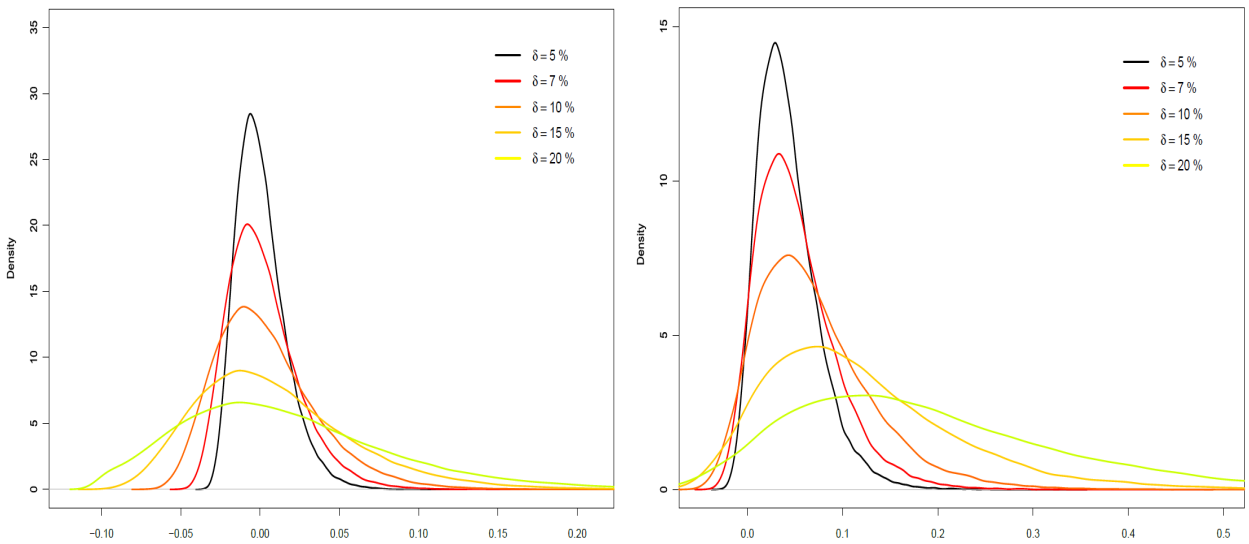
$$\frac{\partial_k \psi(t)\big(F_k(t)+\delta\big)}{S_t^{m,n}+\delta}, \ k = m, \dots, n-1,$$

are of low variability through time and can thus be approximated by their initial values. The approximated dynamics writes

$$\mathrm{d}S_t^{m,n} = \big(S_t^{m,n}+\delta\big) \sum_{k=m}^{n-1} \tilde{\gamma}_k(t) \cdot \boldsymbol{W}_t^S, \ t \leq T_m, \tag{1.42}$$

where $\tilde{\gamma}_k(t) := \frac{\partial_k \psi(0)(F_k(0)+\delta)}{S_0^{m,n}+\delta} \gamma_k(t)$. Thus, under the approximated dynamics (1.42) the shifted swap rate is log-normal and, similarly to Proposition 1.6, a Black formula in which the underlying is the shifted swap rate can be employed to value swaptions.

To illustrate the impact of the shift coefficient, we provide some illustrations in Figure 1.8 on the dependency of terminal distributions of some forward rates regarding the shift. Namely, the 5-year and the 30-year forward rates terminal densities are displayed for different values of the shift, all other parameters being fixed. As expected, the variance of the forward rates increases with the shift. Note that the illustrations are empirical densities (50000 simulations).



(a) Terminal density of 10-year forward rates.

(b) Terminal density of 30-year forward rates.

Figure 1.8: Impact of the shift coefficient on terminal densities of forward rates.

Smile in the DD-LMM When $\delta > 0$, the DD-LMM allows to introduce some dependency in the implied volatility with respect to the strike. First, recall that the implied volatility is defined as being the volatility parameter to input in the *standard* Black formula (see for instance Proposition 1.6) in which $\delta \equiv 0$ is set. When computing swaptions prices under (1.42) with $\delta > 0$ and then inverting it with the standard Black formula, the implied volatility parameter obtained turns out to be strike dependent. We can appreciate this dependency using estimates formulas. Approximations of implied volatilities for vanilla pricing when underlying is modelled through a dynamics whose drift is null have been derived in [HW99a]. Following their methodology based on singular perturbation of vanilla prices, they obtained that the equivalent Black implied volatility for a swaption of maturity T_m and strike K writes under dynamics

(1.42) as:

$$\sigma_{\text{imp}}^{\text{Black}} = \Sigma(T_m)\frac{S_0^{m,n}+K+2\delta}{S_0^{m,n}+K}\Bigg\{1+\frac{(S_0^{m,n}-K)^2}{12S_{av}^2}\Big(1-\frac{1}{(1+\delta/S_{av})^2}\Big) + \frac{\Sigma(T_m)^2T_m\delta}{24S_{av}}\Big(2+\frac{\delta}{S_{av}}\Big)+\cdots\Bigg\}, \tag{1.43}$$

where we set $\Sigma(T_m)^2 = \frac{1}{T_m}\int_0^{T_m}\|\boldsymbol{\gamma}_k(s)\|^2\mathrm{d}s$ and $S_{av} = \frac{1}{2}(S_0^{m,n}+K)$.
Observe that in case when $\delta = 0$, the implied volatility parameter does not depend on K and we recover the well known fact that $\sigma_{\text{imp}}^{\text{Black}} = \Sigma(T_m)$.

To end this paragraph, we provide some illustrations on the impact of the shift coefficient on the DD-LMM and in particular on the induced smile. In Figure 1.9, we provide an example of the replication of a smile provided by the DD-LMM: the latter is calibrated to market data (the shift is fixed to $\delta = 10\%$ during the calibration procedure here) and model volatilities induced by the model are compared to the market ones. It turns out that it does not allow a good fit of market data as the model smile roughly linear with respect to the moneyness. In Figure 1.10, the set of parameters of the DD-LMM is set fixed but the shift coefficient and we analyse the dependency of the smile on swaptions volatilities of different maturities with respect to it; the tenor is fixed to 10 years for all experiments. Quite intuitively, the greater the shift, the riskier the swaptions are. We still observe that the smile induced by the DD-LMM is almost linear which only permits a poor replication of market data. Finally, we illustrate how the shift impact the whole structure of the DD-LMM. Namely, the parameters of the model are calibrated by fixing the shift coefficient to different values. The results are condensed in Table 1.3. Observe that shifts between 10% and 15% account for the best fit of market data (the objective function is computed as the normalized sum of the squared relative distances between market and model volatilities). On our experiment, the parameters a and d turn out to be the most volatile parameters with respect to the value of the shift, followed b and θ. Recall that in present setting, parameters d –resp. a and θ– are proportional to the global level of short-term –resp. long-term– instantaneous volatilities of log-ratios of shifted forward rates (i.e. $\log\big(\tilde{F}_k(t+\Delta t)\big) - \log\big(\tilde{F}_k(t)\big)$); their impact is particularly material on densities of forward rate with large maturities. In Figure 1.11, we provide spot and terminal densities of the forward rates induced by the different parametrization of the Table 1.3. The shift has greater impact on rates with large maturities, both on initial and terminal densities. Initial density of 10-year rates (Figure 1.11a) is quite stable with respect to the shift values; it is not the case of its terminal density (Figure 1.11c). 30-year rates densities have significant dependencies with respect to the shift (Figures 1.11b and 1.11d). In Figure 1.11c, 1.11b and 1.11d, the thickest tails of the distributions are obtained for $\delta = 20\%$: overall, the higher the shift, the thicker are the tails. This is quite intuitive though in Figure 1.11b the dependency with respect to δ is hard to interpret.

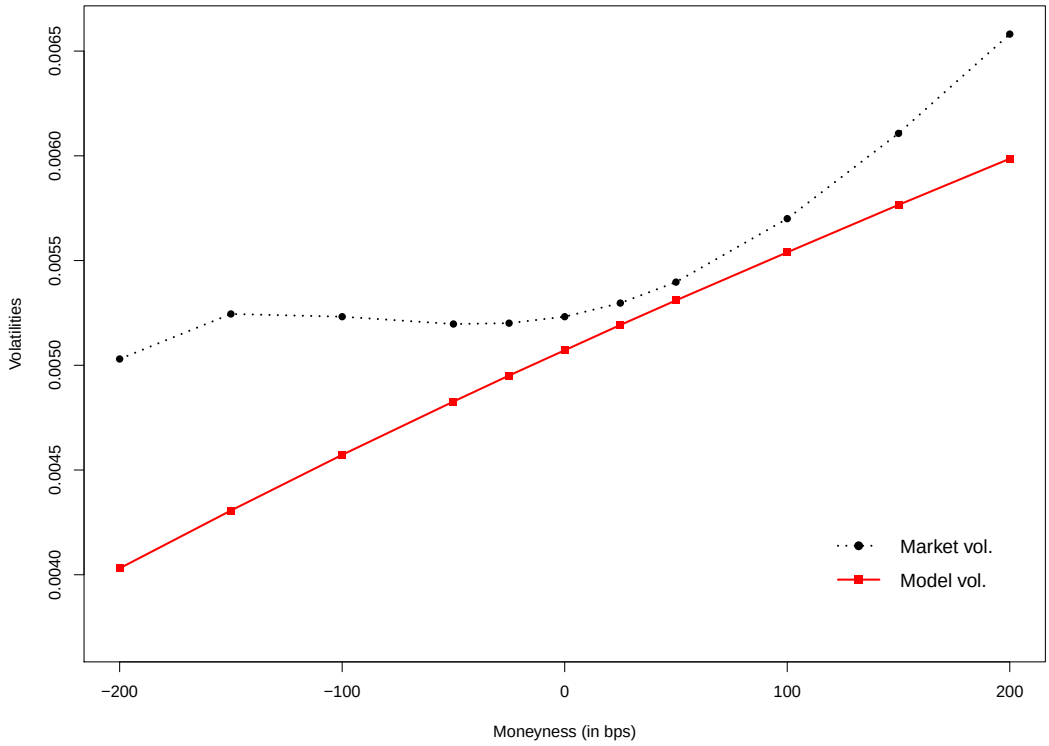


Figure 1.9: Smile induced by the DDLMM calibrated to 12/31/2020 Euro swaptions prices (on the particular 10×10 swaption).

δ (%)	a	b ($\times 10^{-2}$)	c ($\times 10^{-2}$)	d	κ ($\times 10^{-2}$)	θ	Objective func. ($\times 10^{-3}$)
5	$3.07 \cdot 10^{-6}$	3.08	9.58	$4.22 \cdot 10^{-3}$	3.99	$1.52 \cdot 10^{-1}$	6.85
7	$4.03 \cdot 10^{-3}$	2.09	8.86	$1.09 \cdot 10^{-5}$	4.25	$3.04 \cdot 10^{-1}$	6.00
10	$2.43 \cdot 10^{-3}$	1.40	8.55	$4.24 \cdot 10^{-4}$	2.43	$1.18 \cdot 10^{-7}$	5.78
12	$3.19 \cdot 10^{-3}$	1.12	8.29	$8.95 \cdot 10^{-6}$	2.19	$1.25 \cdot 10^{-1}$	5.88
15	$3.11 \cdot 10^{-4}$	0.903	8.38	$1.51 \cdot 10^{-3}$	2.21	$2.13 \cdot 10^{-1}$	5.96
20	$1.84 \cdot 10^{-10}$	0.493	7.22	$5.82 \cdot 10^{-3}$	4.18	$4.38 \cdot 10^{-1}$	7.76

Table 1.3: Calibrated parameters of the DD-LMM; shift δ is fixed to different values.

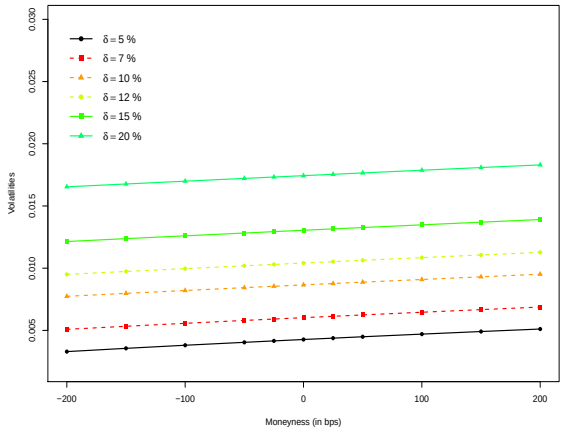
1.5.4.1.2 Constant-Elasticity-of-Variance LMM

The CEV version of the LMM has been introduced in [AP07]. In a similar framework to that of the DD-LMM, the dynamics of the k-th forward rates under the T_{k+1}-forward neutral probability measure $\mathbb{P}^{k+1}$ is assumed to be

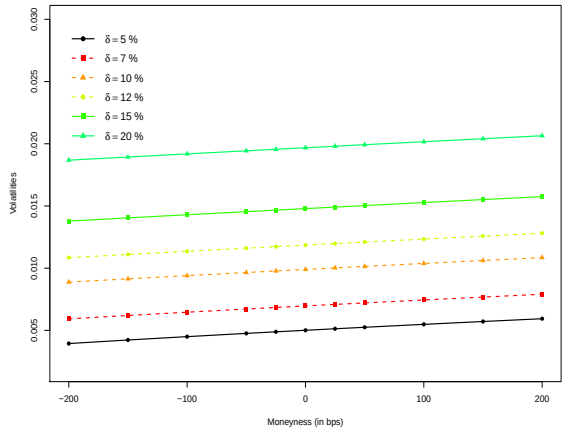
$$\mathrm{d}F_k(t) = \big(F_k(t)\big)^{\alpha} \boldsymbol{\gamma}_k(t) \cdot \mathrm{d}\boldsymbol{W}_t^{k+1}, \ t \leq T_k, \tag{1.44}$$

where $\boldsymbol{\gamma}_k$ is an N_f-dimensional deterministic time dependent vector and $\alpha \geq 0$. Prices for caplet/floorlet can be obtained in closed form as proved in Theorem 3 of [AA00].

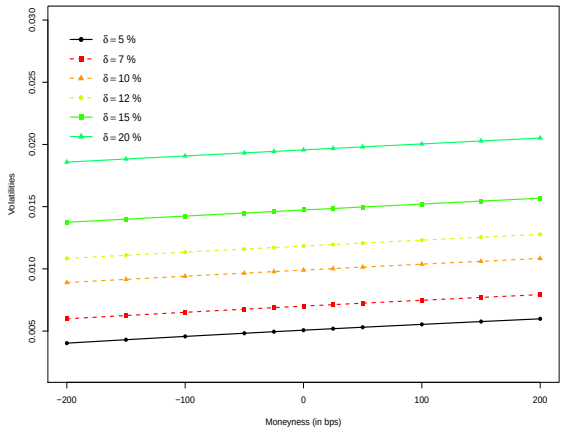
Similarly as for the DD-LMM, application of the Itô's formula yields the dynamics for the swap rate process. Again, the obtained time evolution is too hard to be handled as is states and



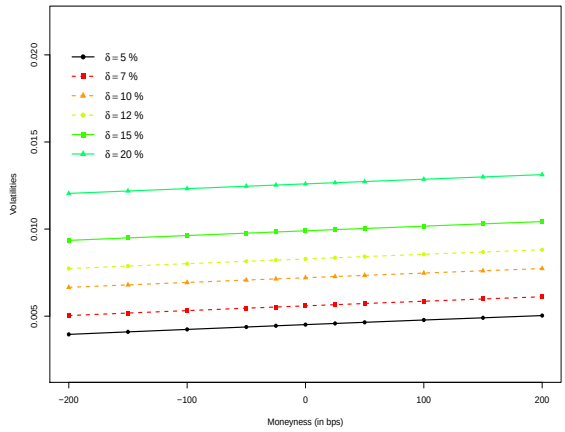
(a) Swaptions prices of maturity 1 year.



(b) Swaptions prices of maturity 5 years.



(c) Swaptions prices of maturity 10 years.



(d) Swaptions prices of maturity 30 years.

Figure 1.10: Impact of the shift coefficient on the smile in the DD-LMM all other parameters being fixed. The smile is computed using theoretical formulas obtained in the DD-LMM.

an approximation based on freezing technique is usually made to obtain the following dynamics, written under $\mathbb{P}^S$:

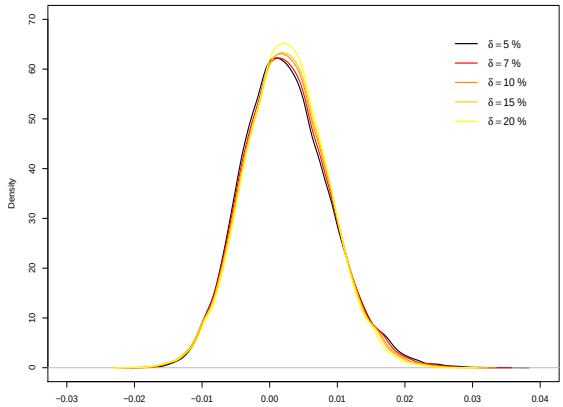
$$\mathrm{d}S_t^{m,n} = \left(S_t^{m,n}\right)^{\alpha} \sum_{k=m}^{n-1} \tilde{\gamma}_k(t) \cdot \mathrm{d}\boldsymbol{W}_t^S, \ t \leq T_m, \tag{1.45}$$

where $(\boldsymbol{W}_t^S)_{t \leq T_m}$ is a Brownian motion and $\tilde{\gamma}_k(t) := \partial_k \psi(0)\big(F_k(0)/S_0^{m,n}\big)^{\alpha}\gamma_k(t)$. In this modelling framework, approximated swaption prices can be derived. The result is proved in [AA00].

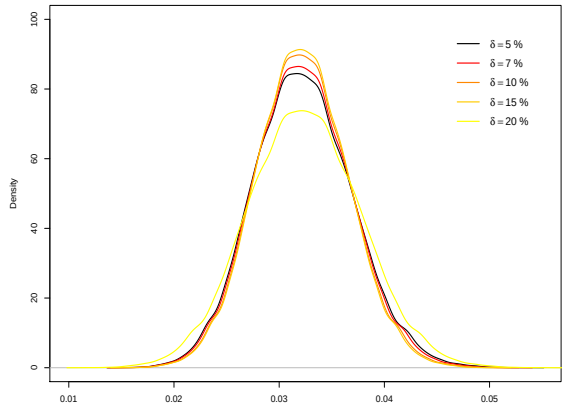
Proposition 1.7 ([AA00])**.** *Denote by* $a = K^{2(1-\alpha)}(1-\alpha)^{-2}/(T_m - t)\tilde{v}_m^2(t)$, $b = (1-\alpha)^{-1}$, $c = (S_t^{m,n})^{2(1-\alpha)}(1-\alpha)^{-2}/(T_m - t)\tilde{v}_m^2(t)$, $\tilde{v}_m^2(t) = \int_t^{T_m} \| \sum_{k=m}^{n-1} \tilde{\gamma}_k(u)\|^2 \mathrm{d}u/(T_m - t)$ *and* $x_{\pm} = \big(\ln(S_t^{m,n}/K) \pm (T_m - t)\tilde{v}_m^2(t)/2\big)/\sqrt{(T_m - t)\tilde{v}_m^2(t)}$. *Then, the swaption price writes*

$$PS(t, T_m, T_n, \boldsymbol{\mathcal{T}'}, K) = \begin{cases} B^S(t)\big(S_t^{m,n}\big[1 - \chi^2(a, b+2, c)\big] - K\chi^2(c, b, a)\big), \ 0 < \alpha < 1, \\ B^S(t)\big(S_t^{m,n}\Phi(x_+) - K\Phi(x_-)\big), \ \alpha = 1, \\ B^S(t)\big(S_t^{m,n}\big[1 - \chi^2(c, -b, a)\big] - K\chi^2(a, 2-b, c)\big), \ 1 < \alpha, \end{cases}$$

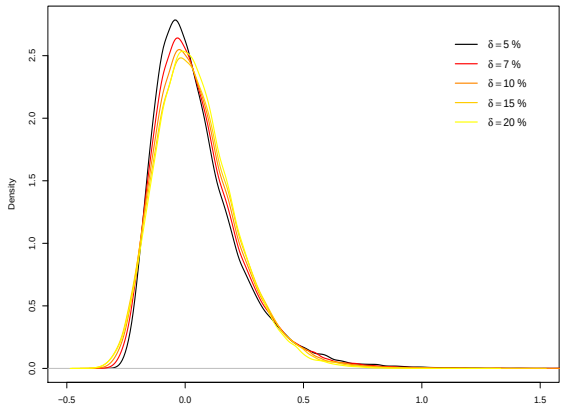
where Φ *is the distribution function of the Gaussian law and* $x \mapsto \chi^2(x, k, \lambda)$ *is that of the non*



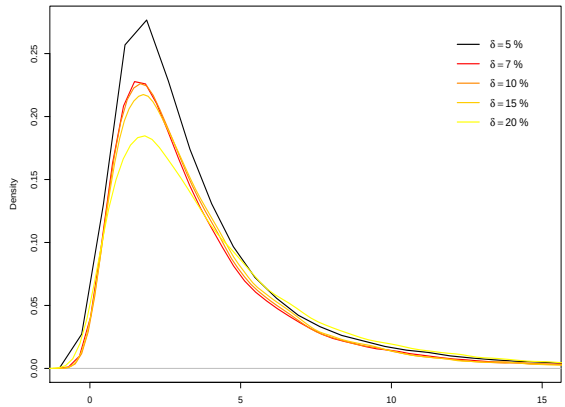
(a) Initial densities of 10-year forward rates.



(b) Initial densities of 30-year forward rates.



(c) Terminal densities of 10-year forward rates.



(d) Terminal densities of 30-year forward rates.

Figure 1.11: Initial and terminal densities of forward rates induced by the DD-LMM calibrated with different shifts. Empirical densities estimated on 50000 paths.

central chi-squared distribution with k degree of freedom and non centrality parameter λ.

Remark 4. *As detailed in [HW99a], this modelling framework allows to replicate market smiles for $\alpha \neq 1$.*

Remark 5. *This modelling framework can be enriched by considering that the shifted forward rate process $\tilde{F}_k(\cdot) = F_k(\cdot) + \delta$ satisfies (1.44).*

1.5.4.2 Displaced Diffusion with Stochastic Volatility LMM

In stochastic volatility models, the instantaneous volatility process is represented by as a stochastic process in order to be able to replicate swaptions prices over a grid of various strikes. The issue is to model this additional risk factor while preserving the tractability of the whole model.

[WZ06] proposed to model the volatility factor using the popular Cox-Ingersoll-Ross dynamics; adding a shift parameter accounting for non-positive interest rates yields the Displaced Diffusion and Stochastic Volatility version of the LMM. It is usually referred to as the DDSVLMM or LMM+ by practitioners and is a very popular model among insurers who aim at replicating market smile.

The modelling framework is very similar to that of the DD-LMM. The volatility structure of the k-th shifted forward rate is now assumed to be given by

$$\boldsymbol{\zeta}_k(t) := \sqrt{V_t}\boldsymbol{\gamma_k}(t),$$

where $(V_t)_{t\leq T^\star}$ lies in the family of Feller processes and $\boldsymbol{\gamma_k}$ is a deterministic time-dependent vector. Under the spot-LIBOR measure, the volatility process is represented by the dynamics

$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}W_t,\ t \leq T_m, \tag{1.46}$$

where κ, θ, ϵ are non-negative parameters and $(W_t)_{t\leq T^\star}$ is a Brownian motion. Imposing in addition the so-called Feller's condition $2\kappa\theta \geq \epsilon^2$ allows to ensure the positivity of the volatility factor through time as long as $V_0 > 0$. [WZ06] allows for correlation between the forward rates process and the volatility process. First, let us recall that the Brownian motion driving the k-th forward rate is denoted by $\left(\boldsymbol{W}_t^{k+1}\right)_{t\geq 0}$. We introduce the time dependent vector $\boldsymbol{\vartheta}_k$ whose j-th component is such that

$$\mathrm{d}\left\langle \boldsymbol{W}_\cdot^{k+1,(j)}, W_\cdot \right\rangle_t = \boldsymbol{\vartheta}_k^{(j)}(t)\mathrm{d}t,$$

where $\boldsymbol{W}^{k+1,(j)}$ stands for the j-th component of the multidimensional vector $\boldsymbol{W}^{k+1}$. The correlation structure between the two processes is captured by the coefficient ρ_k defined as

$$\mathrm{d}\left\langle \frac{\boldsymbol{\gamma_k}(.)}{\|\boldsymbol{\gamma_k}(.)\|} \cdot \boldsymbol{W}_\cdot^{k+1}, W_\cdot \right\rangle_t = \frac{\boldsymbol{\gamma_k}(t)}{\|\boldsymbol{\gamma_k}(t)\|} \cdot \boldsymbol{\vartheta}_k^{(j)}(t)\mathrm{d}t =: \rho_k(t)\mathrm{d}t. \tag{1.47}$$

Under the forward swap measure $\mathbb{P}^S$, the dynamics of the swap rate process follows from the Itô's formula as, for $t \leq T_m$,

$$\begin{aligned} \mathrm{d}S_t^{m,n} &= \sqrt{V_t}\sum_{k=m}^{n-1} \partial_k\psi(t)\big(F_k(t)+\delta\big)\boldsymbol{\gamma_k}(t)\cdot \mathrm{d}\boldsymbol{W}_t^S, \\ \mathrm{d}V_t &= \kappa\big(\theta - \xi(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}W_t^S, \end{aligned} \tag{1.48}$$

where $\boldsymbol{W}^S$ is a multidimensional Brownian motion, W^S is Brownian motion and we set $\xi(t) = 1 + \frac{\epsilon}{\kappa}\sum_{j=m}^{n-1}\alpha_j(t)\sum_{k=1}^{j}\frac{\Delta T_k(F_k(t)+\delta)}{1+\Delta T_k F_k(t)}\rho_k(t)\|\boldsymbol{\gamma_k}(t)\|$. The obtained dynamics is here again too complex for allowing the obtention of swaption prices by closed form formula.

Black framework Still using the freezing approximation method, the previous dynamics can be approximated by a log-normal type dynamics that writes as

$$\begin{aligned} \mathrm{d}S_t^{m,n} &= \sqrt{V_t}\big(S_t^{m,n}+\delta\big)\sum_{k=m}^{n-1}\tilde{\boldsymbol{\gamma_k}}(t)\cdot \mathrm{d}\boldsymbol{W}_t^S, \\ \mathrm{d}V_t &= \kappa\big(\theta - \xi_0(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}W_t^S, \end{aligned} \quad t \leq T_m \tag{1.49}$$

where $\tilde{\boldsymbol{\gamma_k}}(t) := \frac{\partial_k\psi(0)(F_k(0)+\delta)}{S_0^{m,n}+\delta}\boldsymbol{\gamma_k}(t)$ and $\xi_0(t) = 1 + \frac{\epsilon}{\kappa}\sum_{j=m}^{n-1}\alpha_j(0)\sum_{k=1}^{j}\frac{\Delta T_k(F_k(0)+\delta)}{1+\Delta T_k F_k(0)}\rho_k(t)\|\boldsymbol{\gamma_k}(t)\|$.

The deterministic deformation of the volatility structure $(\boldsymbol{\gamma_k})_{k\in\{m,\dots,n-1\}}$ is assumed to be piecewise constant on the grid $[T_j, T_{j+1}[$, $j = 1, \dots, m$. In this parametrization, the log-process of (1.49) belongs to the class of affine process for which analytical knowledge of the moment generating function is known, through resolution of some Riccati equations. The details of the computation can be found in [WZ06] or [LDB20]. Semi-analytical swaptions prices can then

be derived based on the integration of the characteristic function of the log-shifted swap rate. Note that these semi-analytical prices are simply *approximation* prices since they are obtained after *freezing* of (1.48).

In the following, the moment generating function of the log-swap rate process is defined over its definition domain $\mathcal{D} \subset \mathbb{C}$ as

$$\Psi(x; t, S_t^{m,n}, V_t) = \mathbb{E}^S \left[\exp\left(x \ln\left(\frac{S_{T_m}^{m,n} + \delta}{S_0^{m,n} + \delta} \right) \right) \middle| \mathcal{F}_t \right], \ x \in \mathcal{D}.$$

Proposition 1.8 (**Log-normal swaption pricing under DDSVLMM**). *In the DDSVLMM, swaption prices (1.13) express as*

$$PS(t, T_m, T_n, \boldsymbol{\mathcal{T}'}, K) = B^S(0)\big((S_0^{m,n} + \delta)P_1 - (K + \delta)P_2\big),$$

where

$$P_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^{+\infty} \frac{1}{u} Im\left(e^{-iuK^*} \Psi(1 + iu; 0, s_0^{m,n}, v_0) \right) \mathrm{d}u,$$
$$P_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^{+\infty} \frac{1}{u} Im\left(e^{-iuK^*} \Psi(iu; 0, s_0^{m,n}, v_0) \right) \mathrm{d}u,$$

and where we set $K^* = \ln \frac{K+\delta}{S_0^{m,n}+\delta}$.

Proof. The proof of this formula can be found in [DPS00] or in [GP51]. □

Bachelier framework The definitions of the functions $t \mapsto \tilde{\boldsymbol{\gamma}}_{\boldsymbol{k}}(t)$ are based on the assumption of low variability through time of the coefficients $\partial_k \psi(0)(F_k(0) + \delta)/(S_0^{m,n} + \delta$. Based on alternative assumption, those coefficients can be defined as

$$\tilde{\boldsymbol{\gamma}}_{\boldsymbol{k}}(t) := \partial_k \psi(0)(F_k(0) + \delta)\boldsymbol{\gamma}_{\boldsymbol{k}}(t)$$

which yields a normal type dynamics for the swap rate process

$$\begin{aligned} \mathrm{d}S_t^{m,n} &= \sqrt{V_t} \sum_{k=m}^{n-1} \tilde{\boldsymbol{\gamma}}_{\boldsymbol{k}}(t) \cdot \mathrm{d}\boldsymbol{W}_t^S, \\ \mathrm{d}V_t &= \kappa\big(\theta - \xi_0(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}W_t^S, \end{aligned} \qquad t \le T_m. \tag{1.50}$$

[DPS00] provided (semi-) analytical formulas for a number of prices of securities. We detail the derivation of the swaption price under (1.50) and following their work.

In the following, the moment generating function of the swap rate process is defined over its definition domain $\mathcal{D} \subset \mathbb{C}$ as

$$\Psi(x; t, S_t^{m,n}, V_t) = \mathbb{E}^S \left[e^{x S_{T_m}^{m,n}} \middle| \mathcal{F}_t \right], \ x \in \mathcal{D}.$$

Proposition 1.9 (**Normal swaption pricing under DDSVLMM**). *In the DDSVLMM, swaption prices (1.13) express as*

$$PS(t, T_m, T_n, \boldsymbol{\mathcal{T}'}, K) = B^S(0)\big(P_1 - KP_2\big),$$

where

$$P_1 = \frac{1}{2}S_0^{m,n} - \frac{1}{\pi}\int_0^{+\infty} Im\left(\frac{\Psi'(-iu;0,s_0^{m,n},v_0)e^{iuK}}{u}\right)\mathrm{d}u,$$
$$P_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} Im\left(\frac{e^{-iuK}\Psi(iu;0,s_0^{m,n},v_0)}{u}\right)\mathrm{d}u.$$

Proof. We follow the proof in [GP51]. First, we write the classical decomposition $PS(t,T_m,T_n,\boldsymbol{\mathcal{T}'},K) = B^S(0)\Big(\mathbb{E}^S\big[S_{T_m}^{m,n}\mathbb{1}_{\{S_{T_m}^{m,n}\geq K\}}\big] - K\mathbb{E}^S\big[\mathbb{1}_{\{S_{T_m}^{m,n}\geq K\}}\big]\Big)$. Let us denote by $\phi(z) := \mathbb{E}\left[e^{zS_{T_m}^{m,n}}\right] = \Psi(z;0,s_0^{m,n},v_0)$ the characteristic function of $S_{T_m}^{m,n}$ defined for $u \in \mathcal{D}$ and by μ the density function of the random variable $S_{T_m}^{m,n}$ with respect to the Lebesgue measure. First, let us consider $(\eta, A) \in (\mathbb{R}_+)^2$, $\eta \leq A$. Observe that

$$\begin{aligned}\int_\eta^A \mathrm{Im}\left(\frac{\phi'(-ix)e^{ixK}}{x}\right)\mathrm{d}x &= \int_\eta^A\int_{\mathbb{R}} y\frac{e^{ix(K-y)} - e^{-ix(K-y)}}{2ix}\mu(y)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}} y\mu(y)\int_\eta^A \frac{\sin\big(x(K-y)\big)}{x}\mathrm{d}x\mathrm{d}y\end{aligned}$$

where the Fubini theorem is licit since, for all $x \in [\eta, A]$, any $K \in \mathbb{R}$,

$$\left|\frac{\sin\big(x(K-y)\big)}{x}\right| < \frac{1}{\eta}.$$

As in [GP51], consider the sign function $\mathrm{sgn}(x) = \mathbb{1}_{x>0} - \mathbb{1}_{x<0}$ and note that

$$\mathrm{sgn}(K-y) = \frac{2}{\pi}\int_0^\infty \frac{\sin(x(K-y))}{x}\mathrm{d}x$$

holds for any $K \in \mathbb{R}$. Furthermore, note that $(\eta, A) \mapsto \int_\eta^A \frac{\sin\left(x(K-y)\right)}{x}\mathrm{d}x$ is continuous and simultaneously bounded in both of its argument. We finally get that

$$\begin{aligned}\int_0^\infty \mathrm{Im}\left(\frac{\phi'(-ix)e^{ixK}}{x}\right)\mathrm{d}x &= \lim_{(\eta,A)\to(0,\infty)}\int_\eta^A \mathrm{Im}\left(\frac{\phi'(-ix)e^{ixK}}{x}\right)\mathrm{d}x \\ &= \int_{\mathbb{R}} y\mu(y)\lim_{(\eta,A)\to(0,\infty)}\int_\eta^A \frac{\sin\big(x(K-y)\big)}{x}\mathrm{d}x\mathrm{d}y \\ &= \frac{\pi}{2}\int_{\mathbb{R}} \mathrm{sgn}(K-y)y\mu(y)\mathrm{d}y \\ &= \frac{\pi}{2}\Big(\int_{-\infty}^K y\mu(y)\mathrm{d}y - \int_K^\infty y\mu(y)\mathrm{d}y\Big) \\ &= \frac{\pi}{2}\Big(\mathbb{E}\left[S_{T_m}^{m,n}\right] - 2\mathbb{E}\left[S_{T_m}^{m,n}\mathbb{1}_{\{S_{T_m}^{m,n}\geq K\}}\right]\Big) \\ &= \frac{\pi}{2}\Big(S_0^{m,n} - 2\mathbb{E}\left[S_{T_m}^{m,n}\mathbb{1}_{\{S_{T_m}^{m,n}\geq K\}}\right]\Big)\end{aligned}$$

Similar expression for $\mathbb{P}\big(S_{T_m}^{m,n} \geq K\big)$ can be derived to prove the claim. □

Let us finally mention that an alternative method for computing integrals of Fourier transforms of characteristic functions associated with stochastic volatility models such as Heston

one in terms of cosine expansions (that are related to Chebyshev series) has been proposed in [Fan10].

Chapter 2

Freezing approximations for the calibration of the Displaced Diffusion LIBOR Market Model

In present chapter, we discuss in more details the *freezing* techniques used to make the different versions of the LIBOR Market Model tractable. It consists in approximating some relevant stochastic quantities by their initial values to obtain dynamics whose coefficients are therefore deterministic functions of time. Two common techniques are discussed and compare: the so-called *Rebonato* and *Hull & White* approximations.

2.1 Introduction

In the previous chapter, a number of interest rates models derived in the LIBOR Market Model were too complex to be calibrated in their primary forms as a number of stochastic processes are involved in the drift and diffusion coefficients of dynamics defining the models. Yet it is possible to simulate those models thanks to discretization schemes and compute prices using Monte-Carlo methods. However such approaches are time and computationally consuming. To overcome this difficulty approximations can be made to get tractable models –in the sense that they allow the derivation of closed-form formula: this is the so-called *freezing technique*. Based on a low variability assumption of well chosen quantities -allowing to replace some random quantities by their initial values-, some approximating prices can be obtained by closed-form formula. The problem of the induced error arises then. In this section, we provide details on the approximation technique, analyse its relevance and propose some alternative methods.

2.2 Standard freezing method

We retake the notations used in Sections 1.5.3.1 and 1.5.4.1 introducing the *Displaced Diffusion Libor Market Model* (DDLMM). The objective is to compute swaptions prices in the standard DDLMM. Those are quantities proportional to $\mathbb{E}^S[(S_T^{m,n} - K)_+]$. To compute it, Monte-Carlo simulations are necessary as the drift and the diffusion coefficients of the SDE (see the exact modelling dynamics of swap rate (1.41)) defining the DDLMM are state dependent. The use of such method is not practicable. Researchers/practitioners have proposed to somehow approximate the stochastic processes involved in the DDLMM in the hope to be able to easier compute security prices. Those are thus approximating prices that are wanted to be as close as possible of the exact prices. In the following, we focus on swaptions prices.

Beforehand, recall some key fact about the swap rate process. It expresses as a non linear transformation of the forward rates as

$$S_t^{m,n} + \delta = \sum_{j=m}^{n-1} \alpha_j(t)\big(F_j(t) + \delta\big) \tag{2.1}$$

where the stochastic weights are $\alpha_j(t) = \frac{\Delta T_j P(t,T_{j+1})}{\sum_{k=m}^{n-1} \Delta T_k P(t,T_{k+1})}$ and $\delta \geq 0$ is the displacement coefficient. It is stressed that these stochastic weights also depend on the forward rates according to

$$\alpha_j(t) = f_j(F_m(t), \cdots, F_{n-1}(t)) = \frac{\Delta T_j \prod_{k=m}^{j} \frac{1}{1+\Delta T_k F_k(t)}}{\sum_{k=m}^{n-1} \Delta T_k \prod_{l=m}^{k} \frac{1}{1+\Delta T_l F_l(t)}}. \tag{2.2}$$

The dynamics describing the evolution of the swap rate process in the DDLMM under forward swap measure (making the swap rate a martingale) can be derived using Itô's formula:

$$\mathrm{d}\big(S_t^{m,n} + \delta\big) = \mathrm{d}S_t^{m,n} = \sum_{j=m}^{n-1} \partial_j \psi(t)\big(F_j(t) + \delta\big)\boldsymbol{\gamma}_j(t) \cdot \mathrm{d}\boldsymbol{W}_t^S,\ t \leq T_m, \tag{2.3}$$

where we set (see for instance [WZ06] for the details of the computations)

$$\partial_j \psi(t) = \alpha_j(t) + \frac{\Delta T_j}{1 + \Delta T_j F_j(t)} \sum_{l=m}^{j-1} \alpha_l(t)\big(F_l(t) - S_t^{m,n}\big). \tag{2.4}$$

2.2.1 Rebonato

The first approximation is due to Riccardo Rebonato since he was the first to approximate swap rate as linear combination of forward rates in [Reb98]. The basic assumption in this approach is to consider the weights α_j are of low variability through time. The freezing method proposed thus to write swap rate as linear combination of forward rates:

$$S_t^{m,n} + \delta \approx \sum_{j=m}^{n-1} \alpha_j(0)\big(F_j(t) + \delta\big).$$

Recalling that LMM assigns a log-normal volatility structure to the forward rates, the swap rate dynamics under the forward swap measure can thus be written as $\mathrm{d}S_t^{m,n} \approx \sum_{j=m}^{n-1} \alpha_j(0) F_j(t) \boldsymbol{\gamma_j}(t) \cdot \mathrm{d}\boldsymbol{W}_t^S$. Further assumption of low variability through time of the forward rates yields the Rebonato freezing approximation of (2.3) as

$$\mathrm{d}S_t^{m,n} = \sum_{j=m}^{n-1} \alpha_j(0)\big(F_j(0) + \delta\big)\boldsymbol{\gamma_j}(t) \cdot \mathrm{d}\boldsymbol{W}_t^S\ t \leq T_m. \tag{2.5}$$

The obtained approximated swap rate process is a Bachelier type dynamics with time dependent coefficient. In this setting, Bachelier formula can be employed to price swaptions. In particular, recalling that correlation structure between forward rates is given by $\mathrm{d}\left\langle F_k(.) + \delta, F_j(.) + \delta \right\rangle_t =$

$\big(F_k(t)+\delta\big)\big(F_j(t)+\delta\big)\boldsymbol{\gamma_k}(t)\cdot\boldsymbol{\gamma_j}(t)\mathrm{d}t$, the implied volatility parameter under (2.5) writes

$$\big(\sigma^{\mathrm{Reb}}_{m,n}\big)^2=\frac{1}{T_m}\sum_{k,j=m}^{n-1}\alpha_j(0)\alpha_k(0)\big(F_j(0)+\delta\big)\big(F_k(0)+\delta\big)\int_0^{T_m}\boldsymbol{\gamma_j}(t)\cdot\boldsymbol{\gamma_k}(t)\mathrm{d}t. \tag{2.6}$$

Remark 6. *For Black pricing convention, an alternative freezing approximation can be done. Based on the assumption of low variability of the ratios $F_j(t)/S^{m,n}_t$, (2.6) rewrites as*

$$\big(\sigma^{Reb}_{m,n}\big)^2=\frac{1}{T_m}\sum_{k,j=m}^{n-1}\frac{\alpha_j(0)\alpha_k(0)\big(F_j(0)+\delta\big)\big(F_k(0)+\delta\big)}{\big(S^{m,n}_0+\delta\big)^2}\int_0^{T_m}\boldsymbol{\gamma_j}(t)\cdot\boldsymbol{\gamma_k}(t)\mathrm{d}t. \tag{2.7}$$

2.2.2 Hull & White

In [HW99b], authors proposed to approximate directly the dynamics (2.3) derived using Itô's lemma. The method is based on the assumption that the coefficients $\partial_k\psi(t)F_k(t)$ are of low variability. The approximating dynamics writes then

$$\mathrm{d}S^{m,n}_t=\sum_{j=m}^{n-1}\partial_k\psi(0)\big(F_j(0)+\delta\big)\boldsymbol{\gamma_j}(t)\cdot\mathrm{d}\boldsymbol{W}^S_t\ t\le T_m. \tag{2.8}$$

Observe that $\partial_j\psi(0)=\alpha_j(0)+\sum_{k=m}^{n-1}F_k(0)\frac{\partial\alpha_k(0)}{\partial F_j}$ and with (2.2), one can recover the expression of $\partial_j\psi(0)$ as given in (2.4). Again, the obtained dynamics is a Bachelier type one and implied volatility parameter associated to swaption price can be derived as

$$\big(\sigma^{\mathrm{HW}}_{m,n}\big)^2=\frac{1}{T_m}\sum_{k,j=m}^{n-1}\partial_j\psi(0)\partial_k\psi(0)\big(F_j(0)+\delta\big)\big(F_k(0)+\delta\big)\int_0^{T_m}\boldsymbol{\gamma_j}(t)\cdot\boldsymbol{\gamma_k}(t)\mathrm{d}t. \tag{2.9}$$

Remark 7. *For Black pricing convention, an alternative freezing approximation can be done. Based on the assumption of low variability of the ratios $F_j(t)/S^{m,n}_t$, (2.9) rewrites as*

$$\big(\sigma^{HW}_{m,n}\big)^2=\frac{1}{T_m}\sum_{k,j=m}^{n-1}\frac{\partial_j\psi(0)\partial_k\psi(0)\big(F_j(0)+\delta\big)\big(F_k(0)+\delta\big)}{\big(S^{m,n}_0+\delta\big)^2}\int_0^{T_m}\boldsymbol{\gamma_j}(t)\cdot\boldsymbol{\gamma_k}(t)\mathrm{d}t. \tag{2.10}$$

Before going into the technical details of the numerical analysis, we clarify the quantities we will manipulate.

1. **Implied volatilities:** they refer to market data. As discussed in Chapter 1 Section 1.4.3.2, market data often correspond to implied volatilities, that is volatility parameter that should be plugged into a pricing formula to recover the price of the derivative. Following the choice of the pricing formula, we speak of Bachelier or Black volatilities. This information is given by data providers and reflect the current best practices.

2. **Exact volatilities:** those can alternatively referred to as Monte-Carlo volatilities. Under a given set of parameters, primary models can be simulated thanks to discretization scheme; Monte-Carlo methods allow then to compute swaptions prices. Inverting the obtained prices using one of the two mentioned pricing formula allows to recover simulated implied volatilities: we will refer to them as *exact volatilities* in contrast with volatilities obtained thanks to the studied *frozen* approximations. Strictly speaking, they are not exact volatilities as they embed discretization error.

3. **Theoretical volatilities**: those are obtained once the *freezing approximation* has been made. Namely, the refer to the formula (2.6) or (2.9) (or their Black counterparts given in Remarks 6 and 7).

2.2.3 Historical estimations

We propose to first check on historical time series how behave the different *frozen* quantities. The computations below are performed on Euro data between 12/31/1999 and 12/31/2020 and illustrated below in Figure 2.1.

First in Figure 2.1a, we plot the historical time series of the stochastic weights recalled in (2.1) for different maturities and tenors. Namely, (α_1, α_4) (respectively (α_5, α_9), $(\alpha_{10}, \alpha_{19})$ and $(\alpha_{20}, \alpha_{39})$) are associated to 1×5 (respectively 5×5, 10×10, 20×20) swaption. We observe that the low variability assumption of these coefficients over time is admissible on empirical data. The standard deviations of each time series support this assumption. Observe that the gap between each pair of weights goes to zero on late market data: it is a consequence of the late shape of the market curve in which the level of long-term maturities are below mid-term ones.

In Figure 2.1b are plotted the weights induced by the Itô's lemma when deriving the dynamics of the swap rate and defined in (2.4). (∂_2, ∂_4) (respectively (∂_6, ∂_9), $(\partial_{11}, \partial_{19})$) are associated with 1×5 (respectively 5×5, 10×10) swaption. Again the low variability assumption can deemed receivable: the standard deviations of each time series are similar to that obtained in Figure 2.1a. Observe also that the gap between weights of different maturities also goes to zero as time passes.

Forward rates of maturity 1 (resp. 5, 10, 20, 30) year are depicted in Figure 2.1c: standard deviations on historical data are larger than those obtained on the weights but still quite low. The low variability assumption under Bachelier framework as in (2.5) or in (2.8) is more questionable in that case but can be deemed acceptable.

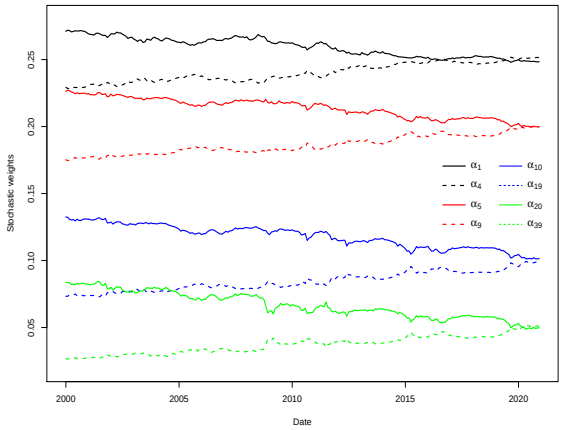
In Figure 2.1d are plotted the ratios $F_1/S^{1,5}$ (resp. $F_4/S^{1,5}$, $F_5/S^{5,5}$, $F_9/S^{5,5}$, $F_{10}/S^{10,10}$, $F_{19}/S^{10,10}$, $F_{20}/S^{20,20}$, $F_{39}/S^{20,20}$). Regarding the low variability assumption of quantities under Black framework, appearing in (2.7) or (2.10), we observe in Figure 2.1d that the approximation can be accepted on ancient market data (before 2015, roughly speaking). But, in late market condition with low and even non-positive rates it is clear that this assumption is no longer valid. This empirical observation contributes to explain why the Bachelier framework is more suited to current low rates environment.

2.2.4 Pricing error

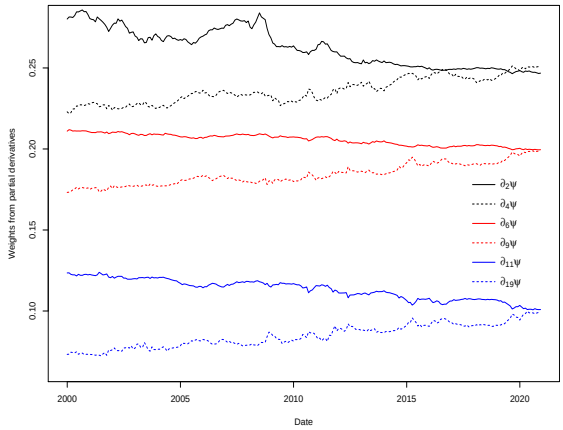
We now turn into the analyze of the error induced by each freezing approximation in term of swaptions pricing and thus in term of error induced on terminal distribution of swap rates. In particular, we estimate the distance between two models in some sense.

2.2.4.1 Bachelier convention

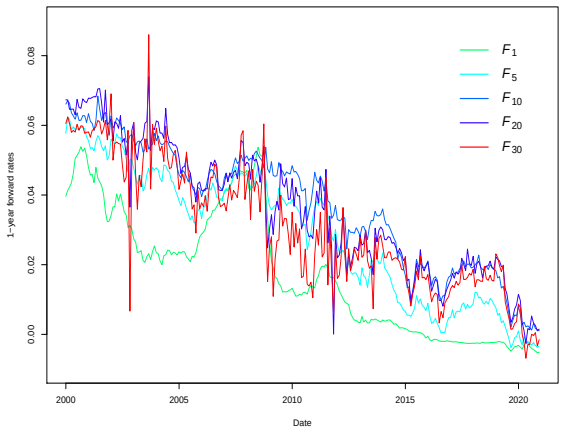
We first propose to numerically justify the previous assumptions from the pricing error point of view. [BM07] and [BDB01] already studied the relevance of the Rebonato and Hull & White freezing approximations. Under a given set of parameters, they compare the exact volatilities to the theoretical ones given in (2.7) and (2.10) (the use of Bachelier convention was not standard at that time). Both volatilities they obtained were close resulting in close prices that validates the freezing approximations. However, their studies relied on the Black (log-normal) pricing convention and was led on derivatives of relatively short maturities. For insurance needs, we



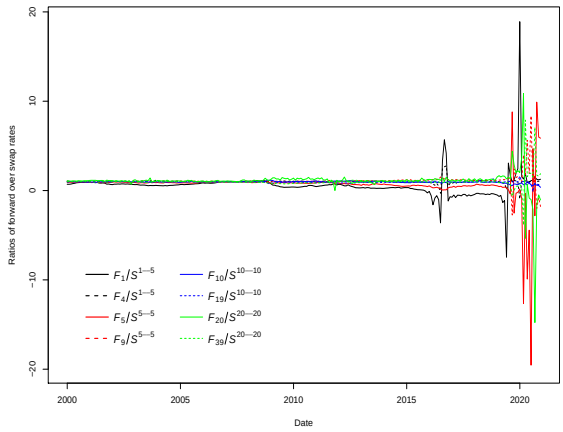
(a) Historical time series of stochastic weights $(\alpha_j(0))_{j\in\{1,4,5,9,10,19,20,29\}}$. Standard deviations of time series range in between 0.0066 and 0.0097.



(b) Historical time series of stochastic weights $(\partial_j\psi(0))_{j\in\{2,4,6,9,11,19\}}$. Standard deviations of time series range in between 0.0035 and 0.012.



(c) Historical time series of 1-year forward rates $(F_j(0))_{j\in\{1,5,10,20,30\}}$. Standard deviations of time series range in between 0.018 and 0.019.



(d) Historical time series of ratios of forward rates over swap rates. Standard deviations of time series range in between 0.097 and 1.98.

Figure 2.1: Empirical check of the different low variability assumptions made in Rebonato and Hull & White methods.

propose to extend their analysis of the *freezing* technique when using the Bachelier pricing formula and in particular for swaptions of long maturities.

The methodology we propose is the following: we calibrate the dynamics (2.3) using the afore mentioned *freezing* approximations (2.5) or (2.8) on market data. An accurate replication –measured by the distance between *frozen* and market volatilities– of them would mean that the proposed approximations are flexible enough to fit market data. In a second time, we simulate the *exact* dynamics (2.3) without any *freezing* approximation to compute Monte-Carlo prices. Inversion of the Bachelier formula allows to compare the exact volatilities to both theoretical and implied ones: the distance between, on one hand, exact and theoretical volatilities –we will refer to as the *freezing* error– and on the other hand, exact and market volatilities –referred to as *calibration* error– allows to appreciate the relevance of the *freezing* approximations through multiple perspectives.

Calibration using *frozen* approximations

For the purpose of this section, we do not calibrate the displacement factor δ which was set to 0.02. Calibration consists in the best replication of 336 swaptions volatilities quoted on June 30th, 2020 in Euro market. Those are composed of both At-The-Money (ATM) and Away-From-the-Money (AFM) data. Maturities and tenors range into $\{1, \ldots, 10, 15, 20, 25, 30\}$ years (in the plots below, those are labelled as $\{1, 2, \ldots, 13, 14\}$). For AFM volatilities, we consider the same range of maturities but we focus on a 10-year reference tenor. The residual strikes[1] (in bps) range into $\pm\{25, 50, 100, 150, 200\}$. The target function used to calibrate the model is the normalized sum of the square of relative differences between market and model prices. Table 2.1 depicted the root mean squared relative error[2] (RMSE) obtained as outputs of the calibration.

Freezing method	RMSE-ATM	RMSE-AFM
Rebonato	0.07039	0.12
Hull & White	0.06996	0.1192

Table 2.1: Comparison of *freezing* techniques for calibration in Bachelier environment.

It turns out that both *freezing* approaches offer similar and accurate replications of market data on ATM data illustrating that the model is parametrized enough to reproduce market data. Not surprisingly, the model poorly perform on AFM data as the model does not induce smile on volatilities.

Simulations of *exact* quantities

The *exact* volatilities displayed here were computed on 10^5 paths. On ATM volatilities, the Hull & White assumption turns out to be the most robust as both the *freezing* and *calibration* errors are smaller than those induced by the Rebonato's approximation. The improvement is particularly material concerning for the *freezing* error though it is also appreciable on calibration error.

Freezing method	Error between M.C. vol. and ..	ATM	AFM
Rebonato	Theoretical vol.	0.05292	0.3211
	Market vol.	0.08543	0.3046
Hull & White	Theoretical vol.	0.02636	0.3225
	Market vol.	0.07503	0.3062

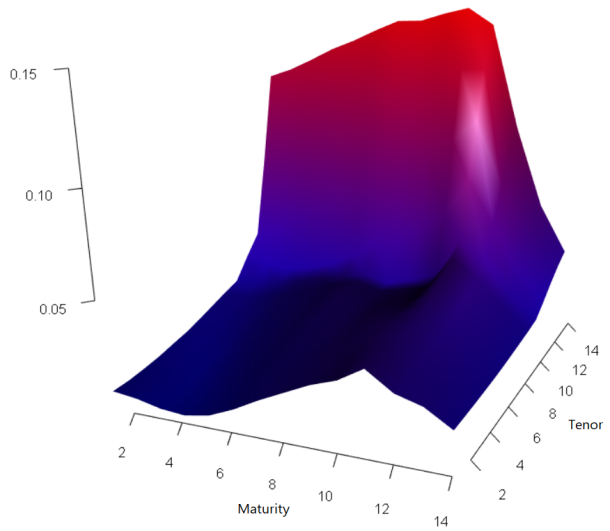
Table 2.2: Distance between exact volatilities and theoretical/market ones.

In Figure 2.2 the relative *freezing* error for both Rebonato and Hull & White assumptions is depicted. On ATM data, the errors between *frozen* volatilities and *exact* ones appear to be increasing functions of the maturity and tenor which seems intuitive. This monotonicity is particularly pronounced for the Rebonato approach. About AFM volatilities, we observe a strong dependency towards the level of moneyness for both approaches: the major part of the AFM error given in Table 2.2 comes from the most extreme strikes which was expected. In

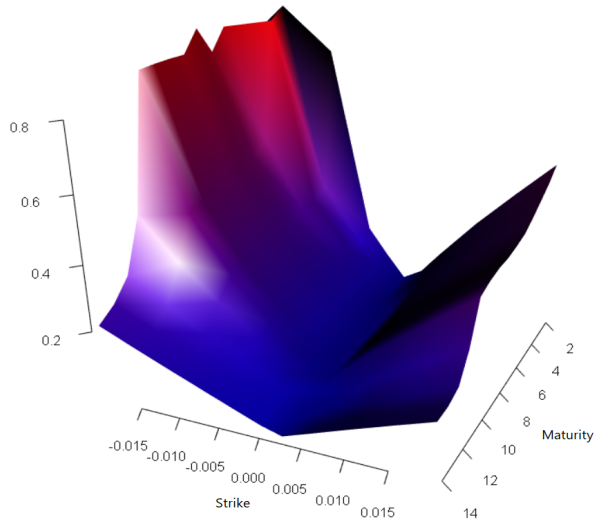
[1] Recall that residual strike of a swaption whose underlying spot value is $S_0^{m,n}$ and of strike K is defined as $k := K - S_0^{m,n}$.

[2] If $(\sigma_i^{Mod})_{i\in\{1,\ldots,N\}}$ stands for the set of closed-form volatilities and $(\sigma_i^{Mkt})_{i\in\{1,\ldots,N\}}$ represents the set of market ones, we define RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{\sigma_k^{\text{Mkt}}-\sigma_k^{\text{Mod}}}{\sigma_k^{\text{Mkt}}}\right)^2}$. The target function used for calibration writes as the square of the RMSE.

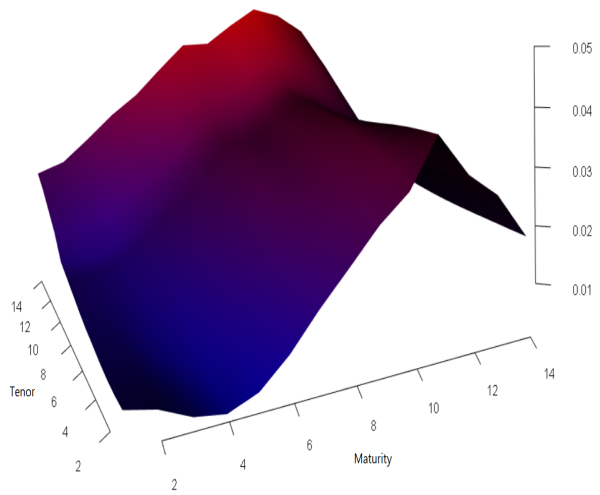
Table B.1, we specify the *freezing* error for some chosen volatilities: the simulated quantities along with bounds of confidence intervals are compared to the theoretical volatilities. Eventually those results demonstrate that the error induced by the *freezing* of some random quantities on prices/volatilities is quite weak.



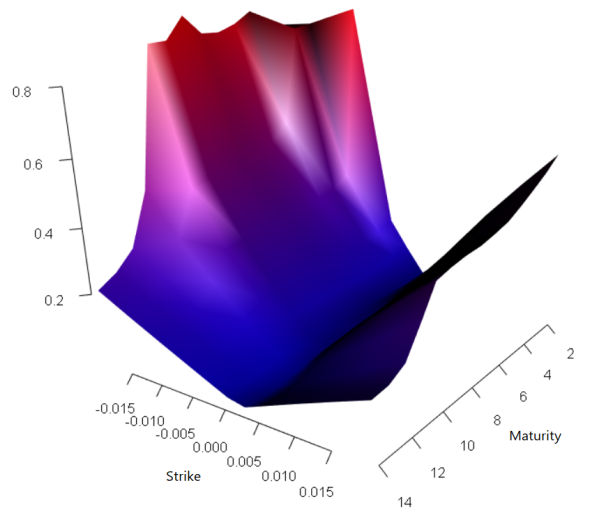
(a) Freezing error on ATM using Rebonato assumption.



(b) Freezing error on AFM using Rebonato assumption.



(c) Freezing error on ATM using Hull & White assumption.



(d) Freezing error on AFM using Hull & White assumption.

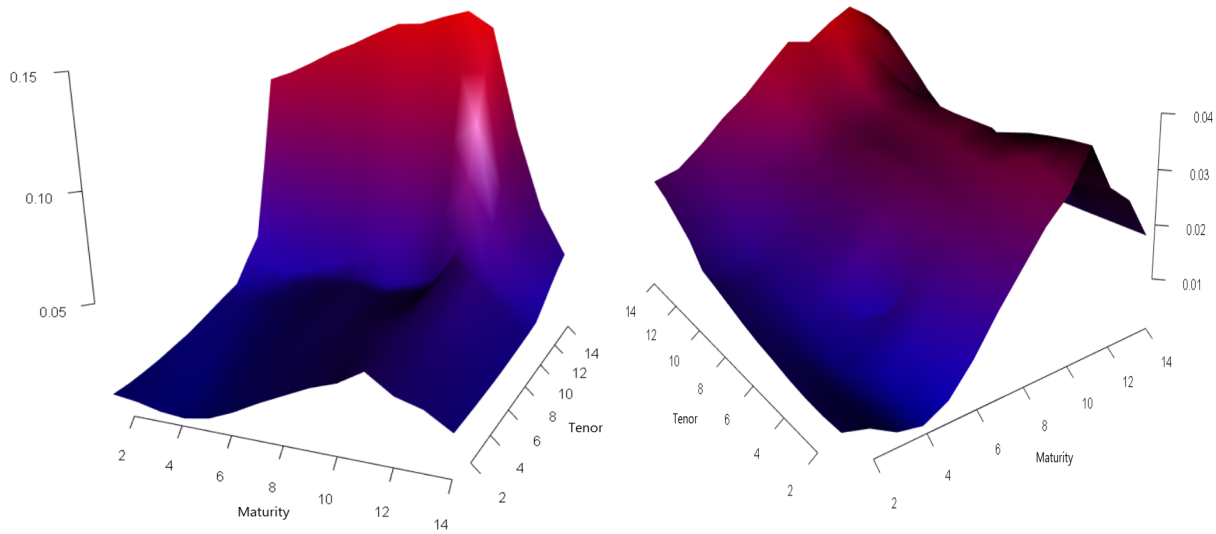
Figure 2.2: Relative *freezing* error for both approximation methods.

Comparison under a fixed set of parameters

Finally, we compare each approximation under a fixed set of parameters. We select a set of parameters obtained as outputs of a calibration procedure. The Hull & White approximation is again the most accurate one in particular on ATM data, as presented in Table 2.3. From results presented below in Table B.2 and Figures 2.3, we observe again that the difference between the two methods is material for long tenors. Those results demonstrate that the Hull & White method appears as the most accurate to approximate exact quantities.

Freezing method	RMSE-ATM	RMSE-AFM
Rebonato	0.05293	0.3211
Hull & White	0.025384636	0.3214

Table 2.3: *Freezing* error under a fixed set of parameters.



(a) Relative error between Rebonato and exact volatilities on ATM data.

(b) Relative error between Hull & White and exact volatilities on ATM data.

Figure 2.3: *Freezing* error under a fixed set of parameters.

2.2.4.2 Log-normal convention

The same study is led in the Black-Scholes framework. Log-normal framework allows to induce some strike dependency and reproduce smile in volatilities. The final conclusion is the same as the one for the Bachelier environment: the Hull & White approximation turns out to be a bit more accurate than the Rebonato one under multiple criterion.

Beforehand, we clarify a point that can seem counter-intuitive: it is quite possible to resort on Black pricing convention while using normal quoted volatilities. Indeed, such volatilities can be converted into Black ones following the scheme: (i) computation of market prices via Bachelier pricing formula using normal quoted volatilities; (ii) computation of log-normal volatilities by inverting the Black formula with market prices as inputs. Conversely, one can use a Bachelier pricing convention while calibrating its model on log-normal prices following a similar procedure. In practice, to circumvent this difficulty and to ensure the consistency the target function to optimize should involve prices. Note that, since market data we use here are Bachelier volatilities, we convert the model volatilities into normal ones for computing the RMSEs.

Calibration using *freezing* approximations

Calibration is performed similarly on the same dataset as previously. The displacement coefficient δ is set to 5% now. Table 2.4 gathers the replication error between theoretical volatilities and market ones. Again both approximation provide similar data replication. Yet contrary to the Bachelier environment, the Rebonato assumption here leads to a slightly better replication of the market data notably for ATM.

Freezing method	RMSE-ATM	RMSE-AFM
Rebonato	0.04822	0.1149
Hull & White	0.04979	0.1143

Table 2.4: Comparison of *freezing* techniques for calibration in Black environment.

Simulations of *exact* quantities

Simulations of log-normal volatilities rely on the Black formula. Table 2.5 gathers the distance of Monte-Carlo Black volatilities to market ones and to *frozen* ones, obtained for each set of parameters as output of the previously calibration process.

In Black environment, Hull & White convention leads to a significant reduction of distances on ATM volatilities; errors on AFM ones are slightly larger. The reduction of *freezing* error is remarkable when relying on this assumption. Regarding the calibration error, Hull & White also provides a slightly better ATM data replication. In Figure 2.4, we provide the relative *freezing* error induced by each method and on each type of data. On ATM data a strong dependency with respect to the tenor and maturity is still observed. Regarding the AFM data now, there is no longer such a strong dependency on the level of moneyness –since the conversion of log-normal towards normal volatilities induced some smile– but rather on the maturity.

Freezing method	Error between Monte-Carlo vol. and ..	ATM	AFM
Rebonato	Theoretical vol.	0.03821	0.05520
	Market vol.	0.05965	0.1308
Hull & White	Theoretical vol.	0.006913	0.07158
	Market vol.	0.05120	0.1368

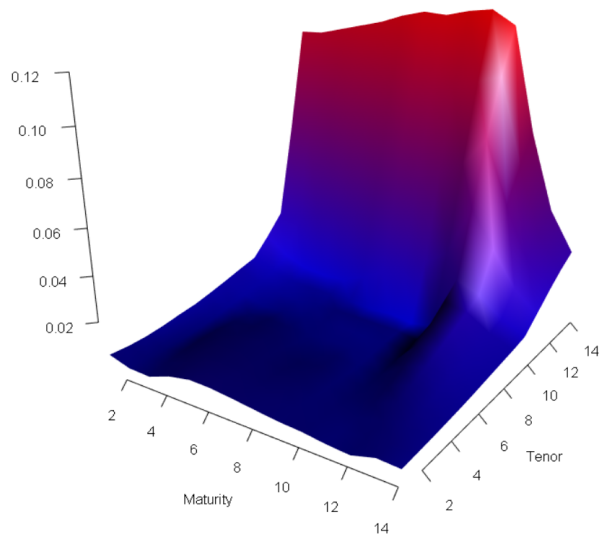
Table 2.5: Distance between exact volatilities and theoretical/market ones.

Comparison under a fixed set of parameters

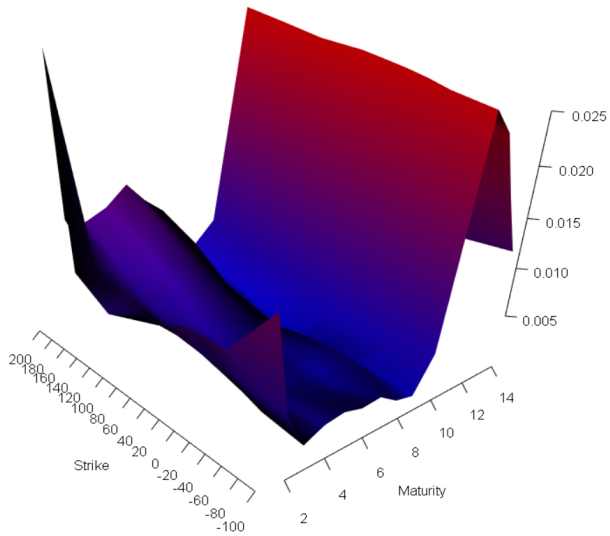
The accuracy of each approximating method on a given set of parameters is summarized in Table 2.6: Hull & White approximation yields a much lower *freezing* error on ATM data and a bit higher one on AFM volatilities. For ATM data (Fig. 2.5a and 2.5b), *freezing* error induced by Rebonato approach turns out to have a stronger dependency with respect to tenor/maturity.

Freezing method	RMSE-ATM	RMSE-AFM
Rebonato	0.03821	0.05520
Hull & White	0.006859	0.06669

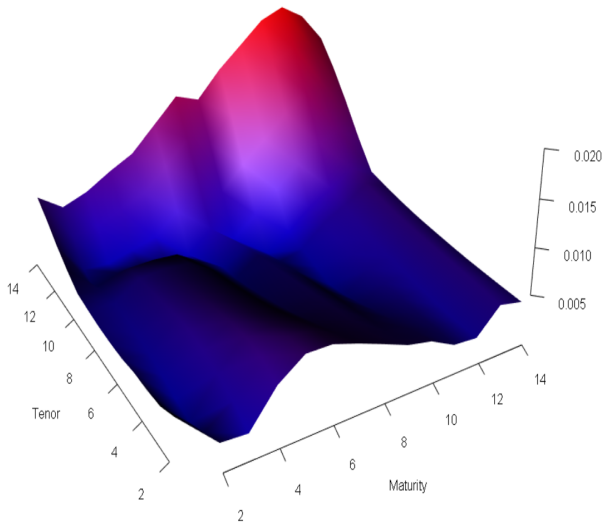
Table 2.6: *Freezing* error under a fixed set of parameters.



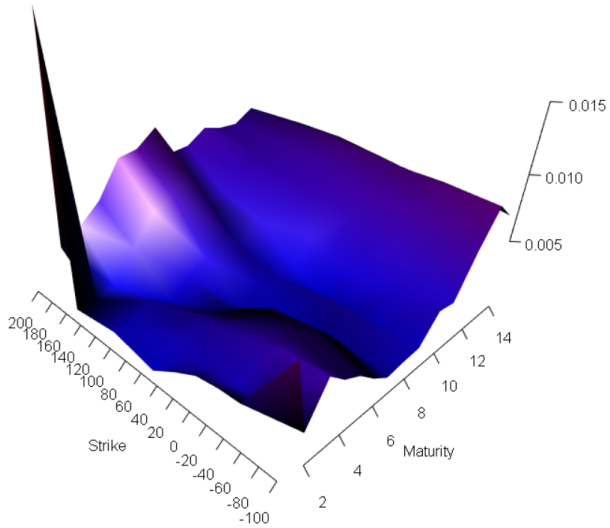
(a) Freezing error on ATM using Rebonato assumption.



(b) Freezing error on AFM using Rebonato assumption.

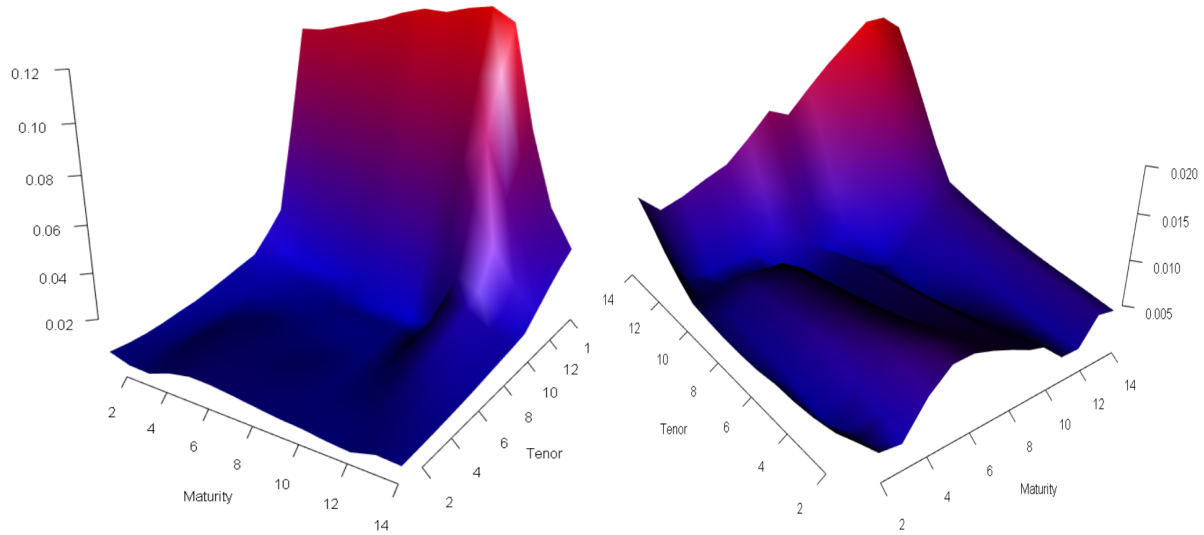


(c) Freezing error on ATM using Hull & White assumption.



(d) Freezing error on AFM using Hull & White assumption.

Figure 2.4: Relative *freezing* error for both approximation methods.



(a) Freezing error on ATM using Rebonato assumption.

(b) Freezing error on ATM using Hull & White assumption.

Figure 2.5: Relative *freezing* error for both approximation methods under a fixed set of parameters.

2.3 Proposed alternative

We first introduce the problem in a very general setting. Let us place ourselves in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\boldsymbol{X} := (\boldsymbol{X}_t)_{t\leq T} \in \mathbb{L}^2(\mathrm{d}t, \mathrm{d}\mathbb{P})$ be a given measurable stochastic process of dimension q whose initial value $\boldsymbol{X}_0 = \boldsymbol{x}_0$ is assumed to be known and such that $t \mapsto \mathbb{E}[\boldsymbol{X}_t]$ is piecewise continuous. Consider $\boldsymbol{Y} := (\boldsymbol{Y}_t)_{t\leq T}$ a stochastic process of dimension n characterized by

$$\mathrm{d}\boldsymbol{Y}_t = \boldsymbol{b}(\boldsymbol{Y}_t, \boldsymbol{X}_t)\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{Y}_t, \boldsymbol{X}_t)\mathrm{d}\boldsymbol{W}_t, \tag{2.11}$$

with $\boldsymbol{Y}_0 \overset{a.s.}{=} \boldsymbol{y}_0 \in \mathbb{R}^n$, $\boldsymbol{b} : \mathbb{R}^n \times \mathbb{R}^q \longrightarrow \mathbb{R}^n$ and $\boldsymbol{\sigma} : \mathbb{R}^n \times \mathbb{R}^q \longrightarrow \mathbb{R}^{n\times n}$ two measurable functions and $(\boldsymbol{W}_t)_{t\leq T}$ a n-dimensional Brownian motion. Computation of $p = \mathbb{E}[f(\boldsymbol{Y}_T)]$, where f is a Lipschitz function making this expectation well defined, is assumed to be not possible in closed form formula. p is then approximated by $\tilde{p} = \mathbb{E}[f(\boldsymbol{Y}_T^\nu)]$ where $\boldsymbol{Y}^\nu$ is an approximation process, characterized by

$$\mathrm{d}\boldsymbol{Y}_t^\nu = \boldsymbol{b}(\boldsymbol{Y}_t^\nu, \boldsymbol{\nu}_t)\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{Y}_t^\nu, \boldsymbol{\nu}_t)\mathrm{d}\boldsymbol{W}_t, \tag{2.12}$$

with $\boldsymbol{Y}_0^\star \overset{a.s.}{=} \boldsymbol{y}_0$ and $(\boldsymbol{\nu}_t)_{t\leq T}$ is an alternative process q-dimensional such that $\boldsymbol{\nu}_0 = \boldsymbol{X}_0$ making the computation of $\tilde{p}$ analytically doable. The problem can be formulated as follows:

$$\inf_{\boldsymbol{\nu}\in\mathcal{U}} (p - \tilde{p})^2 \tag{2.13}$$

where $\mathcal{U}$ is the set of admissible approximation process. Of course, if the set of admissible approximation process $\mathcal{U}$ is composed of all progressively measurable process, the problem (2.13) admits a trivial solution $\boldsymbol{\nu}_t^* = \boldsymbol{X}_t$. In the following, we will thus restrict ourselves to deterministic (piecewise) continuous functions:

$$\mathcal{U} = \Big\{\boldsymbol{g} : t \in [0, T] \longrightarrow \boldsymbol{g}_t \in \mathbb{R}^q \text{ deterministic: } \boldsymbol{g}_0 = \boldsymbol{x}_0, \int_0^T \|\boldsymbol{g}_s\|^2 \mathrm{d}s < \infty\Big\}.$$

A priori estimates
Let $\boldsymbol{\nu} \in \mathcal{U}$ be given. Since f is a Lipschitz, we have

$$(p - \tilde{p})^2 \leq C_{\mathrm{Lip(f)}}\mathbb{E}[\|\boldsymbol{Y}_T - \boldsymbol{Y}_T^\nu\|^2].$$

Roughly speaking, we expect that the closest the processes $\boldsymbol{X}$ and $\boldsymbol{\nu}$ are (in some sense), the better the approximation $\tilde{p} \approx p$ is. This intuition is illustrated below.

Theorem 2.1. *Assume that the functions $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ are Lipschitz. Then, there exists some positive constant $K > 0$ such that*

$$\mathbb{E}[\|\boldsymbol{Y}_T - \boldsymbol{Y}_T^\nu\|^2] \leq K \int_0^T \mathbb{E}[\|\boldsymbol{X}_s - \boldsymbol{\nu}_s\|^2]\mathrm{d}s. \tag{2.14}$$

Proof. From triangle and Jensen's inequality, we have

$$\|\boldsymbol{Y}_T - \boldsymbol{Y}_T^\nu\| \leq \int_0^T \|\boldsymbol{b}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{b}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\|\mathrm{d}s + \left\|\int_0^T \big(\boldsymbol{\sigma}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{\sigma}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\big)\mathrm{d}\boldsymbol{W}_s\right\|,$$

and with Jensen's inequality again

$$\|\boldsymbol{Y}_T - \boldsymbol{Y}_T^\nu\|^2 \leq 2T \int_0^T \|\boldsymbol{b}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{b}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\|^2\mathrm{d}s + 2\left\|\int_0^T \big(\boldsymbol{\sigma}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{\sigma}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\big)\mathrm{d}\boldsymbol{W}_s\right\|^2.$$

Taking the expectation, using the Fubini theorem, the Lipschitz property of the coefficients and the Burkholder-Davis-Gundy inequality, we get:

$$\begin{aligned}
\mathbb{E}[\|\boldsymbol{Y}_T - \boldsymbol{Y}_T^\nu\|^2] &\leq 2C^2_{\mathrm{Lip}(\boldsymbol{b})} T \int_0^T \mathbb{E}\big[\|(\boldsymbol{Y}_s, \boldsymbol{X}_s) - (\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\|^2\big] \mathrm{d}s \\
&\quad + 2\mathbb{E}\left[\sup_{t\leq T} \left\| \int_0^t \big(\boldsymbol{\sigma}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{\sigma}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\big) \mathrm{d}\boldsymbol{W}_s \right\|^2\right] \\
&\leq 2C^2_{\mathrm{Lip}(\boldsymbol{b})} T \int_0^T \mathbb{E}\big[\|(\boldsymbol{Y}_s, \boldsymbol{X}_s) - (\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\|^2\big] \mathrm{d}s \\
&\quad + 2C_{\mathrm{BDG}} \int_0^T \mathbb{E}\left[\|\boldsymbol{\sigma}(\boldsymbol{Y}_s, \boldsymbol{X}_s) - \boldsymbol{\sigma}(\boldsymbol{Y}_s^\nu, \boldsymbol{\nu}_s)\|^2\right] \mathrm{d}s \\
&\leq C\Big(\int_0^T \mathbb{E}\big[\|\boldsymbol{Y}_s - \boldsymbol{Y}_s^\nu\|^2\big] \mathrm{d}s + \int_0^T \mathbb{E}\big[\|\boldsymbol{X}_s - \boldsymbol{\nu}_s\|^2\big] \mathrm{d}s \Big)
\end{aligned}$$

where we set $C = 2\max\left(C^2_{\mathrm{Lip}(\boldsymbol{b})}T, C_{\mathrm{BDG}}C^2_{\mathrm{Lip}(\boldsymbol{\sigma})}\right)$ for the last inequality. The Gronwall's lemma gives the result. □

We can look in more detail the auxiliary problem

$$\inf_{\boldsymbol{\nu}\in\mathcal{U}} J(\boldsymbol{\nu}), \tag{2.15}$$

where we set the functional $J(\boldsymbol{\nu}) := \int_0^T \mathbb{E}[\|\boldsymbol{X}_s - \boldsymbol{\nu}_s\|^2]\mathrm{d}s$. We resort on the theory of variational calculus: the Lagrangian function $(t, \boldsymbol{\nu}) \in [0,T]\times\mathcal{U} \mapsto L(t,\boldsymbol{\nu}) := \mathbb{E}[\|\boldsymbol{X}_t - \boldsymbol{\nu}\|^2] = \mathbb{E}[\|\boldsymbol{X}_t\|^2] - 2\boldsymbol{\nu}\cdot \mathbb{E}[\boldsymbol{X}_t] + \|\boldsymbol{\nu}\|^2$ is continuous, partially differentiable and convex with respect to its second variable. According to Proposition 1.4.3 of [Kie18], any local minimizer of J is a global minimizer. Still according to [Kie18], Proposition 1.9.1, any local minimizer must satisfy the Euler-Lagrange equation that is

$$\frac{\partial}{\partial \nu} L(t,\boldsymbol{\nu}) = 0,\ t\in[0,T], \tag{2.16}$$

Consider $\boldsymbol{\nu} \in \mathcal{U}$ and a small perturbation of it $\delta\boldsymbol{\nu}$ such that $\boldsymbol{\nu} + \delta\boldsymbol{\nu} \in \mathcal{U}$. Straightforward computations show that

$$\forall t \in [0,T],\ L(t, \boldsymbol{\nu} + \delta\boldsymbol{\nu}) = L(t,\boldsymbol{\nu}) + 2\delta\boldsymbol{\nu}\cdot(\boldsymbol{\nu} - \mathbb{E}[\boldsymbol{X}_t]) + \|\delta\boldsymbol{\nu}\|^2.$$

Combined with (2.16) we obtain a minimizer $\boldsymbol{\nu}^*$ of J

$$\forall t \in [0,T], \boldsymbol{\nu}_t^* = \mathbb{E}[\boldsymbol{X}_t].$$

It is a rather intuitive result: approximation of the process $\boldsymbol{X}$ by its expectation is a relevant candidate for our approximation problem (2.13). This is not fully satisfying for our initial problem since

$$\inf_{\nu\in\mathcal{U}} (p - \tilde{p})^2 \leq \inf_{\nu\in\mathcal{U}} J(\boldsymbol{\nu}).$$

However it allows to propose an alternative to the common usages presented above.

Mean approximation on *forward* rates

We applied to proposed alternative in present rates context by replacing the forward rates by their covariation process to obtain closed-form formulas similar to (2.6) and (2.9). Note that

the present proposed alternative aims at reducing the *freezing* error; regarding other types of errors there is no reason *a priori* to believe they would be reduced.

2.3.1 Rebonato-like

We propose to approximate the Rebonato implied volatility as

$$\left(\sigma_{m,n}^{\text{Reb}}\right)^2 = \frac{1}{T_m} \sum_{k,j=m}^{n-1} \alpha_j(0)\alpha_k(0) \int_0^{T_m} \mathbb{E}\left[\left(F_k(t)+\delta\right)\left(F_j(t)+\delta\right)\right] \boldsymbol{\gamma_j}(t)\cdot\boldsymbol{\gamma_k}(t)\mathrm{d}t. \tag{2.17}$$

The expectation appearing in the integral above can be taken under any equivalent to the Risk-Neutral one probability measure since we rely on the covariation process to compute it. This process starting at 0 is uniquely determined up to indistinguishability. By definition, the process $\left(\left(F_k(t)+\delta\right)\left(F_j(t)+\delta\right) - \langle F_k(.)+\delta, F_j(.)+\delta\rangle_t\right)_{t\leq 0}$ is a (local) martingale and thus $\mathbb{E}\left[\left(F_k(t)+\delta\right)\left(F_j(t)+\delta\right)\right] = \mathbb{E}\left[\langle F_k(.)+\delta, F_j(.)+\delta\rangle_t\right] + \left(F_k(0)+\delta\right)\left(F_j(0)+\delta\right)$. With $\langle F_k(.)+\delta, F_j(.)+\delta\rangle_t = \int_0^t \left(F_k(s)+\delta\right)\left(F_j(s)+\delta\right)\boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s$, we use Fubini theorem and solve a first degree ordinary differential equation satisfied by $t \mapsto \mathbb{E}\left[\left(F_k(t)+\delta\right)\left(F_j(t)+\delta\right)\right]$ to get $\mathbb{E}\left[\left(F_k(t)+\delta\right)\left(F_j(t)+\delta\right)\right] = \left(F_k(0)+\delta\right)\left(F_j(0)+\delta\right)\exp\left(\int_0^t \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s\right)$. Together with (2.17) gives

$$\left(\sigma_{m,n}^{\text{Reb}}\right)^2 = \frac{1}{T_m} \sum_{k,j=m}^{n-1} \alpha_j(0)\alpha_k(0)\left(F_k(0)+\delta\right)\left(F_j(0)+\delta\right)\left(\exp\left(\int_0^{T_m} \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s\right)-1\right). \tag{2.18}$$

2.3.2 Hull & White-like

The formula (2.9) can be similarly adapted so that the approximated implied volatility writes

$$\left(\sigma_{m,n}^{\text{HW}}\right)^2 = \frac{1}{T_m} \sum_{k,j=m}^{n-1} \partial_j\psi(0)\partial_k\psi(0)\left(F_k(0)+\delta\right)\left(F_j(0)+\delta\right)\left(\exp\left(\int_0^{T_m} \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s\right)-1\right). \tag{2.19}$$

Note that when the integrated variance of the forward rates is small, a Taylor series expansion shows that the proposed versions of the alternative *freezing* coincides with the standard initial value *freezing*. Namely, when $\int_0^{T_m} \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s << 1$, $\exp\left(\int_0^{T_m} \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s\right) - 1 \approx \int_0^{T_m} \boldsymbol{\gamma_k}(s)\cdot\boldsymbol{\gamma_j}(s)\mathrm{d}s$.

Note also that the obtention of log-normal type formulas can be easily obtained by simply dividing the theoretical volatilities in (2.18) and (2.19) by $S_0^{m,n}+\delta$ (still based on the low variability assumption of the ratios $\alpha_k(t)/\left(S_t^{m,n}+\delta\right)$ or $\partial_k\psi(t)/\left(S_t^{m,n}+\delta\right)$).

2.3.3 Numerical results

We simply provide results relative to Bachelier pricing convention here.

Calibration using *freezing* proposal

In Table 2.7 we give the distance between theoretical volatilities and market ones. The distinction between the two approaches is not really significant. Note that the proposed approach somewhat degrades the calibration process when compared to the standard initial value freezing method.

Freezing method	RMSE-ATM	RMSE-AFM
Rebonato	0.08536	0.1273
Hull & White	0.08361	0.1273

Table 2.7: Proposed *freezing* method for calibration in Bachelier environment.

Simulation of *exact* quantities

Table 2.8 gathers *freezing* and calibration errors induced by our proposal. As previously, the Hull & White approximation turns out to be a bit more accurate on ATM data and a bit less on AFM ones. Contrary to what could have been expected, the proposal is not satisfying as the induced *freezing* error is substantially greater with the proposed freezing than the one displayed in Table 2.2 regarding ATM data. On AFM ones, the proposal is slightly more accurate.

Freezing method	Error between M.C. vol. and ..	RMSE-ATM	RMSE-AFM
Rebonato	Theoretical vol.	0.1136	0.3121
	Market vol.	0.1043	0.2963
Hull & White	Theoretical vol.	0.09738	0.3142
	Market vol.	0.09380	0.2981

Table 2.8: Distance between exact volatilities and theoretical/market ones.

Appendix B provides some detailed results on the comparison between both methods from the *freezing* error point of view.

Chapter 3

On the Jacobi process and polynomial expansions

This chapter comprises a survey on the Jacobi process and on polynomial expansions techniques for probability density functions. The first is a bounded stochastic process with the mean reversion property and having a number of interesting properties. Second one is particularly adapted to distributions satisfying the moment problem, such as that of the Jacobi process. These two concepts will be at the core of Chapter 4.

Notations The following notations are valid until the end of present thesis.

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying usual conditions. In a financial context, $\mathbb{P}$ can be viewed as the historical probability measure whereas the filtration will represent market information (quoted prices, observed smile, etc.). The latter is assumed to be generated by a multivariate Brownian motion. For two (local) martingales $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$, $\langle X., Y. \rangle_t$ will denote their quadratic covariation at time t. $Z \overset{d}{=} Z'$ means that distributions of Z and Z' are in fact the same whereas $Z \overset{a.s.}{=} Z'$ stands for almost sure equality. $\sigma(Z)$ is the sigma-algebra generated by the random variable Z. $\mathbb{E}[\cdot]$ is the expectation associated to $\mathbb{P}$, $\mathbb{E}_x[X_T]$ is the conditional expectation of X_T given the starting point $X_0 = x$. We will denote with bold font $\boldsymbol{u}$ the vectors; the canonical scalar product between two vectors will be denoted $\boldsymbol{u} \cdot \boldsymbol{v}$. Unless otherwise stated, $\|\boldsymbol{u}\|$ will represent the ($\mathcal{L}^2$-) norm induced by the scalar product i.e. $\|\boldsymbol{u}\| = \sqrt{\boldsymbol{u} \cdot \boldsymbol{u}}$.
For $(a, b) \in \mathbb{R}^2$, $a < b$, $[a, b]$ (resp. (a, b)) stands for the set $\{x \in \mathbb{R} : a \leq x \leq b\}$ (resp. $\{x \in \mathbb{R} : a < x < b\}$). $(a, b]$ and $[a, b)$ are defined similarly.

Considering E a subset of $\mathbb{R}^d$, $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ and a smooth function f defined over E, we will denote by $\partial_\alpha f$ the following differentiation operator

$$x = (x_1 \ldots, x_d) \in E \mapsto \partial_\alpha f(x) = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_1} f(x).$$

We also introduce the following functional space

$$\mathcal{C}_{pol}^{\infty}(E) = \left\{ f : E \to \mathbb{R} : f \in \mathcal{C}^{\infty}, \forall \alpha \in \mathbb{N}^d, \exists C_\alpha > 0, e_\alpha \in \mathbb{N}^*, \forall x \in E, |\partial_\alpha f(x)| \leq C_\alpha (1 + \|x\|^{e_\alpha}) \right\}.$$

Some «special» functions will be needed in this work. We will denote the Gamma function by $z \in \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\} \mapsto \Gamma(z) = \int_0^\infty u^{z-1} e^{-u} \mathrm{d}u$. The beta function is then defined as $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Name	Parameters	Density
Normal or Gaussian $\mathcal{N}(\mu, \sigma^2)$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Gamma $\Gamma(k, \theta)$	$(k, \theta) \in (\mathbb{R}_+^*)^2$	$\frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-\theta x} \mathbb{1}_{x>0}$
Beta $\beta(a, b)$	$(a, b) \in (\mathbb{R}_+^*)^2$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \mathbb{1}_{0<x<1}$
Chi-squared with d degree $\chi^2(d)$	$d \in \mathbb{N}^*$	$\frac{x^{d/2-1} e^{-x/2}}{2^{d/2}\Gamma(d/2)} \mathbb{1}_{x>0}$
Exponential $\mathcal{E}(\lambda)$	$\lambda > 0$	$\lambda e^{-\lambda x} \mathbb{1}_{\{x>0\}}$

Table 3.1: Common continuous probability distributions.

3.1 Jacobi process: an overview

3.1.1 Introduction

Stochastic processes taking values in finite domain have been originally introduced by biologists to model gene frequencies. To give some insights on their motivations, let us consider a population in which two types of alleles $\{A, B\}$ of a given gene exist. Sewall Wright and Ronald Fisher first studied a discrete time evolution model to describe the spread of alleles among the population (the interested reader could refer to [Wri31] and references therein) in which the number of individuals owning a given allele is distributed following a binomial distribution. Namely, denote by X_n the number of individuals in the population of size N with allele type A at time n. The sequence $(X_n)_{n\in\mathbb{N}}$ is modelled by a Markov chain with transition probabilities given by

$$\mathbb{P}(X_{n+1} = k | X_n = i) = \binom{N}{k} \left(\frac{i}{N}\right)^k \left(1 - \frac{i}{N}\right)^{N-k}.$$

for $(k, i) \in \{1, 2, \ldots, N\}^2$. It turns out that this discrete time model can be extended to a continuous diffusion setup. The limit process is employed as a continuous in time gene frequencies model. It can be shown (see [KT81]) that the limit process satisfy a stochastic differential equation of the form

$$\mathrm{d}X_t = (a - kX_t)\mathrm{d}t + \sigma\sqrt{X_t(1 - X_t)}\mathrm{d}W_t,$$

where W is a Brownian motion and non negative parameters a, k allows to model mutations and selection in the alleles evolutions. The process $(X_t)_{t\geq 0}$ defined this way is usually referred to as the Wright-Fisher process. When $0 \leq a \leq k$ and $\sigma \in \mathbb{R}$, it can be shown that this process is well defined and remains in the interval $[0, 1]$ almost surely (see Theorem 3.1 below) allowing then to interpret X as a frequency. Translation of the defined process $(X_t)_{t\geq 0}$ allows to consider more general diffusion processes living in any bounded interval. In particular, the process taking values in $[-1, 1]$ has been named *Jacobi process* due to its connection with Jacobi polynomials (named after Carl Gustav Jacob Jacobi; see dedicated Section 3.1.3 below for details on the mentioned connection). Thereafter, Jacobi process will more generally refer to any $[x_*, x^\star]$-valued (with $-\infty \leq x_* < x^\star \leq +\infty$) process with Wright-Fisher type dynamics.

In finance modelling, Jacobi process has also been employed with multiple objectives. Modelling of stochastic correlation can be done using Jacobi process lying in $[-1, 1]$. In [Ma09], the author worked with one equity dynamics for domestic economy and one for foreign economy; those assets are correlated with a stochastic correlation factor which dynamics is given by a $[-1, 1]$-valued Jacobi process. Semi analytical option prices were obtained in this framework. In a similar –but not equivalent– modelling context, [VV12] used a Jacobi process to model stochastic leverage. In an interest rates context, [DS02] proposed the Jacobi dynamics for mod-

elling the evolution of the short rate; they derived semi analytical formula for bond prices that is barely tractable. In his thesis, [Kap10] proposed a survey and an extension of the work of [DS02] –along with a discrete time interest rate modelling based on the Ehrenfest model (see for instance [KM65] on the topic). [DJDW01] and more recently [LS07], used a Jacobi dynamics for modelling the logarithm of the exchange rate between currencies in a specific zone. Even more recently, [AFP17] employed to Jacobi dynamics in a stochastic volatility type model to represent the evolution of the volatility factor.

Based on the last mentioned article, the Jacobi process will be central in Chapter 4 of present work. In this section, we propose to give a brief overview about the Jacobi process summarizing its main theoretical properties and some of the related extensions that can be found in the literature on the topic.

3.1.2 Definition and existence

Let us consider two (extended) real numbers $x_*, x^\star$ satisfying $-\infty \leq x_* < x^\star \leq +\infty$ and define the bounding function

$$Q(x) = (x^\star - x)(x - x_*). \tag{3.1}$$

Observe that $\frac{(x^\star - x_*)^2}{4} \geq Q(x) \geq 0$ for all x lying in $[x_*, x^\star]$. Consider then the following stochastic differential equation

$$X_t^{x_0} = x_0 + \int_0^t (a - kX_s^{x_0})\,\mathrm{d}s + \epsilon \int_0^t \sqrt{Q(X_s^{x_0})}\mathrm{d}B_s,\ t \geq 0, \tag{3.2}$$

where $x_0 \in [x_*, x^\star]$, $(a, k) \in \mathbb{R}^2$ that are such that $kx_* \leq a \leq kx^\star$, $\epsilon \in \mathbb{R}_+$and $(B_t)_{t\geq 0}$ is a one-dimensional Brownian motion.

Theorem 3.1. *There exists a unique strong solution satisfying (3.2).*

In our terminology, the «$[x_*, x^\star]$-valued» Jacobi process refers to this unique solution of (3.2).

Proof. Let us fix $x_0 \in [x_*, x^\star]$ and define the following functions:

$$f : x \in \mathbb{R} \mapsto a - k\mathrm{tr}(x),\ g : x \in \mathbb{R} \mapsto \epsilon\sqrt{Q(\mathrm{tr}(x))},\ \text{and } \mathrm{tr} : x \in \mathbb{R} \mapsto \min(\max(x, x_*), x^\star). \tag{3.3}$$

Consider then the stochastic differential equation

$$\tilde{X}_t^{x_0} = x_0 + \int_0^t f(\tilde{X}_s^{x_0})\mathrm{d}s + \int_0^t g(\tilde{X}_s^{x_0})\mathrm{d}B_s. \tag{3.4}$$

Continuity and boundedness of maps f and g implies weak existence of a solution to the SDE (3.4) following, for instance, [KS91], Theorem 5.4.22, p. 323. As a polynomial function, Q is locally Lipschitz; it is in particular over $[x_*, x^\star]$. It is also straightforward to check that the truncating function $x \mapsto \mathrm{tr}(x)$ is Lipschitz. Using in addition the 1/2-Hölder regularity of the square-root function, we eventually get that for some constant $C > 0$, for all $(x, y) \in \mathbb{R}^2$,

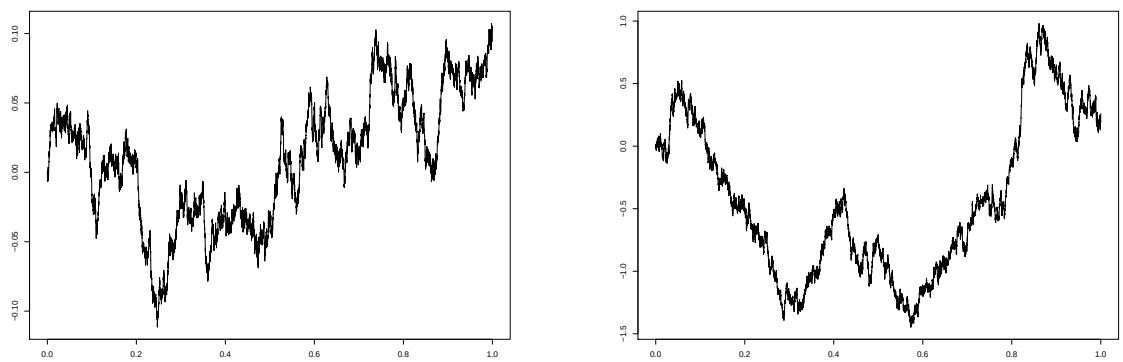
$$|g(x) - g(y)|^2 \leq \epsilon^2 |Q(\mathrm{tr}(x)) - Q(\mathrm{tr}(y))| \leq \epsilon^2 C|x - y|.$$

The pathwise uniqueness of solution to (3.4) follows thanks to Yamada-Watanabe theorem (see [YW+71], Theorem 1). Strong existence of solution to (3.4) is deduced from weak existence and pathwise uniqueness using Corollary 5.3.23 in [KS91] or Corollary 1 in [YW+71].

The remaining of the proof is dedicated to justification that solution $\tilde{X}^{x_0}$ of (3.4) remains in the interval $[x_*, x^\star]$ almost surely when $x_0 \in [x_*, x^\star]$. In such a case, the truncating function tr could be «removed» from the function f and g meaning that $\tilde{X}^{x_0}$ actually satisfies (3.2). We could proceed using Yamada functions to prove that the processes $X^{x_0}_{1,t} := \tilde{X}^{x_0}_t - x_*$ and $X^{x_0}_{2,t} := x^\star - \tilde{X}^{x_0}_t$ remains non-negative almost surely, as in Theorem 6.1.1 of [Alf15]. Alternatively, we propose here to use comparison theorems. The constant process $t \mapsto x_*$ satisfies (3.4) with $a = kx_*$ and $x_0 = x_*$; now, for a given $a \geq kx_*$, observe that $a - kx \geq k(x_* - x)$, $x \in \mathbb{R}$. Theorem 5.2.18 in [KS91] (p. 293) gives that $\mathbb{P}(\tilde{X}^{x_0}_t \geq x_*; \forall t \geq 0) = 1$. A similar reasoning is done for the upper bound, observing that $t \mapsto x^\star$ satisfies (3.4) with $x_0 = x^\star$ and $a = kx^\star$. Finally, $\mathbb{P}(x_* \leq \tilde{X}^{x_0}_t \leq x^\star; \forall t \geq 0) = 1$ and then $\tilde{X}^{x_0}$ satisfies (3.4) without truncating function in f and g: that is, $\tilde{X}^{x_0}$ satisfies (3.2). □

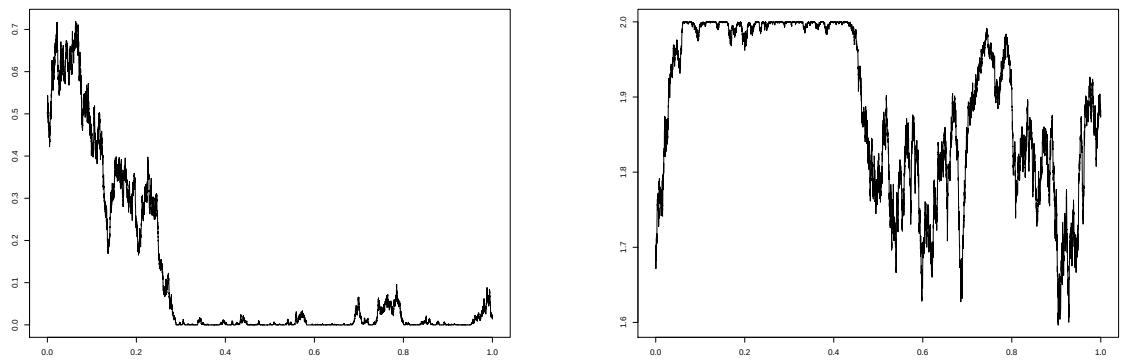
Remark 8. *In [AFP17], the authors get weak existence of $[x_*, x^\star]$-valued solution to (3.2) using an affine transformation of the unit interval $[0, 1]$ and weak existence of $[0, 1]$-valued solution as result of Theorem 2.1, [LP17]. Apart from that, our justification of strong existence and uniqueness is similar to theirs.*

Remark 9. *It is worth mentioning the particular cases when $a = kx^\star$ and $x_0 = x^\star$ or when $x_* = x^\star$ that make the process time independent: $\forall t \geq 0$, $X^{x_0}_t \overset{a.s.}{=} x_0$.*



(a) $x^\star = -x_* = 0.5$, $x_0 = 0.0$, $a = 0.2$, $k = 0.7$ and $\epsilon = 0.4$.

(b) $x^\star = 5$, $x_* = -3$, $x_0 = 0.0$, $a = 2$, $k = 1.5$ and $\epsilon = 0.4$.



(c) $x^\star = 2$, $x_* = 0$, $x_0 = 0.5$, $a = 0.9$, $k = 0.9$ and $\epsilon = 1.2$.

(d) $x^\star = 2$, $x_* = 0$, $x_0 = 1.7$, $a = 1.4$, $k = 0.9$ and $\epsilon = 1.2$.

Figure 3.1: Illustration of Jacobi paths for various parametrizations.

3.1.3 Distributional properties

Density function Before the development of stochastic integration with respect to Brownian motion by Kiyoshi Itô, diffusions were essentially studied through their transition semi-groups. For common diffusion process it amounts to study their infinitesimal generator. In this regard, it is particularly interesting to know the spectral composition of these functional operators as it is the case of infinitesimal generator associated to Jacobi diffusion. Indeed, the spectral representation is particularly suited for compact valued process as described in Section V.8 of [BW09] (p. 408 and p. 504). [Ma09] and [Alf15] applied this methodology to $[-1,1]$-valued Jacobi process while [DS02] focused on $[0,1]$-valued process to derive the transition density function. We propose here to follow the same approach for a generic domain $[x_*, x^\star]$.

We will see that the marginal density functions of the Jacobi process express in term of (extended) Jacobi polynomials. Jacobi polynomials polynomial functions are orthogonal in the space of square integral functions over $[-1,1]$ with respect to a proper weighting function. Composed with translations, they define *extended* Jacobi polynomials that can alternatively be referred to as *modified* Jacobi polynomials following the terminology of [Kap10].

In the following, we will denote by z the translation defined by $z : x \in [x_*, x^\star] \mapsto \frac{2}{(x^\star - x_*)}(x - x_*) - 1 \in [-1,1]$. The infinitesimal generator of diffusion (3.2) $\mathcal{A}$ applied to a twice differentiable function $f \in \mathcal{C}^2([x_*, x^\star])$ writes

$$\mathcal{A}f(x) = (a - kx)f'(x) + \frac{\epsilon^2}{2}(x - x_*)(x^\star - x)f''(x). \tag{3.5}$$

Let us denote by $(P_n^{(a,b)})_{n\in\mathbb{N}}$ the Jacobi polynomials parametrized by $(a,b) \in \mathbb{R}^2$. They can be defined as the orthonormal basis of the Hilbert space $\mathcal{L}^2_w = \left\{ f : \mathbb{R} \to \mathbb{R} \text{ measurable} : \int_{\mathbb{R}} f(x)^2 w(x)\mathrm{d}x < \infty \right\}$ with $w(x) = (1-x)^a(1+x)^b \mathbb{1}_{\{-1\le x\le 1\}}$ equipped with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{L}^2_w}$ (see below Section 3.4 and Equation 3.53). The interested reader can refer to [AS64] for a survey of the properties of the these polynomials.

Definition 1. *The n-th modified Jacobi polynomials is defined as*

$$Q_n^{(\tilde{a},\tilde{b})} : x \in [x_*, x^\star] \mapsto P_n^{(\tilde{a},\tilde{b})}(z(x)).$$

Basic properties of the modified Jacobi polynomials derive directly from that of standard ones.

Lemma 3.2. *Modified Jacobi polynomials $\left(Q_n^{(\tilde{a},\tilde{b})}\right)_{n\in\mathbb{N}}$ form an orthogonal basis of $\mathcal{L}^2_\pi$ with $\pi(x) = \left(\frac{2}{x^\star - x_*}\right)^{\tilde{a}+\tilde{b}} (x^\star - x)^{\tilde{a}}(x - x_*)^{\tilde{b}} \mathbb{1}_{\{x_*\le x\le x^\star\}}$ with $\tilde{b} = \frac{2(a-kx_*)}{\epsilon^2(x^\star - x_*)} - 1$ and $\tilde{a} = \frac{2(kx^\star - a)}{\epsilon^2(x^\star - x_*)} - 1$. Moreover, for each $n \in \mathbb{N}$, $Q_n^{(\tilde{a},\tilde{b})}$ is an eigenvector of the operator $\mathcal{A}$ associated with eigenvalue $\lambda_n = -kn - n(n-1)\frac{\epsilon^2(x^\star - x_*)}{4}$:*

$$\mathcal{A}Q_n^{(\tilde{a},\tilde{b})} = \lambda_n Q_n^{(\tilde{a},\tilde{b})}.$$

Proof. Let us first prove the second claim, it is straightforward to get

$$\frac{\mathrm{d}Q_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}x}(x) = \frac{2}{x^\star - x_*}\frac{\mathrm{d}P_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}z}(z) \text{ and } \frac{\mathrm{d}^2 Q_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}x^2}(x) = \left(\frac{2}{x^\star - x_*}\right)^2 \frac{\mathrm{d}^2 P_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}z^2}(z)$$

with $z = z(x) \in [-1, 1]$. We thus obtain

$$\mathcal{A}Q_n^{(\tilde{a},\tilde{b})}(x) = \frac{\epsilon^2}{2}\left\{(1-z^2)\frac{\mathrm{d}^2 P_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}z^2}(z) + \left[\frac{4a - 2k(x^\star + x_*)}{\epsilon^2(x^\star - x_*)} - \frac{2k}{\epsilon^2}z\right]\frac{\mathrm{d}P_n^{(\tilde{a},\tilde{b})}}{\mathrm{d}z}(z)\right\}.$$

The result is a consequence of the differential equation satisfied by the standard Jacobi polynomials $P_n^{a,b}$ is $(1-x^2)f'' + (b - a - (a+b+2)x)f' + n(n+a+b+1)f = 0$ (see [AS64]).

Now, observe that the family $\left(Q_n^{(\tilde{a},\tilde{b})}\right)_{n\in\mathbb{N}}$ forms an orthogonal system. Indeed, since $\lambda_n \neq \lambda_m$ for $n \neq m$, we have by definition of the scalar product associated to π:

$$\begin{aligned}\left\langle Q_n^{(\tilde{a},\tilde{b})}, Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi &= \int_{x_*}^{x^\star} Q_n^{(\tilde{a},\tilde{b})}(x)Q_m^{(\tilde{a},\tilde{b})}(x)\pi(x)\mathrm{d}x \\ &= \frac{1}{\lambda_n}\int_{x_*}^{x^\star} \mathcal{A}Q_n^{(\tilde{a},\tilde{b})}(x)Q_m^{(\tilde{a},\tilde{b})}(x)\pi(x)\mathrm{d}x = \frac{1}{\lambda_n}\left\langle \mathcal{A}Q_n^{(\tilde{a},\tilde{b})}, Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi.\end{aligned}$$

Similarly, we get $\left\langle Q_n^{(\tilde{a},\tilde{b})}, Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi = \frac{1}{\lambda_m}\left\langle Q_n^{(\tilde{a},\tilde{b})}, \mathcal{A}Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi$. However, from [BW09] (section V.8), we know that $\left\langle Q_n^{(\tilde{a},\tilde{b})}, \mathcal{A}Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi = \left\langle \mathcal{A}Q_n^{(\tilde{a},\tilde{b})}, Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi$. We deduce then that $\left\langle Q_n^{(\tilde{a},\tilde{b})}, Q_m^{(\tilde{a},\tilde{b})}\right\rangle_\pi = 0$. Moreover, using again the properties of standard Jacobi polynomials, we have that

$$\begin{aligned}&\int_{x_*}^{x^\star} Q_n^{(\tilde{a},\tilde{b})}(x)Q_m^{(\tilde{a},\tilde{b})}(x)(x^\star - x)^{\tilde{a}}(x - x_*)^{\tilde{b}}\mathrm{d}x \\ &= \left(\frac{x^\star - x_*}{2}\right)^{\tilde{a}+\tilde{b}+1}\int_{-1}^{1} P_n^{(\tilde{a},\tilde{b})}(u)P_m^{(\tilde{a},\tilde{b})}(u)(1-u)^{\tilde{a}}(1+u)^{\tilde{b}}\mathrm{d}u \\ &= \frac{(x^\star - x_*)^{\tilde{a}+\tilde{b}+1}}{2n+\tilde{a}+\tilde{b}+1}\frac{\Gamma(n+\tilde{a}+1)\Gamma(n+\tilde{b}+1)}{n!\Gamma(n+\tilde{a}+\tilde{b}+1)}\mathbb{1}_{m=n},\end{aligned}$$

whence the result.

□

Those properties are useful to derive the transition density for the process $(X_t^{x_0})_{t\geq 0}$ using a spectral decomposition of it.

Proposition 3.3. *Retake the parameters $\tilde{a}$, $\tilde{b}$, the eigenvalues $(\lambda_n)_{n\in\mathbb{N}}$ and the weighting function π introduced in the previous Lemma (3.2). The transition probability density of the process defined in (3.2) is given by*

$$p_{X^{x_0}}(t; x_0, y) = \pi(y)\sum_{n=0}^{\infty}\frac{1}{c_n}Q_n^{(\tilde{a},\tilde{b})}(x_0)Q_n^{(\tilde{a},\tilde{b})}(y)e^{\lambda_n t},\ t > 0, y \in [x_*, x^\star], \tag{3.6}$$

where the scaling coefficients $(c_n)_{n\in\mathbb{N}}$ are given by

$$c_n = \frac{1}{n!}\frac{2^{\tilde{a}+\tilde{b}}(x^\star - x_*)}{2n+\tilde{a}+\tilde{b}+1}\frac{\Gamma(n+\tilde{a}+1)\Gamma(n+\tilde{b}+1)}{\Gamma(n+\tilde{a}+\tilde{b}+1)}.$$

In particular, the stationary distribution of X^{x_0} obtained when sending $t \to \infty$ in (3.6) is the

non-standard beta[1] *distribution over* $[x_*, x^\star]$ *with parameters* $\tilde{a}+1$ *and* $\tilde{b}+1$*:*

$$p_\infty(y) = \frac{\pi(y)}{c_0} = \frac{1}{(x^\star - x_*)^{\tilde{a}+\tilde{b}+1}} \frac{(x^\star - x)^{\tilde{a}}(x - x_*)^{\tilde{b}}}{B(\tilde{a}+1, \tilde{b}+1)}$$

with B *the Beta function.*

Proof. According to [BW09] (see theoretical complements of section V.8 starting on p.504), the density of transition probability can be expanded using the orthonormal family $\left(Q_n^{(\tilde{a},\tilde{b})}\right)_{n\in\mathbb{N}}$ as

$$p_{X^x}(t; x, y) = \pi(y) \sum_{n=0}^{\infty} \frac{1}{c_n} Q_n^{(\tilde{a},\tilde{b})}(x) Q_n^{(\tilde{a},\tilde{b})}(y) e^{\lambda_n t}, \ t > 0, x, y \in]x_*, x^\star[. \tag{3.7}$$

Stationary density is simply deduced from the obtained one by sending $t \to \infty$. For any $n > 0$, $\frac{1}{c_n} Q_n^{(\tilde{a},\tilde{b})}(x_0) Q_n^{(\tilde{a},\tilde{b})}(y) e^{\lambda_n t} \xrightarrow[t\to+\infty]{} 0$; furthermore, the convergence of the series in (3.7) is uniform in t over any interval of the form $(a, +\infty)$ for any $a > 0$. We thus get that $\lim_{t\to\infty} p_{X^{x_0}}(t; x_0, y) = \frac{1}{c_0}\pi(y)$ which gives the result. □

Hitting properties In Figures 3.1c and 3.1d, we observe that the simulated Jacobi processes attain its lower, respectively upper, boundaries. This behaviour depends on the values of parameters defining (3.2). We take a closer look at this property here.

To prove the existence of the process X^{x_0} defined as solution of the diffusion (3.2), we used that $\mathbb{P}(\forall t \geq 0; X_t^{x_0} \in [x_*, x^\star]) = 1$. A common question that arises for modelling purposes, especially for financial applications, is to determine whether the process will attain the boundary $\{x_*, x^\star\}$ or not. Feller's tests for explosions allows to handle this question, and we refer to the Section 5.5.C in [KS91] for more details.

We introduce the stopping times corresponding to the first moment when the process reaches boundaries of its living domain. Denote by $\tau_{x_*} = \inf\{t \geq 0 : X_t^{x_0} = x_*\}$, $\tau_{x^\star} = \inf\{t \geq 0 : X_t^{x_0} = x^\star\}$ and by $\tau_{ex} = \tau_{x_*} \wedge \tau_{x^\star}$ the first exit time of the process X^{x_0}.

Proposition 3.4. *Assume that* $x_0 \in]x_*, x^\star[$ *and*

$$\epsilon^2(x^\star - x_*) \leq 2\min(a - kx_*, kx^\star - a) \tag{3.8}$$

Then the process X^{x_0} *defined in (3.2) never reaches the boundaries of its domain:*

$$\mathbb{P}(\tau_{ex} = \infty) = 1.$$

The following proof is a direct application of the general results of [KS91]. It is often found in the literature to prove that the CIR process never reaches 0 under Feller's condition. Due to its shortness, we recall it in our framework.

Proof. Denote by $b(x) = a - kx$ and $\sigma(x) = \epsilon\sqrt{Q(x)}$ the drift and diffusion functions of the process (3.2). Observe first that $\forall x \in]x_*, x^\star[$, $\sigma(x)^2 > 0$; then that for any $x \in]x_*, x^\star[$ there is an $\eta > 0$ such that $\int_{x-\eta}^{x+\eta} \frac{1+|b(u)|}{\sigma^2(u)} \mathrm{d}u < \infty$. This second assertion comes from fact that $\sigma^2(x^*) = 0 \iff x^* \in \{x_*, x^\star\}$, thus for any $x \in]x_*, x^\star[$, we can find a $\eta > 0$ such that $[x-\eta, x+\eta] \subset]x_*, x^\star[$; the continuous function $u \mapsto \frac{1+|b(u)|}{\sigma^2(u)}$ is thus integrable over the compact $[x-\eta, x+\eta]$.

[1] The beta distribution over $[0, 1]$ in sometimes named the *standard* beta distribution in contrast with *non-standard* beta distribution defined over arbitrary interval.

Now, fix a $c \in \mathbb{R} \setminus \{x_*, x^\star\}$, $c \notin$ and define the scale function

$$s(x) = \int_c^x \exp\left(-2\int_c^y \frac{b(s)}{\sigma^2(s)}\mathrm{d}s\right)\mathrm{d}y =: \int_c^x f(y)\mathrm{d}y.$$

Straightforward computations show that, for any $x \in \mathbb{R}$, for a given $y \in [c, x]$,

$$f(y) = \exp\left(-2\int_c^y \frac{b(s)}{\sigma^2(s)}\mathrm{d}s\right) = \left(\frac{y - x_*}{c - x_*}\right)^{\frac{-2(a-kx_*)}{\epsilon^2(x^\star - x_*)}} \left(\frac{x^\star - y}{x^\star - c}\right)^{\frac{2(a-kx^\star)}{\epsilon^2(x^\star - x_*)}}.$$

Condition (3.8) writes $\epsilon^2(x^\star - x_*) \leq \min(a - kx_*, kx^\star - a)$ that is equivalent $\max\left(\frac{-2(a-kx_*)}{\epsilon^2(x^\star - x_*)}, \frac{2(a-kx^\star)}{\epsilon^2(x^\star - x_*)}\right) \leq -1$. Then the function f is neither integrable at x_*^+ nor at $x^{\star\,-}$:

$$s(x) \xrightarrow[x \nearrow x^\star]{} \infty \text{ and } s(x) \xrightarrow[x \searrow x_*]{} \infty.$$

The conclusion follows from Proposition 5.5.29, p.345, in [KS91]. □

In the following, we will refer to condition (3.8) as the Feller condition. When it is not satisfied, the hitting probability can be explicitly computed (still following Section 5.5.C of [KS91]). Indeed if $\mathbb{P}(T_{ex} < \infty) = 1$, the probabilities of hitting x_* and $x^\star$ are now given by:

$$\mathbb{P}\left(\lim_{t \nearrow \tau_{ex}} X_t^{x_0} = x^\star\right) = 1 - \mathbb{P}\left(\lim_{t \nearrow \tau_{ex}} X_t^{x_0} = x_*\right) = \frac{s(x^{\star\,-}) - x_0}{s(x^{\star\,-}) - s(x_*^+)}.$$

3.1.4 Study of the moments

Random variable characterized by their moments are of particular interest for the work led in this thesis; we have discussed about such variables in dedicated Section 3.2. We will see below that Jacobi process is always characterized by its moments regardless the choice of the boundaries $\{x_*, x^\star\}$. Positive and negative moments (i.e. moments with negative exponents) of the Jacobi process are studied below in the five following cases: (i) $-\infty < x_* < x^\star < \infty$, (ii) $0 = x_* < x^\star < \infty$, (iii) $-\infty < x_* < x^\star = +\infty$, (iv) $-\infty = x_* < x^\star < \infty$ and finally (v) $x_* = -\infty$ and $x^\star = +\infty$. Note that when considering infinite bounds, a rescaling is necessary. Cases (i) and (ii) will be treated using polynomial property while cases (iii), (iv) and (v) are exploited using their affine property to derive their characteristic function.

Remark 10. *Jacobi process usually refers to a bounded stochastic process (i.e.* $\max\left(|x_*|, |x^\star|\right) < \infty$*). Yet, unbounded processes can be seen as particular Jacobi process with infinite bounds. Two processes are well known in the literature: set* $(x_*, x^\star) = (0, +\infty)$ *yields the Cox-Ingersoll-Ross process and* $(x_*, x^\star) = (-\infty, +\infty)$ *defines the Ornstein-Uhlenbeck. In both cases, some rescaling should be done as we will see below.*

3.1.4.1 Moments computations

3.1.4.1.1 $-\infty < x_* < x^\star < +\infty$ As a bounded process, it is straightforward to get that all positive marginal moments of X^{x_0} are finite: for all $t \geq 0$,

$$X_t^{x_0} \in \bigcap_{p \in \mathbb{N}} \mathbb{L}^p.$$

It is well known that distributions with bounded supports are determined by their moments (see Section 3.2), and so is the Jacobi process. To compute its moments in this case, the theory

of polynomial processes elaborated in Section 3.1.9 can be employed. Indeed straightforward computations show that if $\mathcal{A}$ denotes the infinitesimal generator of diffusion (3.2), the set of polynomial functions of a given degree is invariant by $\mathcal{A}$: if $P(x) = \sum_{i=0}^{k} c_i x^i \in \mathcal{P}_k(\mathbb{R})$ for some $k \in \mathbb{N}$, then

$$\begin{aligned}\mathcal{A}P(x) = &-\sum_{i=0}^{k} i\Big(k + \frac{\epsilon^2}{2}(i-1)\Big)c_i x^i + \sum_{i=0}^{k} i\Big(a + \frac{\epsilon^2}{2}(i-1)(x_* + x^\star)\Big)c_i x^{i-1} \\ &- \frac{\epsilon^2 x_* x^\star}{2}\sum_{i=0}^{k} i(i-1)c_i x^{i-2} \in \mathcal{P}_k(\mathbb{R}),\end{aligned}$$

where $\mathcal{P}_k(\mathbb{R})$ is the set of polynomial function defined over $\mathbb{R}$ of degree $k \in \mathbb{N}$. Consequently the Jacobi process is polynomial. The detail of the computation of marginal expectation of X^{x_0} is given below as illustration. In the basis $\{1, x\}$ of $\mathcal{P}_1(\mathbb{R})$, the action of $\mathcal{A}$ can be represented by the $\mathbb{R}^{2\times 2}$ matrix

$$A^{(1)} = \begin{pmatrix} 0 & a \\ 0 & -k \end{pmatrix}.$$

The upper triangular form of $A^{(1)}$ eases the computation of its exponential. Indeed observe that

$$A^{(1)} = \begin{pmatrix} 1 & -a/k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & -k \end{pmatrix}\begin{pmatrix} 1 & a/k \\ 0 & 1 \end{pmatrix}$$

and that the left sided matrix is the inverse of the right sided one. Then for $t \geq 0$

$$\exp\left(tA^{(1)}\right) = \begin{pmatrix} 1 & \frac{a}{k}\left(1 - e^{-kt}\right) \\ 0 & e^{-kt} \end{pmatrix}$$

and thus using property of polynomial processes (see Equation 3.33 below)

$$\mathbb{E}\left[X_t^{x_0}\right] = (1, x_0)\cdot \exp\left(tA^{(1)}\right)(0,1)^T = x_0 e^{-kt} + \frac{a}{k}\left(1 - e^{-kt}\right).$$

Note that the expectation does not depend on x_* and $x^\star$ and is equal to the expectation of a CIR process or of an Ornstein-Uhlenbeck one.

In the particular sub-case when $x_* > 0$, we can study the negative moments of X^{x_0}: for any $\alpha \in \mathbb{R}_+$, at any time $t \geq 0$, $\frac{1}{(X_t^{x_0})^\alpha} \leq \frac{1}{x_*^\alpha}$ hence the finiteness of all negative moments of the Jacobi process. Their computation is more complicated. First we use a particular representation of the hypergeometric Euler function ${}_2F_1$ (see for instance [AS64] for more details) to write, for $s, t \geq 0$

$$\begin{aligned}&\int_{x_*}^{x^\star} (x^\star - u)^s (u - x_*)^t u^{-\alpha} \mathrm{d}u \\ &= \frac{(x^\star - x_*)^{s+t+1}}{x_*^\alpha}\int_0^1 x^t (1-x)^s \left(1 - \frac{x_* - x^\star}{x_*}x\right)^{-\alpha} \mathrm{d}x \\ &= \frac{(x^\star - x_*)^{s'-1}}{x_*^\alpha}\int_0^1 x^{t'-1}(1-x)^{s'-t'-1}\left(1 - \frac{x_* - x^\star}{x_*}x\right)^{-\alpha} \mathrm{d}x \\ &= \frac{(x^\star - x_*)^{s'-1}}{x_*^\alpha} B(t', s'-t')\,{}_2F_1\left(\alpha, t', s'; \frac{x_* - x^\star}{x_*}\right) \qquad (3.9)\end{aligned}$$

where we introduced $t' = t + 1$ and $s' = s + t' + 1$ and recall that B is the beta function. Note that this representation is valid for $s' > t' > 0$ and $\frac{x_* - x^\star}{x_*} < 1$. This second constraint is always

satisfy for $x^\star > 0$. We then use the explicit expression of the transition density obtained in (3.6). Let $t \geq 0$ and $\alpha \in \mathbb{R}_+$.

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{(X_t^{x_0})^\alpha}\right] &= \int_{x_*}^{x^\star} \frac{1}{y^\alpha}\pi(y)\sum_{n=0}^{\infty}\frac{1}{c_n}Q_n^{(\tilde{a},\tilde{b})}(x_0)Q_n^{(\tilde{a},\tilde{b})}(y)e^{\lambda_n t}\mathrm{d}y \\
&= \sum_{n=0}^{\infty}\frac{e^{\lambda_n t}}{c_n}Q_n^{(\tilde{a},\tilde{b})}(x_0)\int_{x_*}^{x^\star}\frac{1}{y^\alpha}\pi(y)Q_n^{(\tilde{a},\tilde{b})}(y)\mathrm{d}y \\
&= \sum_{n=0}^{\infty}\frac{2^{\tilde{a}+\tilde{b}}}{c_n}\frac{e^{\lambda_n t}Q_n^{(\tilde{a},\tilde{b})}(x_0)}{(x^\star - x_*)^{\tilde{a}+\tilde{b}+n}}\sum_{k=0}^{n}(-1)^k\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k} \\
&\quad \times \int_{x_*}^{x^\star}(x^\star - y)^{\tilde{a}+k}(y-x_*)^{\tilde{b}+n-k}y^{-\alpha}\mathrm{d}y \\
&= \sum_{n=0}^{\infty}\frac{2^{\tilde{a}+\tilde{b}}}{c_n}\frac{(x^\star - x_*)^2}{x_*^p}e^{\lambda_n t}Q_n^{(\tilde{a},\tilde{b})}(x_0)\sum_{k=0}^{n}(-1)^k\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k} \\
&\qquad \times \int_0^1 (1-y)^{\tilde{a}+k}y^{\tilde{b}+n-k}\left(1-zy\right)^{-\alpha}\mathrm{d}y
\end{aligned}$$

where we set $z = \frac{x_* - x^\star}{x_*}$ and the permutation between sum and integral has been made possible since for all $n \in \mathbb{N}$, $[x_*, x^\star] \ni y \mapsto \pi(y)Q_n^{(\tilde{a},\tilde{b})}(y)/y^\alpha$ can be uniformly bounded by $\frac{1}{x_*^\alpha}\|\pi Q_n^{(\tilde{a},\tilde{b})}\|_\infty$ so that the series $\sum_{n\geq 0} c_n^{-1}Q_n^{(\tilde{a},\tilde{b})}(x_0)\|\pi Q_n^{(\tilde{a},\tilde{b})}\|_\infty e^{\lambda_n t}x_*^{-p} < \infty$. Hence using (3.9), we get that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{(X_t^{x_0})^\alpha}\right] = 2^{\tilde{a}+\tilde{b}}\frac{(x^\star - x_*)^2}{x_*^\alpha}\sum_{n=0}^{\infty}\frac{e^{\lambda_n t}Q_n^{(\tilde{a},\tilde{b})}(x_0)}{c_n}\times \\
&\sum_{k=0}^{n}(-1)^k\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k}B(\tilde{b}+n-k+1,\tilde{a}+k+1)\,{}_2F_1\left(\alpha,\tilde{b}+n-k+1,\tilde{a}+\tilde{b}+n+2;z\right).
\end{aligned} \tag{3.10}$$

3.1.4.1.2 $0 = x_* < x^\star < \infty$ The study of positive moments remains as in the previous case. Computations coming from the theory of polynomial processes can be adapted by setting $x_* = 0$.

However, negative moments are not necessarily finite anymore. In this particular case,

$$\begin{aligned}
\pi(y) &= \left(\frac{2}{x^\star}\right)^{\tilde{a}+\tilde{b}} \times (x^\star - y)^{\tilde{a}}y^{\tilde{b}}, \\
Q_n^{(\tilde{a},\tilde{b})}(y) &= \frac{1}{(x^\star)^n}\sum_{k=0}^{n}\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k}(y-x^\star)^k y^{n-k}.
\end{aligned}$$

Hence the expression:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{(X_t^{x_0})^\alpha}\right] &= \int_0^{x^\star} \frac{1}{y^\alpha}\pi(y)\sum_{n=0}^{\infty}\frac{1}{c_n}Q_n^{(\tilde{a},\tilde{b})}(x_0)Q_n^{(\tilde{a},\tilde{b})}(y)e^{\lambda_n t}\mathrm{d}y \\
&= 2^{\tilde{a}+\tilde{b}}\sum_{n=0}^{\infty}\frac{e^{\lambda_n t}Q_n^{(\tilde{a},\tilde{b})}(x_0)}{c_n(x^\star)^{\tilde{a}+\tilde{b}+n}}\sum_{k=0}^{n}(-1)^k\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k}\int_0^{x^\star} y^{\tilde{b}+n-k-\alpha}(x^\star-y)^{\tilde{a}+k}\mathrm{d}y \\
&= \frac{2^{\tilde{a}+\tilde{b}}}{(x^\star)^{1-\alpha}}\sum_{n=0}^{\infty}\frac{e^{\lambda_n t}Q_n^{(\tilde{a},\tilde{b})}(x_0)}{c_n}\sum_{k=0}^{n}(-1)^k\binom{n+\tilde{a}}{n-k}\binom{n+\tilde{b}}{k}B(\tilde{b}-\alpha+1+n-k,\tilde{a}+k+1),
\end{aligned}$$

that is well defined for $\alpha < \tilde{b}+1$ that can equivalently be written as

$$\alpha\epsilon^2 x^\star < 2a.$$

For $\alpha = 1$, this last constraint is in particular implied by the Feller condition (3.8) when $x_* = 0$. This result will be proved in another way later in this thesis.

3.1.4.2 Characteristic function

As defined in (3.2), the Jacobi process naturally lives in the compact $[x_*, x^\star]$. By rescaling the diffusion coefficient, one can study the limiting cases $x_* = -\infty$ and $x^\star = +\infty$. In these particular cases, the theory of affine processes can be employed: beside being polynomial, the process $(X_t^{x_0})_{t\geq 0}$ is also affine.

Affine processes have been well studied in the literature, notably with applications to interest-rates and credit risk modelling. We refer the interested reader to [Alf06] for a thorough work on this topic. The characteristic function of X^{x_0} is explicitly known and the moments can be derived by differentiation.

3.1.4.2.1 $x_* \in \mathbb{R}$ **and** $x^\star = +\infty$ Let $\tilde{\epsilon} \geq 0$ and set $\epsilon = \frac{\tilde{\epsilon}}{\sqrt{x^\star}}$ in (3.2). Since $\tilde{\epsilon}\sqrt{\frac{Q(x)}{x^\star}} \xrightarrow[x^\star\to+\infty]{} \tilde{\epsilon}\sqrt{x-x_*}$, we consider in this paragraph the stochastic process defined by

$$X_t^{x_0} = x_0 + \int_0^t (a - kX_s^{x_0})\,\mathrm{d}s + \tilde{\epsilon}\int_0^t \sqrt{X_s^{x_0}-x_*}\mathrm{d}B_s, \tag{3.11}$$

where $(B_t)_{t\geq 0}$ is a Brownian motion. The infinitesimal generator associated to the process (3.11) writes

$$\mathcal{A}f(x) = (a-kx)f'(x) + \frac{\tilde{\epsilon}^2}{2}(x-x_*)f''(x)$$

when applied to a twice differentiable function.

Proposition 3.5. *Let $T > 0$. The characteristic function of $X_T^{x_0}$ is well-defined over the domain*

$$\mathcal{D} = \left\{z \in \mathbb{C} : Re(z) > -\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}\right\}$$

and is given by

$$\mathbb{E}\left[e^{-zX_T^{x_0}}\right] = \left(\frac{\tilde{\epsilon}^2 z}{2k}(1-e^{-kT})+1\right)^{-\frac{2a}{\tilde{\epsilon}^2}}\exp\left(\frac{ze^{-kT}}{1+\frac{\tilde{\epsilon}^2 z}{2k}(1-e^{-kT})}(x_*-x_0)\right). \tag{3.12}$$

Proof. We can adapt the proof of [Alf15], Proposition 1.2.4. Fix $z \in \mathbb{R}_+$, such that the function $F(t, X_T^{x_0}; z) = \mathbb{E}\left[e^{-zX_T^{x_0}} | \mathcal{F}_t\right]$ is well defined. First assume the existence of two smooth functions $A(t,z)$ and $B(t,z)$ such that F can be written as

$$F(t, X^{x_0}; z) \overset{a.s.}{=} e^{A(T-t,z)+B(T-t,z)X_t^{x_0}}. \tag{3.13}$$

The process $\left(F(t, X_T^{x_0}; z)\right)_{0 \le t \le T}$ is a martingale, and Itô's formula indicates that F is necessary solution of the partial differential equation (PDE):

$$\begin{cases} \frac{\partial F}{\partial t} + (a - kx)\frac{\partial F}{\partial x} + \frac{\tilde{\epsilon}^2}{2}(x - x_*)\frac{\partial^2 F}{\partial x^2} = 0, \\ F(T, x; z) = e^{-zx}. \end{cases}$$

It results in the following system of ordinary differential equations (ODEs) satisfied by functions A and B,

$$\begin{cases} \frac{\partial A}{\partial t} &= aB - \frac{\tilde{\epsilon}^2 x_*}{2}B^2, \\ \frac{\partial B}{\partial t} &= \frac{\tilde{\epsilon}^2}{2}B^2 - kB, \end{cases}$$

associated with initial conditions $A(0,z) = 0$ and $B(0,z) = -z$. This is a Riccati system of differential equations. $C = 1/B$ satisfies the ODE $\partial_t C - \kappa C + \frac{\tilde{\epsilon}^2}{2} = 0$ coupled with $C(0) = -1/z$ that is solved to get that $B(t,z) = \frac{-ze^{-kt}}{1+\frac{\tilde{\epsilon}^2 z}{2k}(1-e^{-kt})}$. Using this expression in the differential equation satisfied by A yields

$$A(t,z) = -\frac{2a}{\tilde{\epsilon}^2}\log\left(\frac{\tilde{\epsilon}^2 z}{2k}(1 - e^{-kt}) + 1\right) + \frac{zx_* e^{-kt}}{1 + \frac{\tilde{\epsilon}^2 z}{2k}(1 - e^{-kt})}.$$

Hence eventually the expression of the characteristic function

$$\begin{aligned} \mathbb{E}\left[e^{-zX_T^{x_0}}\right] &= F(0, x_0; z) \\ &= \left(\frac{\tilde{\epsilon}^2 z}{2k}(1 - e^{-kT}) + 1\right)^{-\frac{2a}{\tilde{\epsilon}^2}} \exp\left(\frac{ze^{-kT}x_*}{1 + \frac{\tilde{\epsilon}^2 z}{2k}(1 - e^{-kT})} - \frac{ze^{-kT}x_0}{1 + \frac{\tilde{\epsilon}^2 z}{2k}(1 - e^{-kT})}\right). \end{aligned}$$

To prove that the assumed form (3.13) allows to uniquely identified F, we use a Feynman-Kac type argument. Based on dynamics of Jacobi process (3.11) and thanks to Itô's formula applied to the process $F(t, X; z)$ provides $F(T, X^{x_0}; z) = F(0, X^{x_0}; z) + \tilde{\epsilon}\int_0^T B(T - s, z)F(s, X^{x_0}; z)\sqrt{X_s^{x_0} - x_*}\mathrm{d}B_s$ that is

$$e^{-zX_T^{x_0}} = e^{A(T,z)+B(T,z)x_0} + \tilde{\epsilon}\int_0^T B(T - s, z)F(s, X^{x_0}; z)\sqrt{X_s^{x_0} - x_*}\mathrm{d}B_s. \tag{3.14}$$

For $z \in \mathbb{R}_+$ and any $0 \le t \le T$, $B(T - t, z) \le 0$ and $A(T - t, z) \le \frac{2kx_*}{\tilde{\epsilon}^2(e^{k(T-t)}-1)}$ so that $F(s, X^{x_0}; z) \le e^{\frac{2kx_*}{\tilde{\epsilon}^2(e^{k(T-s)}-1)}}$ for all $0 \le s \le T$. Consequently, Itô's integral in (3.14) has zero expectation. Hence $\mathbb{E}\left[e^{-zX_T^{x_0}}\right] = e^{A(T,z)+B(T,z)x_0}$ which allows to identify F over $\mathbb{R}_+$.

It remains to determine the set of convergence of the characteristic function. The definition domain $\mathcal{D}$ contains the set of positive real numbers since $X_T^{x_0} \ge x_*$. The right-hand side of (3.12) is an analytic function for $z > -\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}$. Since characteristic functions are analytic in the interior of their domain of convergence (see for instance [Fil09], Lemma 10.8), both sides of

(3.12) must coincide over $(-\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}, \infty)$. Using the right-hand side expression in (3.12) and fact that $\frac{2a}{\tilde{\epsilon}^2} \geq 0$, we have using monotone convergence theorem that

$$\mathbb{E}\left[\exp\left(\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}X_T^{x_0}\right)\right] \to \infty$$

as $z \searrow -\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}$. Now for $z \in \mathbb{C}$, $\mathbb{E}\left[|e^{-zX_T^{x_0}}|\right] = \mathbb{E}\left[\exp\left(-\mathrm{Re}(z)X_T^{x_0}\right)\right] < \infty$ if and only if $\mathrm{Re}(z) > -\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}$. Both sides of (3.12) are analytic over $(-\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}, \infty[$, we deduce that the identity (3.12) extends to the whole domain

$$\left\{z \in \mathbb{C} : \mathrm{Re}(z) > -\frac{2k}{\tilde{\epsilon}^2(1-e^{-kT})}\right\}.$$

□

Remark 11. *Setting $x_* = 0$ in (3.11) corresponds to the Cox-Ingersoll-Ross process. It has been well studied in the literature, notably in [Alf15]: Proposition 1.2.4 (p.7) provides the expression of the characteristic function of the CIR process and its domain of definition. They can be recovered by setting $x_* = 0$ in Proposition 3.5.*

As discussed previously it is straight to check the determination of the distribution of X^{x_0} by its moments when $\max(|x_*|, |x^\star|) < \infty$. Yet, the property still hold in present case.

Proposition 3.6. *Set $x_* \in \mathbb{R}$ and $x^\star = +\infty$. The process X^{x_0} defined by (3.11) is determined by its moments.*

Proof. See proof of Lemma 4.1 of [FL16]. □

3.1.4.2.2 $x_* = -\infty$ **and** $x^\star \in \mathbb{R}$ Let $\tilde{\epsilon} \geq 0$ and set $\epsilon = \frac{\tilde{\epsilon}}{\sqrt{-x_*}}$ with $x_* \leq 0$ in (3.2). Since $\tilde{\epsilon}\sqrt{\frac{Q(x)}{-x_*}} \xrightarrow[x_* \to -\infty]{} \tilde{\epsilon}\sqrt{x^\star - x}$, we consider in this paragraph the stochastic process defined by

$$X_t^{x_0} = x_0 + \int_0^t (a - kX_s^{x_0})\,\mathrm{d}s + \tilde{\epsilon}\int_0^t \sqrt{x^\star - X_s^{x_0}}\mathrm{d}B_s, \tag{3.15}$$

where $(B_t)_{t\geq 0}$ is a Brownian motion. The computations of case (iii) can be adapted.

3.1.4.2.3 $x_* = -\infty$ **and** $x^\star = +\infty$ We now take $\tilde{\epsilon} \geq 0$ and set $\epsilon = \frac{\tilde{\epsilon}}{\sqrt{-x_* x^\star}}$ with $x_* \in \mathbb{R}_-$ and $x^\star \in \mathbb{R}_+$. Since $\tilde{\epsilon}\sqrt{\frac{Q(x)}{-x_* x^\star}} \to 1$ as $(x_*, x^\star) \to (-\infty, +\infty)$, we consider in this paragraph the stochastic process defined by

$$X_t^{x_0} = x_0 + \int_0^t (a - kX_s^{x_0})\,\mathrm{d}s + \tilde{\epsilon}B_t, \tag{3.16}$$

where $(B_t)_{t\geq 0}$ is a Brownian motion. The study reduces then to the that of an Ornstein-Uhlenbeck process introduced in [UO30]. (3.16) can be solved explicitly for $k > 0$:

$$X_t^{x_0} = x_0 e^{-kt} + \frac{a}{k}(1 - e^{-kt}) + \tilde{\epsilon}\int_0^t e^{-k(t-s)}\mathrm{d}B_s,\ t \geq 0.$$

By definition of the Wiener integral, X^{x_0} is a Gaussian process and is thus determined by its first two moments. It is straightforward to get that

$$X_t^{x_0} \sim \mathcal{N}\left(x_0 e^{-kt} + \frac{a}{k}(1-e^{-kt}), \frac{\tilde{\epsilon}^2}{2k}(1-e^{-2kt})\right),$$

and deduce that

$$\mathbb{E}\left[e^{-zX_t^{x_0}}\right] = \exp\left(-z\left(x_0 e^{-kt} + \frac{a}{k}(1-e^{-kt})\right) + \frac{\tilde{\epsilon}^2 z^2}{4k}(1-e^{-2kt})\right),\ z \in \mathbb{C}.$$

3.1.5 Link with Cox-Ingersoll-Ross process

We recall here the connection between the $[0,1]$-valued Jacobi and the Cox-Ingersoll-Ross processes derived notably in Section 6.1.3 (p. 191) of [Alf15]. This connection can be seen as an extension to continuous time framework the following well-known identity connecting gamma and beta distributions.

Lemma 3.7. *If $X_1 \sim \Gamma(k_1, \theta)$ and $X_2 \sim \Gamma(k_2, \theta)$, then $\frac{X_1}{X_1+X_2} \sim \beta(k_1, k_2)$ and is independent of $X_1 + X_2$.*

Proof. Set $Y = X_1 + X_2$, $Z = X_1/Y$ and let $h : \mathbb{R}^2 \to \mathbb{R}$ be bounded and measurable such that $\mathbb{E}\big[|h(Y,Z)|\big] < \infty$. Computing $\mathbb{E}\big[h(Y,Z)\big]$ by a change of variable allows to determine the density f of the couple (Y,Z) as

$$f(y,z) = \frac{\theta^{k_1+k_2}}{\Gamma(k_1)\Gamma(k_2)} e^{-\theta y} y^{k_1+k_2-1} z^{k_1-1}(1-z)^{k_2-1} \mathbb{1}_{\{y>0\}} \mathbb{1}_{\{0<z<1\}}.$$

□

Here comes the counterpart claim result on stochastic processes.

Proposition 3.8 ([Alf15])**.** *Let $\big((W_t^i)_{t\geq 0}\big)_{i=1,2}$ be two independent Brownian motions. Let be two CIR processes defined as*

$$Y_t^i = y_i + a_i t + \epsilon \int_0^t \sqrt{Y_s^i} \mathrm{d}W_s^i,\ i = 1,2,$$

where $a_1, a_2, y_1, y_2, \epsilon$ are non-negative parameters satisfying $\epsilon^2 \leq 2(a_1+a_2)$ and $y_1 + y_2 > 0$. Then, $Y = Y^1 + Y^2$ is a CIR process starting from $y_1 + y_2$, with parameters $a = a_1 + a_2$, $k = 0$ and ϵ, that never reaches zero. Define then

$$X_t^{x_0} = \frac{Y_t^1}{Y_t^1 + Y_t^2} \text{ and } \tau(t) = \int_0^t \frac{1}{Y_s} \mathrm{d}s. \tag{3.17}$$

for $t \geq 0$. $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ is a bijective function and by denoting $\mathcal{T} = \tau^{-1}$, the process $\left(X_{\mathcal{T}(t)}^{x_0}\right)_{t\geq 0}$ is a $[0,1]$-valued Jacobi process starting from $x_0 = \frac{y_1}{y_1+y_2}$ and with parameters $k = a_1 + a_2$, $a = k_1$, ϵ and that is independent of $(Y_t)_{t\geq 0}$.

This allows to deduce an alternative representation of the density function of the Jacobi process. To do so we resort on the definition of the noncentral beta distribution introduced in [Seb63].

Proposition 3.9. *Let $t \geq 0$. Let us denote by $c_t^1 = \frac{4a_1}{\epsilon^2(1-e^{-a_1 t})}$, $d_t^1 = c_t^1 e^{-a_1 t}$, $c_t = \frac{4(a_1+a_2)}{\epsilon^2(1-e^{-(a_1+a_2)t})}$ and $d_t = c_t e^{-(a_1+a_2)t}$. Conditionally on $\{\tau(t)\}$, $\frac{c^1_{\mathcal{T}(t)}}{c_{\mathcal{T}(t)}} X^{x_0}_{\mathcal{T}(t)}$ has a non central beta distribution $\beta'\left(\frac{a_1}{\epsilon^2}, \frac{a_1+a_2}{\epsilon^2}, d^1_{\mathcal{T}(t)} y_1, d_{\mathcal{T}(t)}(y_1+y_2)\right)$.*

Proof. $c_t^1 Y_t^1$ is distributed following a non central chi-square distribution (see for instance Proposition 1.2.11 in [Alf15]) with $\frac{2a_1}{\epsilon^2}$ degrees of freedom and non-centrality parameter $d_t^1 y_1$. The same stands for $c_t Y_t$ with respective parameters $\frac{2(a_1+a_2)}{\epsilon^2}$ and $d_t(y_1+y_2)$. The claim follows directly from the definition of the non-central beta distribution β'. □

Let us now consider the time-changed process $\tilde{X}^{x_0}_{\mathcal{T}(t)} := \frac{c^1_{\mathcal{T}(t)}}{c_{\mathcal{T}(t)}} X^{x_0}_{\mathcal{T}(t)}$. Properties of the non-central beta distribution can be employed to obtain that, roughly speaking, the marginal distribution of $\tilde{X}^{x_0}_{\mathcal{T}(\cdot)}$ can write as a beta distribution whose parameters are Poisson variates.

Proposition 3.10. *Let $t \geq 0$ and (N_1, N_2) be a couple of independent Poisson random variables of respective parameters $\left(\frac{d^1_{\mathcal{T}(t)} y_1}{2}, \frac{d_{\mathcal{T}(t)}(y_1+y_2)}{2}\right)$. Then, for any bounded measurable function f, we have*

$$\mathbb{E}\left[f\left(\tilde{X}^{x_0}_{\mathcal{T}(t)}\right)\right] = \mathbb{E}\left[\frac{\Gamma\left(\frac{a_1}{\epsilon^2} + N_1 + \frac{a_1+a_2}{\epsilon^2} + N_2\right)}{\Gamma\left(\frac{a_1}{\epsilon^2} + N_1\right)\Gamma\left(\frac{a_1+a_2}{\epsilon^2} + N_2\right)} \int_0^1 f(x) x^{\frac{a_1}{\epsilon^2}+N_1}(1-x)^{\frac{a_1+a_2}{\epsilon^2}+N_2} \mathrm{d}x\right].$$

Proof. Let (N_1, N_2) be a couple of independent Poisson random as above. From representation (3.17), we know that $\tilde{X}^{x_0}$ writes as a ratio of random variables distributed following non-central chi-square distributions. Furthermore, such distributions can be conveniently represented using independent Poisson and standard chi-square variables. Indeed, the density of the non-central chi-square distribution with k degrees of freedom and non-centrality parameter λ writes:

$$f_X(x; k, \lambda) = \sum_{n=0}^{+\infty} \frac{e^{-\lambda/2}(\lambda/2)^n}{n!} \frac{x^{\frac{k}{2}+n-1} e^{-x/2}}{2^{k/2+n}\Gamma\left(\frac{k+2n}{2}\right)}.$$

Furthermore considering $N \sim \mathcal{P}(\lambda/2)$ and Z a variate following a chi-square distribution with $k+2n$ degrees of freedom conditional on $\{N = n\}$; then Z is chi-square distributed with k degrees of freedom and non-centrality parameter λ. The claimed result is a consequence of Lemma 3.7. □

In particular, the density function of the distribution of $X^{x_0}_{\mathcal{T}(t)}$ conditional on $\{\tau(t)\}$ writes as

$$\begin{aligned} f_{\tilde{X}^{x_0}_{\mathcal{T}(t)}}(x) &= e^{-(y_1 d^1_{\mathcal{T}(t)} + (y_1+y_2) d_{\mathcal{T}(t)})/2} \times \\ &\sum_{i,j=0}^{+\infty} \frac{\left(d^1_{\mathcal{T}(t)} y_1/2\right)^i \left(d_{\mathcal{T}(t)}(y_1+y_2)/2\right)^j}{i!j!} \frac{x^{\frac{a_1}{\epsilon^2}+i-1}(1-x)^{\frac{a_1+a_2}{\epsilon^2}+j-1}}{B\left(\frac{a_1}{\epsilon^2}+i, \frac{a_1+a_2}{\epsilon^2}+j\right)} \mathbb{1}_{0 \leq x \leq 1}. \end{aligned}$$

The inverse time deformation $\mathcal{T} = \tau^{-1}$ can be obtained using the generalized inverse function:

$$\mathcal{T}(t) = \inf\{s : \tau(s) > t\}. \tag{3.18}$$

Remark 12. *For a given $\omega \in \Omega$, $\mathcal{T}$ is a strictly increasing and continuous function. For such functions, generalized inverse function as defined in (3.18) and standard inverse function*

(when it exists) coincide. The interested reader can refer to [EH13] and references therein for a discussion on the topic.

For the sake of completeness, we provide some details on the time deformation τ. As previously discussed, Y is a CIR process starting from $y_1 + y_2$ whose parameters are $a = k_1$, $k = a_1 + a_1$ and ϵ. Since $\epsilon^2 \leq 2(a_1 + a_2)$, Y does not reach zero in finite time which allows to define the inverse process $1/Y$, starting from $1/(y_1 + y_2)$. Itô's formula provides that $1/Y$ satisfies the following SDE

$$\mathrm{d}X_t = \left(k - (a - \epsilon^2)X_t\right) X_t \mathrm{d}t - \epsilon X_t^{3/2} \mathrm{d}W_t \tag{3.19}$$

where $(W_t)_{t\geq 0}$ is a Brownian motion. The process defined by (3.19) is usually referred to as the 3/2 process in the literature. In [AG99] the authors proposed to use this dynamics for short rate modelling and derive some closed-form formula for Zero-Coupon bond prices. [CS07] studied stochastic volatility dynamics in interest rates context and studied variance swap derivatives; in their work, the variance process is modelled using the previous 3/2 dynamics. Authors have been able to derive explicit expression for the joint characteristic function of the log-driver and the integrated variance processes. Authors actually extended the works of [Hes97] and [Lew] in order to get semi-analytical formulas for prices of derivatives on realized variance. In [GV19], authors worked in equity environment and employed similar modelling where the variance process follows a 3/2 dynamics; the model they considered is sometimes named the «3/2 Heston model». They adapted the analytical expressions of [CS07] and [AG99] to get closed-form formulas for stock derivatives. We also mention the survey [KN12] that gathers several results on the 3/2 Heston model.

In present context, we are looking for the distribution of the integrated 3/2 process $\left(\int_0^t \frac{1}{Y_s}\mathrm{d}s\right)_{t\geq 0}$. The results of [CS07] are of particular interest then. The following assertion specifies the density function of the process (3.19) using the inversion formula (Fourier transform) often used for pricing of derivatives. The expression of the characteristic function of the 3/2 process is thus required: Proposition 1 in [AG99] (p.742) gives the Laplace transform of the 3/2 dynamics; Theorem 3 in [CS07] additionally provides the characteristic function of the integrated 3/2 model. Both obtained their results using the PDEs solved by the Laplace/characteristic functions.

In the following, the confluent hypergeometric function (or Kummer's function, named after Ernst Kummer) will be needed. It is defined for all $(a, z) \in \mathbb{C}^2$ and $b \in \mathbb{C}\backslash\{\mathbb{Z}^- \cup \{0\}\}$ $(b \neq 0, -1, -2, \dots)$ as

$$M(a, b, z) = \sum_{i=0}^{+\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}$$

where $(a)_n = a(a+1)(a+2) \times \cdots \times (a+n-1)$ is the standard Pochhammer symbol (introduced by Leo Pochhammer). We denote the characteristic function of the integrated 3/2 process by

$$\phi\left(u; \int_0^t X_s \mathrm{d}s\right) = \mathbb{E}\left[\exp\left(iu \int_0^t X_s \mathrm{d}s\right)\right] \tag{3.20}$$

defined for $u \in \mathbb{R}$ and where i is the imaginary unit satisfying $i^2 = -1$.

The following result is nothing else than an application of the inversion formula for identifying density function based on the explicit knowledge of the characteristic function. It is a very common technique for derivatives pricing in mathematical finance to resort on what is actually a Fourier transform applied to the characteristic function of some integrated process. In case

of the 3/2 process, the mentioned characteristic function can be expressed as

$$\phi\left(u; \int_0^t X_s \mathrm{d}s\right) = \frac{\Gamma(\beta-\alpha))}{\Gamma(\beta)} x(0,t)^\alpha M\left(\alpha, \beta, -x(0,t)\right), \ u \in \mathbb{R}, \tag{3.21}$$

where

$$x(0,T) = \frac{2kY_0}{\epsilon^2(e^{kT}-1)}, \ \alpha = -\left(\frac{1}{2} + \frac{a-\epsilon^2}{\epsilon^2}\right) + \sqrt{\left(\frac{1}{2} + \frac{a-\epsilon^2}{\epsilon^2}\right)^2 - \frac{2iu}{\epsilon^2}},$$

$$\text{and } \beta = 2\alpha + 2\left(1 + \frac{a-\epsilon^2}{\epsilon^2}\right) = 1 + 2\sqrt{\left(\frac{1}{2} + \frac{a-\epsilon^2}{\epsilon^2}\right)^2 - \frac{2iu}{\epsilon^2}}$$

have been introduced. The interested reader can refer to [AG99], [CS07] or [KN12] for a justification of this analytical expression.

3.1.6 Integrated Jacobi process

In some modelling applications, the integrated Jacobi process may be useful:

$$I_t = \int_0^t X_s^{x_0} \mathrm{d}s$$

where $(X_t^{x_0})_{t\geq 0}$ is defined in Equation (3.2). The Laplace function $\phi(z; I_t) = \mathbb{E}\left[\exp\left(-zI_t\right)\right]$ of I is evaluated in $z = 1$ in [DS02] and derived in a semi-analytical way. Following the technique of [Duf01], we can extend the expression to any $z \geq 0$ taking advantage of the formula obtained by [DS02].

Let us denote by $Y_t^{y_0} = zX_t^{x_0}$ for some $z \geq 0$. We have

$$Y_t^{y_0} = y_0 + \int_0^t (\bar{a} - kY_u^{y_0})\mathrm{d}u + \epsilon \int_0^t \sqrt{(Y_u^{y_0} - \bar{x}_*)(\bar{x}^\star - Y_u^{y_0})}\mathrm{d}W_u$$

with $y_0 = zx_0$, $\bar{a} = za$, $\bar{x}_* = zx_*$ and $\bar{x}^\star = zx^\star$. In the following expression, we will need the quantities:

$$A = \frac{2}{\epsilon^2}\frac{\bar{a} - k\bar{x}_*}{\bar{x}^\star - \bar{x}_*} = \frac{2}{\epsilon^2}\frac{a - kx_*}{x^\star - x_*}, \ B = \frac{2}{\epsilon^2}\frac{k\bar{x}^\star - \bar{a}}{\bar{x}^\star - \bar{x}_*} = \frac{2}{\epsilon^2}\frac{kx^\star - a}{x^\star - x_*} \text{ and for } n \in \mathbb{N},$$

$$\Lambda_n = kn + \frac{\epsilon^2}{2}n(n-1), \ K_n = \frac{(A+B+2n-1)\Gamma(A+n)\Gamma(A+B+n-1)}{n!\Gamma(A)^2\Gamma(B+n)},$$

$$(x)_k = \frac{\Gamma(x+k)}{\Gamma(x)} \text{ and } \psi_n(x) = \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{(A+B+n-1)_k}{(A)_k} x^k.$$

Proposition 3.11. *The Laplace function of the integrated process writes, for $z \geq 0$:*

$$\mathbb{E}\left[\exp\left(-zI_t\right)\right] = \mathbb{E}\left[\exp\left(-\int_0^t Y_u^{y_0}\mathrm{d}u\right)\right]$$

$$= e^{-zx_*t}\left(1 + \sum_{n\geq 1} z^n (x^\star - x_*)^n \times \sum_{(v_n,\ldots,v_1)\in\mathcal{V}^n} \psi_{v_n}\left(\frac{x_0 - x_*}{x^\star - x_*}\right) \prod_{j=n}^1 K_{v_j} q(v_j, v_{j-1}) S_t^{\otimes n}(\Lambda_{v_n}, \ldots, \Lambda_{v_1})\right)$$

where

$$\mathcal{V}^n = \{(v_n, \dots, v_1) \in \mathbb{N}^n : |v_j - v_{j-1}| \leq 1, 1 \leq j \leq n, v_0 = 0\},$$

$$S_t^{\otimes n}(\lambda_n, \dots, \lambda_1) = \int_0^t \int_{s_n}^t \cdots \int_{s_2}^t \exp\left(-\sum_{j=n}^{1} \lambda_j (s_j - s_{j+1})\right) \mathrm{d}s_1 \dots \mathrm{d}s_n,$$

$$q(v_j, v_{j-1}) = \begin{cases} \frac{\left(2v(A+B+v-1)+A(A+B-2)\right)\Gamma(A)^2 v!\Gamma(B+v)}{(A+B+2v)(A+B+2v-1)(A+B+2v-2)\Gamma(A+v)\Gamma(A+B+v-1)} & \text{if } v_j = v_{j-1}, \\ \frac{v!\Gamma(A)^2\Gamma(B+v)}{(A+B+2v-1)(A+B+2v-2)(A+B+2v-3)\Gamma(A+v-1)\Gamma(A+B+v-2)} & \text{otherwise,} \end{cases}$$

with $v = \max(v_j, v_{j-1})$.

Proof. Application of Theorem 3.1 in [DS02]. □

Remark 13. *The obtained formula is hard to use in practice. Section 5 in [DS02] is dedicated to the derivation of relevant numerical approximations of it for practical uses.*

3.1.7 Simulation of the Jacobi process

We briefly discuss in present paragraph how the Jacobi process can be simulated in practice focusing on two common methods: discretization scheme for simulating solution of (3.2) and weak approximations designed for evaluating marginal distributions of $X_t^{x_0}$.

3.1.7.1 Discretization schemes

We begin with presentation of some relevant discretization schemes for simulation of Jacobi process paths. The present study is very linked to that of the simulation of the Cox-Ingersoll-Ross process (a common reference on the subject is also [Alf15] and references therein). It has constituted an active research field in numerical methods for stochastic processes due to the presence of the square root function in diffusion coefficient of (3.2) making standard numerical schemes found in the literature irrelevant here.

The Euler-Maruyama scheme is probably the most common discretization scheme in the literature. For discretization of SDE with smooth enough coefficients (i.e. with linear growth and Lipschitz regularity in the space variable), the Euler scheme is known to well perform with a strong convergence rate of $O(N^{-1/2})$ (see for instance [K+88]) and weak convergence rate of $O(N^{-1})$ for very irregular function (see [TT90]), by denoting the grid time step as $1/N$. Numerous studies have been lead in order to extend these results to a wider class of coefficients: [HMS02] provided some results on strong convergence with simply locally Lipschitz coefficient and polynomial drifts; [GR11] studied the case of Hölder regularity; [Y+02] studied irregular coefficients and focused on the weak convergence of the scheme.

In [LKD10] the authors proposed to *truncate* the Euler scheme for ensuring the proper definition of the scheme applied to the simulation of the CIR and proved the strong convergence of this scheme. Following their work, we first introduce a scheme we will refer to as "Full Truncation Euler" scheme (FTE). We want to simulate the Jacobi process (3.2) up to a time horizon T. On a fixed time grid $t_k = \frac{kT}{N}$, the proposed FTE scheme writes, for $k \in \{1, \dots, N-1\}$,

$$\text{(FTE): } X_{t_{k+1}}^N = X_{t_k}^N + (a - k\mathrm{tr}(X_{t_k}^N))(t_{k+1} - t_k) + \epsilon\sqrt{Q(\mathrm{tr}(X_{t_k}^N))}(W_{t_{k+1}} - W_{t_k}),$$

where we reuse the truncating function $\mathrm{tr}(x) = \min(\max(x, x_*), x^\star)$ defined in (3.3). To analyze

this scheme, it is convenient to introduce its continuous in time version

$$\bar{X}_t^N = x_0 + \int_0^t \left(a - k\mathrm{tr}(\bar{X}_{\tau_s}^N)\right) \mathrm{d}s + \epsilon \int_0^t \sqrt{Q\left(\mathrm{tr}(\bar{X}_{\tau_s}^N)\right)} \mathrm{d}W_s, \tag{3.22}$$

where τ_t denotes the last time in the grid before t: for all $t \in [t_k, t_{k+1}[$, $\tau_t = t_k$. Note that $\bar{X}^N$ satisfying (3.22) is well defined.

Pathwise convergence

The fact that the Jacobi process remains bounded over time can be used to get a strong pathwise convergence rate of order «almost» 1/2.

Proposition 3.12 ([Gyö98]). *Assume that the Feller condition (3.8) holds. Then, for every $T \geq 0$ and $\eta > 0$, there exists some finite positive random variable U such that the following holds almost surely:*

$$N^{1/2-\epsilon} \sup_{t \leq T} |X_t^{x_0} - \bar{X}_t^N| \leq U.$$

Proof. Under the Feller condition (3.8), the state space of the considered Jacobi process is $D =]x_*, x^\star[$. Consider the compact sets defined by $D_k = [x_* + 1/k, x^\star - 1/k]$ for $k \in \mathbb{N}^*$ such that $D = \cup_{k \in \mathbb{N}^*} D_k$. As a composition of Lipschitz functions, $x \in D \mapsto (a - k\mathrm{tr}(x))$ is also Lipschitz, in particular locally Lipschitz; the diffusion function $x \in D \mapsto \sqrt{Q(\mathrm{tr}(x))}$ is also a locally Lipschitz function as its derivative is locally bounded. A straightforward application of Theorem 2.4 of [Gyö98] gives that, as N goes to infinity, the approximating diffusion (3.22) converges almost surely uniformly in time over bounded intervals toward a process $(\bar{X}^{x_0})_{t \geq 0}$ solution of

$$\bar{X}_t^{x_0} = x_0 + \int_0^t \left(a - k\mathrm{tr}(\bar{X}_s^{x_0})\right) \mathrm{d}s + \epsilon \int_0^t \sqrt{Q\left(\mathrm{tr}(\bar{X}_s^{x_0})\right)} \mathrm{d}W_s.$$

As discussed for the existence of the Jacobi process (see proof of Proposition 3.1), the truncating function can be omitted in the above stochastic differential equation. The claim is then proved. □

Remark 14. *As pointed out in [DNS12], this result holds also true for simulation of the CIR process using the symmetrized Euler scheme proposed in [BD04], the partial truncated scheme of [DD+98] or the scheme introduced in [HM05].*

Remark 15. *Jacobi paths illustrated in Figure 3.1 were obtained following (3.22) discretization scheme.*

$\mathbb{L}^p$ convergence

Better convergence rates can be obtained in $\mathbb{L}^p$ spaces. The result presented in this section is obtained by considering a drift implicit Euler scheme also often named backward Euler scheme. Such implicit schemes are studied in general setting in [Alf12] and [NS14]. This technique allows to ensure that the numerical scheme is well defined under rather general assumptions.

For simplicity, we focus in this paragraph on $[0,1]$-valued Jacobi process $(Y_t^{y_0})_{0 \leq t \leq T}$. The results presentend in this section can apply to any $[x_*, x^\star]$-valued Jacobi process which is easily seen when considering the translated process $\left(X_t^{(x^\star - x_*)y_0 + x_*} = (x^\star - x_*)Y_t^{y_0} + x_*\right)_{0 \leq t \leq T}$. The following proposition is proved in [NS14]. Let us consider the transformed stochastic process defined by $Z_t^{z_0} \overset{a.s.}{=} 2\arcsin(\sqrt{Y_t^{y_0}})$ for any time $t \geq 0$. Itô's formula provides that

$$dZ_t^{z_0} = f(Z_t^{z_0})\mathrm{d}t + \epsilon \mathrm{d}W_t. \tag{3.23}$$

where f is the function defined over $]0,\pi[$ by

$$f(x) = \left(a - \frac{\epsilon^2}{2}\right)\cot\left(\frac{x}{2}\right) - \left(k - a - \frac{\epsilon^2}{4}\right)\tan\left(\frac{x}{2}\right),$$

where $x \mapsto \tan(x)$ is the tangent function and $\cot = 1/\tan$ is the cotangent function. Let us consider the drift implicit Euler scheme $(\bar{Z}_{t_k})_{k=1,\dots,N}$ defined by

$$\bar{Z}^N_{t_{k+1}} = \bar{Z}^N_{t_k} + f(\bar{Z}^N_{t_{k+1}})(t_{k+1} - t_k) + \epsilon(W_{t_{k+1}} - W_{t_k}),\ k = 1, \dots, N-1$$

for approximating solution of (3.23).

Proposition 3.13 ([NS14])**.** *Assume that* $\epsilon^{-2}\min\{a,(k-a)\} > 1$ *and let* $1 \le p < \frac{4}{3}\epsilon^{-2}\min\{a,(k-a)\}$*. Then*

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|Y^{y_0}_t - \sin^2\left(\bar{Z}^N_t/2\right)\right|^p\right] \le \frac{C_p}{N^p}.$$

Proof. A straightforward adaptation of the computations made in Section 3.5 of [SN14] allows to check the required conditions to apply Proposition 3 of [Alf12]. The result follows. □

Remark 16. *Under the relaxed condition* $2a > \epsilon^2$*, a strong convergence of order* $1/2$ *has been shown in [NS14] for the simulation of the CIR process.*

Remark 17. *The presented rate is actually the best known convergence rate for the CIR process under quite general assumptions.*

3.1.7.2 Weak approximation

The interested reader to Section 6.1.5 (p.197) in [Alf15] and references therein for a fuller discussion on weak approximation schemes with square-root like diffusive functions.

For some practical purposes there is no systematic needs to strongly approximate the paths of the Jacobi process and it may be enough to accurately approximate quantities writing as $\mathbb{E}\left[f(X^{x_0}_T)\right]$ for proper function f. In this section, we present a second order weak scheme for approximating a $[-1,1]$-Jacobi valued process $(Y^{y_0})_{0\le t\le T}$; as previously mentioned, we can get back to a $[x_*, x^\star]$-Jacobi process by setting $X^{x_0}_t = \frac{2}{x^\star - x_*}(Y^{y_0}_t - x_*) - 1$ and $x_0 = \frac{2}{x^\star - x_*}(y_0 - x_*) - 1$. The infinitesimal generator of $(Y^{y_0})_{0\le t\le T}$ applied to a twice differentiable function f writes

$$\mathcal{A}f(x) = (a - kx)f'(x) + \frac{\epsilon^2}{2}(1-x)(x-1)f''(x).$$

It is decomposed in two sub-infinitesimal generators $\mathcal{A}f = \mathcal{A}^1 f + \mathcal{A}^2 f$ with $\mathcal{A}^1 f(x) = \left(a - (k - \frac{\epsilon^2}{2})x\right)f'(x)$ and $\mathcal{A}^2 f(x) = -\frac{\epsilon^2}{2}xf'(x) + \frac{\epsilon^2}{2}(x - x_*)(x^\star - x)f''(x)$. Ordinary differential equation associated to $\mathcal{A}^1$ is a linear first order differential equation which is solved by $f^1(x,t) = xe^{-(k-\frac{\epsilon^2}{2})t} + a\frac{1-e^{-(k-\frac{\epsilon^2}{2})t}}{(k-\frac{\epsilon^2}{2})}$ if $(k - \frac{\epsilon^2}{2}) \neq 0$, $f^1(x,t) = xe^{-(k-\frac{\epsilon^2}{2})t} + at$ otherwise. Observe that $f^1(x,t) \in [-1,1]$ for any $0 \le t \le T$ if and only if $a - (k - \epsilon^2/2) \le 0$ and $a + (k - \epsilon^2/2) \le 0$. This is in particular the case when Feller condition (3.8) is satisfied. Regarding the second generator $\mathcal{A}^2$, it is associated with the stochastic differential equation

$$\mathrm{d}X_t = -\frac{\epsilon^2}{2}\mathrm{d}t + \epsilon\sqrt{1 - X_t^2}\mathrm{d}W_t.$$

Its solution is given by $X_t^{x_0} = \sin(\tilde{x}_0 + \epsilon W_t)$ with $\tilde{x}_0$ satisfying $\sin(\tilde{x}_0) = x_0$. The Ninomiya-Victoir scheme introduced in [NV08] writes in present context as
$\bar{Y}_t = f^1(\sin\left(\arcsin\left(f^1(x, t/2)\right) + \epsilon W_t\right), t/2)$ and turns out to have a weak error of order 2.

Proposition 3.14. *For $f \in \mathcal{C}^\infty_{pol}([-1,1])$, there exists a real number $C > 0$ and an integer $n_0 \in \mathbb{N}$ such that*

$$\forall n \geq n_0,\ |\mathbb{E}\left[f(\bar{Y}_{t_n})\right] - \mathbb{E}\left[f(Y_T^{y_0})\right]| \leq C/n^2.$$

Proof. See section 6.1.5 in [Alf15]. Numerical schemes when the Feller condition (3.8) is not satisfied are also discussed. □

3.1.8 Extensions

3.1.8.1 Time-dependent coefficients

To get a richer parametrization of the Jacobi process, one can consider the following time-dependent stochastic differential equation

$$X_t^{x_0} = x_0 + \int_0^t \left(a(s) - k(s)X_s^{x_0}\right)\mathrm{d}s + \int_0^t \epsilon(s)\sqrt{Q(X_s^{x_0})}\mathrm{d}B_s, \tag{3.24}$$

where now $a : \mathbb{R}_+ \longrightarrow \mathbb{R}$, $k : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$. Assuming that a, k and ϵ are measurable, bounded and that those bounds satisfy some inequalities is enough to get the existence and uniqueness of the process defined in (3.24) (we can proceed as in non-time dependent case). In particular, we introduce the respective lower and upper bounds of the functions a, k and ϵ: for any $t \geq 0$, $\underline{a} \leq a(t) \leq \bar{a}$, $\underline{k} \leq k(t) \leq \bar{k}$ and $0 < \underline{\epsilon} \leq \epsilon(t) \leq \bar{\epsilon}$. The mentioned inequalities write: $\underline{a} \geq \bar{k}x_*$ and $\bar{a} \leq \underline{k}x^\star$.

The following condition is a continuous in time version of the Feller condition (3.8). It ensures the process (3.24) to remain bounded over time.

Proposition 3.15. *Assume that $x_0 \in (x_*, x^\star)$ and*

$$x^\star - x_* \leq 2\min\left(\frac{\underline{k}x^\star - \bar{a}}{\underline{\epsilon}^2}, \frac{\underline{a} - \bar{k}x_*}{\bar{\epsilon}^2}\right). \tag{3.25}$$

Then X^{x_0} defined in (3.24) never reaches its boundaries:

$$\mathbb{P}\left(\forall t \in [0,T] : x_* < X_t^{x_0} < x^\star\right) = 1.$$

The following proof uses standard argument: to get rid of the time dependency in the Itô integral, a time change is made; comparison theorem for SDEs allows to get the result since coefficients are bounded. This procedure is for instance applied in [BGM10] for the study of time-dependent CIR process.

Proof. Let the time change function $t \mapsto \phi_t$ be defined by

$$t = \int_0^{\phi_t} \epsilon(s)^2\mathrm{d}s.$$

Since $\underline{\epsilon} > 0$, ϕ is a strictly increasing time change; moreover, it is continuous by definition and so is its inverse ϕ^{-1}. Let us consider now the process

$$\tilde{B}_t = \int_0^{\phi_t} \epsilon(s)\mathrm{d}B_s.$$

Proposition 5.1.5 of [RY13] (p.170) provides that $\langle \tilde{B}_\cdot \rangle_t = \int_0^{\phi_t} \epsilon(s)^2 \mathrm{d}s = t$ and that $(\tilde{B}_t)_{0 \le t \le \phi_T^{-1}}$ is a continuous local martingale. From Lévy's characterization theorem we deduce that $(\tilde{B}_t)_{0 \le t \le \phi_T^{-1}}$ is a Brownian motion.

Now, for $t \in [0, \phi_T^{-1}]$, we set $\tilde{X}_t^{x_0} = \tilde{X}_{\phi_t}^{x_0}$ and according to Propositions 5.1.4 and 5.1.5 in [RY13], we write

$$\tilde{X}_t^{x_0} = x_0 + \int_0^t \frac{1}{\epsilon(\phi_s)^2} \big(a(\phi_s) - k(\phi_s)\tilde{X}_s^{x_0}\big)\mathrm{d}s + \int_0^t \sqrt{Q(\tilde{X}_s^{x_0})}\mathrm{d}\tilde{B}_s.$$

To study $\tilde{X}^{x_0}$ we priori introduce the two stochastic processes defined by time-independent SDEs:

$$X_t^1 = x_0 + \int_0^t \left(\bar{\epsilon}^{-2}\big(\underline{a} - \bar{k}X_s^1\big)\right) \mathrm{d}s + \int_0^t \sqrt{Q(X_s^1)}\mathrm{d}B_s, \tag{3.26}$$

$$X_t^2 = x_0 + \int_0^t \left(\underline{\epsilon}^{-2}\big(\bar{a} - \underline{k}X_s^2\big)\right) \mathrm{d}s + \int_0^t \sqrt{Q(X_s^2)}\mathrm{d}B_s. \tag{3.27}$$

Feller condition (3.8) that $\mathbb{P}\left(\forall t \in [0, \phi_T^{-1}] : X_t^1 > x_*\right) = 1$ and $\mathbb{P}\left(\forall t \in [0, \phi_T^{-1}] : X_t^2 < x^\star\right) = 1$. Now, since for any $(x,t) \in [x_*, x^\star] \times \mathbb{R}_+$, $\big(1/\underline{\epsilon}^2\big)\big(\bar{a} - \underline{k}x\big) \ge \epsilon(t)^{-2}\big(a(t) - k(t)x\big) \ge \big(1/\bar{\epsilon}^2\big)\big(\underline{a} - \bar{k}x\big)$ we can use twice comparison theorem (see for instance Theorem 5.2.18 in [KS91] p.293) to get that $\mathbb{P}\left(\forall t \in [0, \phi_T^{-1}] : x^\star > X_t^2 \ge \tilde{X}_t^{x_0}\right) = 1$ and $\mathbb{P}\left(\forall t \in [0, \phi_T^{-1}] : \tilde{X}_t^{x_0} \ge X_t^1 > x_*\right) = 1$. Hence $\mathbb{P}\left(\forall t \in [0, \phi_T^{-1}] : x^\star > \tilde{X}_t^{x_0} > x_*\right) = 1$ which proves the claim. □

The Jacobi process lies in the family of *polynomial processes* introduced in [EP11] or [CKRT12]: moments of such processes can be analytically computed and account for an extension of the affine processes. Time inhomogeneous polynomial processes have been introduced in [dCAH18] as extension of the previous works. Below in Section 3.1.9, we present them in general framework requiring less regularity than in [dCAH18].

3.1.8.2 Rough Jacobi process

Fractional Brownian motion (fBM for short) is a popular example of non Markovian Gaussian process offering a rather good tractability. Its regularity is monitored by the Hurst index usually denoted $H \in]0,1[$: the case $H = 1/2$ allows to recover the standard Brownian motion. $H > 1/2$ produces paths that move «slower» than the standard Brownian motion; rather, the case $H < 1/2$ produces *rough* paths that move «faster» than standard Brownian motion. Some key modelling properties of the fractional Brownian motion (fBM) have been established in [MVN68]. This process can be differently represented as integrals of some kernels with respect to the standard Brownian motion (sometimes named Volterra-Itô integrals): an overview of commonly used kernels can be found in [She14]. Quite a few properties of standard Itô's calculus can be extended to the stochastic calculus based on fractional Brownian motion and non semi-martingales. [D+99] proposed an extension of Itô's formula, Itô-Clark-Ocone representation and Girsanov theorem in particular case when $H > 1/2$. [DHPD00] defined integrals with respect to fBM analogously to Itô and Stratanovitch integrals using Wick product and derived some Itô type formula. [R+02] also provide a number of theoretical results on stochastic differential equations driven by fBM in case $H > 1/2$. Note that to establish the fractional stochastic calculus, the key tool of the mentioned works is the Malliavin calculus.

In financial modelling, a fractional Black-Scholes model has been introduced in [EVDH03] and [HØ03]. Following a similar approach, they both derived a Black-Sholes type formula for

options prices whose underlying is modelled using a geometric fBM. However, using such process to model underlying is not consistent with usual arbitrage free assumption that is closely linked to Itô's calculus. In a fractional environment the classical arbitrage free requirement should be conveniently modified to ensure the consistence of the pricing method; however, the practical economic notion of arbitrage in practice is very linked to the standard Itô's calculus (since it involves self-financing strategies). This point has been discussed and illustrated in [BH05]. We also refer to [SV03] for a discussion on this topic.

The theory of rough paths ($H < 1/2$) for financial modelling has recently gained in popularity to account for a number of stylized facts. The modelling of stochastic volatility using rough paths has lately spread among practitioners endorsed by strong empirical studies. The work of [GJR18] is key for motivating the use such models. Authors notably argued for modelling based on fBM with Hurst index around 10% to simulate volatility factor. The counterpart of the standard Heston model has became popular: this is the so-called rough Heston model. It was introduced in [EER19], in which authors derived an analytical formula for the characteristic function of the log-price process. The authors also address the question of hedging using this model in [EER+18]. [AJEE19] proposed an approximation of rough Heston model well-suited for numerical implementations that is based on Markovian approximating structure. A rather general theoretical framework is used in [JLP+19] defining rough stochastic process as solution of convolution type Volterra equations. Authors take advantage of this representation to deduce a number of properties on the considered processes.

An alternative approach has been chosen by [D+14], who introduced the $[0,1]$-valued fractional Jacobi process for $H > 1/2$ and [MYT18], who introduced the fractional Cox-Ingersoll-Ross process ($H \in (0,1)$). Both defined the mentioned processes as solutions of stochastic differential equations driven by fractional Brownian motions with time independent diffusive terms which eases the study. In this section, we briefly present the methodology of [D+14] to give insights on the fractional Jacobi process in case when $H > 1/2$.

Let us consider a fBM $(B_t^H)_{t\geq 0}$ with Hurst index $H \in (1/2, 1)$. Let us consider the following equation

$$X_t^{x_0,H} = x_0 + \int_0^t (a - kX_s^{x_0,H})\mathrm{d}s + \int_0^t \epsilon\sqrt{X_s^{x_0,H}(1 - X_s^{x_0,H})}\mathrm{d}B_s^H \tag{3.28}$$

where $x_0 \in [0,1]$.

Proposition 3.16 ([D+14]). *For $0 < a < k$, there exists a unique solution satisfying (3.28) taking its values in $(0,1)$ almost surely for any time $t > 0$. Moreover, if $x_0 \in (0,1)$, there exists $\eta > 0$ such that the solution of (3.28) takes its values in $[\eta, 1-\eta]$ almost surely for any time $t > 0$.*

Proof. We sketch the proof of Proposition 2.1 and Lemma 2.1 in [D+14]. The author worked on the following stochastic differential equation

$$Y_t^{y_0,H} = y_0 + \int_0^t \frac{2a - k(1 + \sin(Y_s^{y_0}))}{\cos(Y_s^{y_0,H})}\mathrm{d}s + \epsilon B_t^H \tag{3.29}$$

where $y_0 \in [-\pi/2, \pi/2]$, which admits a solution when $0 < a < k$ over $[0, \tau[$ where τ is the first hitting time of $\{-\pi/2, \pi/2\}$. The proof takes advantage of the regularity of the drift function $f(x) = \frac{2a-k(1+\sin(x))}{\cos(x)}$ over $]-\pi/2, \pi/2[$ and of the β-Hölder regularity of the fBM for $\beta < H$ to show that $\tau = \infty$ almost surely. The unicity comes from fact that diffusive term in (3.29) is constant and that f is decreasing over $]-\pi/2, \pi/2[$. Existence of $\eta' > 0$ such that $Y^{y_0,H} \in [-\pi/2 + \eta', \pi/2 - \eta']$ almost surely follows from continuity of the path. Finally, those results transpose to $X^{x_0,H}$ defined in (3.28) by the change of variable $X^{x_0,H} = \frac{1}{2}(\sin(Y^{y_0,H}) + 1)$

using for instance Itô's formula in [DHPD00] (Theorem 4.5). □

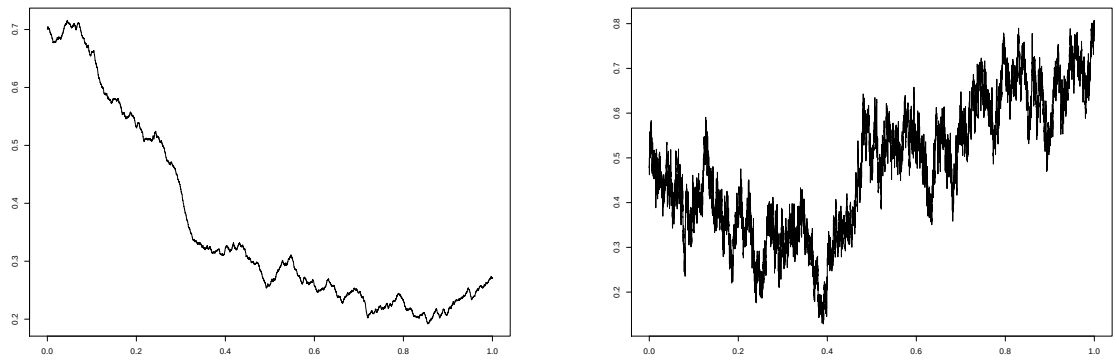
For volatility modelling, it can be interesting to consider a rough Jacobi process, that is a stochastic process driven by a fBM of Hurst index $H \in]0, 1/2[$. Following the framework of [EER19] and [JLP+19], we can consider the following Volterra equation

$$X_t^{x_0,H} = x_0 + \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}(a - kX_s^{x_0,H})\mathrm{d}s + \epsilon \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\sqrt{Q\big(\mathrm{tr}(X_s^{x_0,H})\big)}\mathrm{d}B_s \tag{3.30}$$

where $x_0 \in [x_*, x^\star]$, $\alpha = H + 1/2 \in]1/2, 1[$, Q is the bounding function defined in (3.1), $x \mapsto \mathrm{tr}(x) = \min\left(\max\left(x, x_*\right), x^\star\right)$ is the truncating function introduced in the proof of Theorem 3.1 and $(B_t)_{t\geq 0}$ is a standard Brownian motion.

Proposition 3.17 ([JLP+19])**.** *(3.30) admits a continuous weak solution for any $x_0 \in [x_*, x^\star]$.*

Proof. Direct application of Theorem 3.4 of [JLP+19] since (i) $x \mapsto a - kx$ and $x \mapsto \sqrt{Q\big(\mathrm{tr}(x)\big)}$ are continuous with linear growth and (ii) the suitable kernel $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ admits a resolvent of the first kind (see [JLP+19] for the details). □



(a) $x_* = 0$, $x^\star = 1$, $x_0 = 0.7$, $a = 0.5$, $k = 0.6$, $\epsilon = 0.8$ and $H = 3/4$.

(b) $x_* = 0$, $x^\star = 1$, $x_0 = 0.5$, $a = 0.5$, $k = 0.6$, $\epsilon = 0.4$ and $H = 1/4$.

Figure 3.2: Illustration of fractional Jacobi paths for various parametrizations.

3.1.8.3 Matrix valued process

A natural extension of the considered framework so far is to consider the multi-dimensional process, namely the matrix Jacobi process. The curious reader about random matrices and stochastic processes can refer to [Dou05] and references therein. According to this thesis, the origin of the study of matrix-valued stochastic processes comes from statistical considerations of John Wishart back in the 1930's and from physics matters of Eugene Wigner in the 1950's. Both were led to wonder how can the distributions of the eigenvalues of a matrix whose components are random variables be described.

In financial modelling, matrix processes has naturally appeared when considering multi-dimensional quantities or correlated drivers. For instance, [GS04] proposed a multidimensional assets models extending the standard Heston model. In their settings, log-prices processes are driven by a matrix volatility process that is a Wishart process. [GJ06] studied the multidimensional Jacobi process and discussed its use to represent dynamics of a discrete probability

distribution. [DFGT07] also proposed a multi-dimensional equity model but here the variance-covariance matrix describing the asset vector is assumed to evolves stochastically. In a similar framework, [Pal15] studied an Ornstein-Uhlenbeck variance-covariance process. [Ben10] worked on one dimensional asset process driven by a stochastic volatility matrix.

In this paragraph, we briefly present some of the results of [Dou05] on matrix Jacobi process (Chapter 9). Let us introduce some notations: $\mathcal{M}_{m,n}$ is the set of $m \times n$ real matrices, M^T denotes the transpose of M, $\mathbb{S}_n$ is the set of real symmetric matrices of size $n \in \mathbb{N}$, $\mathcal{SO}_n$ represents the special orthogonal group (set of orthogonal matrices with determinant 1), $\mathcal{S}_n$ be the set of skew-symmetric matrices[2], $\tilde{\boldsymbol{O}}_m \in \mathcal{M}_{m,m}$ is the matrix which components are all equal to zero, Id_m is the identity matrix of size m, $\Pi_{m,p} = \{M \in \mathcal{M}_{m,p} | M^T M \leq Id_m\}$ where inequality is understood componentwise, $\hat{\mathbb{S}}_m = \{M \in \mathbb{S}_m | \tilde{\boldsymbol{O}}_m < M < Id_m\}$, $\bar{\mathbb{S}}_m = \{M \in \mathbb{S}_m | \tilde{\boldsymbol{O}}_m \leq M \leq Id_m\}$, $\hat{\mathbb{S}}'_m$ (resp. $\bar{\mathbb{S}}'_m$) is the subset of $\hat{\mathbb{S}}_m$ (resp. $\bar{\mathbb{S}}_m$) which eigenvalues are all distinct, and denote by $\pi_{m,p}$ the mapping $M \in \mathcal{SO}_n \mapsto \pi_{m,p}(M) \in \mathcal{M}_{m,n}$ which preserved the upper left corner of M (m first lines and p first columns).

The two following results allow to well define the multidimensional Jacobi process.

Proposition 3.18 ([Dou05])**.** *Let $\boldsymbol{B}$ be a Brownian motion on $\mathcal{SO}_n$. $X = \pi_{m,p}(\boldsymbol{B})$ is a diffusion on $\Pi_{m,p}$ whose infinitesimal generator is $\frac{1}{2}\Delta_{n,m,p}$ where*

$$\Delta_{n,m,p} F = \sum_{\substack{1 \leq i,i' \leq m \\ 1 \leq j,j' \leq p}} \left(\mathbb{1}_{\{i=i',j=j'\}} - X_{i,j'} X_{i',j} \right) \frac{\partial^2 F}{\partial X_{i,j} \partial X_{i',j'}} - (n-1) \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq p}} X_{i,j} \frac{\partial F}{\partial X_{i,j}}$$

for regular enough function F.

Theorem 3.19 ([Dou05])**.** *Let X be the diffusion governed by generator $\frac{1}{2}\Delta_{n,m,p}$ as defined in Proposition 3.18 and define $J = XX^T$. Then J is a diffusion on $\bar{\mathbb{S}}_m$. If $p \geq m+1$, $q \geq m+1$ and $J_0 \in \hat{\mathbb{S}}_m$, then J satisfies the following stochastic differential equation*

$$\mathrm{d}J = \sqrt{J}\mathrm{d}B\sqrt{Id_m - J} + \sqrt{Id_m - J}\mathrm{d}B^T\sqrt{J} + (pId_m - (p+q)J)\,\mathrm{d}t$$

with B a Brownian motion on $\mathcal{M}_{m,m}$ and $q = n - p$. J is called Jacobi process of dimensions (p,q).

This allows to defined matrix Jacobi process with integer dimensions. The extension to the case of arbitrary (p,q) can be done according to the following theorem.

Theorem 3.20 ([Dou05])**.** *Let X be in $\bar{\mathbb{S}}_m$ and consider the following SDE:*

$$\mathrm{d}J = \sqrt{J}\mathrm{d}B\sqrt{Id_m - J} + \sqrt{Id_m - J}\mathrm{d}B^T\sqrt{J} + (pId_m - (p+q)J)\,\mathrm{d}t, \;\; J_0 = X, \tag{3.31}$$

where $(p,q) \in \mathbb{R}^2$, $J \in \bar{\mathbb{S}}_m$ and B a Brownian motion on $\mathcal{M}_{m,m}$.

i. If $\min(p,q) \geq m+1$ and $X \in \hat{\mathbb{S}}_m$, (3.31) has a unique strong solution in $\hat{\mathbb{S}}_m$.

ii. If $\min(p,q) > m-1$ and $X \in \bar{\mathbb{S}}'_m$, (3.31) has a unique solution in law in $\bar{\mathbb{S}}_m$.

iii. If the eigenvalues of J are initially distinct they remain so at any time and can be ordered $\lambda_1 > \cdots > \lambda_m$. They satisfy the following differential equation:

$$\mathrm{d}\lambda_i(t) = 2\sqrt{\lambda_i(t)(1-\lambda_i(t))}\mathrm{d}b_i(t) + \Big\{ p - (p+q)\lambda_i(t) + \sum_{j \neq i} \frac{\lambda_i(t)(1-\lambda_j(t)) + \lambda_j(t)(1-\lambda_i(t))}{\lambda_i(t) - \lambda_j(t)} \Big\} \mathrm{d}t,$$

[2] A square matrix M which satisfies $M^T = -M$ is said to be skew-symmetric

for any $1 \leq i \leq m$ *and where* $b_1, \ldots, b_m$ *are independent real Brownian motions.*

A number of properties on the matrix Jacobi process can be found in [Dou05] and references therein: dynamics of determinant of J, study of boundaries hitting times and local times, derivation of the invariant measure, Girsanov relation or extension to Hermitian matrices.

3.1.9 Time inhomogeneous polynomial diffusions

In the following of this thesis, we will rely on the polynomial property of some stochastic processes to perform Gram-Charlier expansions of their densities. The class of polynomial processes comprises that of affine processes. Indeed for the later one is able to know the characteristic function via the resolution of so-called Ricatti equations while the polynomial property simply offers the ability to compute the moment through matrix exponentiation.

[EP11] introduced the concept of *polynomial jump-diffusions* as solutions of SDEs with possible jumps whose coefficients are of linear growth over the state-space $\mathbb{R}$. Such diffusions are particularly tractable as they allow to readily compute their moments which is necessary in a number of numerical methods; for instance, [EP11] exploited this property to price some double barriers options based on method of moments. [CKRT12] introduced the corresponding *polynomial processes* that are Markov processes by directly characterizing their semi-groups. They provide the same tractability regarding the computation of moments. As solutions of SDEs are non-necessarily Markov, [FL16] proposed to work on *polynomial diffusions* using simply the Itô's formula to establish the main properties of this class of diffusions. They extended the concept of *polynomial diffusions* of [EP11] to more general state-spaces while [FL20] provided a similar extension of the notion of *polynomial jump-diffusions*. To characterize the *polynomial property*, all these works are based on a common idea: the sets of polynomial functions of a given degree are left invariant by the semi-group or infinitesimal generator of the studied processes.

All the mentioned works are set in time-homogeneous frameworks. [dCAH18] extended the framework of [CKRT12] to time-inhomogeneous Markov processes essentially requiring additional continuously differentiable assumptions. In this thesis, we will work with stochastic processes defined as solutions of SDEs whose coefficients are time-varying. This time dependency may be non-regular. Typically, some coefficients will be set piecewise constant. We propose in this section an adaptation of the frameworks of [dCAH18] and [FL16] to account for the time inhomogeneity –possibly non regular– of SDEs. However, our extension is quite straight as we will work in full state-space. While discontinuous processes are quite used in financial modelling (see the standard work [Tan03] on this topic), we will focus in this thesis on continuous diffusions as jumping polynomial processes are little used in practice to this day to our knowledge.

Let $(d, d') \in (\mathbb{N}^*)^2$, $T > 0$ be a finite time horizon, b and σ be two multi-dimensional measurable functions defined over $[0, T] \times \mathbb{R}^d$ and respectively taking values in $\mathbb{R}^d$ and $\mathbb{R}^{d \times d'}$. Let $(W_t)_{0 \leq t \leq T}$ be a d'-dimensional Brownian motion and Y be a squared integrable random variable of $\mathbb{R}^d$: $\mathbb{E}[\|Y\|_d^2] < \infty$. We consider the following d-dimensional SDE:

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \ t \leq T, X_0 \overset{a.s.}{=} Y. \tag{3.32}$$

The regularity in time of coefficients b and σ does not impact the existence/uniqueness of solutions to (3.32) as illustrated in the following result proved in Theorem 5.2.1 of [Øks03]. In the following, we consider a matrix norm $|||\cdot|||$.

Theorem 3.21 ([Øks03])**.** *Assume that* b *and* σ *are of linear growth and Lipschitz in the space*

variable that is: for any $0 \leq t \leq T$, $(x, y) \in (\mathbb{R}^d)^2$,

$$\|b(t,x)\| + |||\sigma(t,x)||| \leq C(1 + \|x\|)$$
$$\|b(t,x) - b(t,y)\| + |||\sigma(t,x) - \sigma(t,y)||| \leq C\|x - y\|$$

for some positive constant C. *Additionally assume that* Y *is independent of the sigma-algebra generated by the Brownian motion* $(W_t)_{0\leq t\leq T}$. *Then (3.32) admits a unique solution whose paths are almost-surely continuous, that is adapted to the sigma-algebra generated by* Y *and* W, $\sigma(Y, (W_s)_{0\leq s\leq t})$, *and squared integrable:*

$$\mathbb{E}\left[\int_0^T \|X_s\|^2 \mathrm{d}s\right] < \infty.$$

The polynomial property we want to characterize is an extension of what [FL16] proposed in a time homogeneous framework. Let us consider the time-dependent infinitesimal generator associated to (3.32): for a regular enough function $f : \mathbb{R}^d \mapsto \mathbb{R}$,

$$\mathcal{A}_t f(x) = b(t,x) \cdot \nabla f(x) + \frac{1}{2}\mathrm{tr}\big((\sigma(t,x)\nabla f(x))(\sigma(t,x)\nabla f(x))^T\big), x \in \mathbb{R}^d, t \in [0,T].$$

Let us denote by $\mathcal{P}_n(\mathbb{R}^d)$ the set of polynomial of degree at most n defined over $\mathbb{R}^d$.

Lemma 3.22. *Consider a family of operators* $\mathcal{G} := \big(\mathcal{G}_t\big)_{t\in[0,T]}$ *of the form* $\mathcal{G}_t = \bar{b}(t,x) \cdot \nabla f(x) + \frac{1}{2}\mathrm{tr}\big((\bar{\sigma}(t,x)\nabla f(x))(\bar{\sigma}(t,x)\nabla f(x))^T\big)$. *The following are equivalent:*

(i) $\mathcal{G}$ *maps* $\mathcal{P}_n(\mathbb{R}^d)$ *to itself for any* $n \in \mathbb{N}$: $\forall t \in [0,T]$, $\mathcal{G}_t(\mathcal{P}_n(\mathbb{R}^d)) \subset \mathcal{P}_n(\mathbb{R}^d)$;

(ii) for any $t \in [0,T]$, $x \mapsto \bar{b}(t,x) \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $x \mapsto \bar{\sigma}(t,x)\bar{\sigma}(t,x)^T \in \mathcal{P}_2(\mathbb{R}^d)$.

Proof. Implication (ii) $\Longrightarrow$ (i) is straight. The reverse implication follows by applying the family $\mathcal{G}$ to monomials of degrees one and two respectively. □

Definition 2. *A solution to (3.32) is said to be a time-inhomogeneous polynomial diffusion if it is associated with an infinitesimal generator that maps* $\mathcal{P}_n(\mathbb{R}^d)$ *to itself for any* $n \in \mathbb{N}$.

The main property associated with polynomial diffusions is the matrix exponential representation of their moments. We present it in a time inhomogeneous setting. Fix $n \in \mathbb{N}$, let N denote the dimension of $\mathcal{P}_n(\mathbb{R}^d)$ and $B_N = (1, b_1(x), \cdots, b_N(x))$ be a basis of $\mathcal{P}_n(\mathbb{R}^d)$. For a polynomial function $p \in \mathcal{P}_n(\mathbb{R}^d)$, there exists a unique vector $\overrightarrow{p}$ such that $p(x) = B_N(x) \cdot \overrightarrow{p}$. At every date $t \in [0,T]$, the operator $\mathcal{A}_t$ is linear and its action on the basis B_N can thus be represented as a unique matrix in $\mathbb{R}^{(N+1)\times(N+1)}$ denoted by A_t that is such that for any $p \in \mathcal{P}_n(\mathbb{R}^d)$, $\mathcal{A}_t p(x) = B_N(x) \cdot A_t \overrightarrow{p}$. Coefficients of A_t are regular functions of the diffusion coefficients b and σ. In particular, the time regularity of $t \mapsto A_t$ is that of $t \mapsto b(t,\cdot)$ and $t \mapsto \sigma(t,\cdot)$. We assume that b and σ are regular enough so that for any $x \in \mathbb{R}^d$, $\int_0^T \Big(\big\|b(s,x)\big\| + |||\sigma(s,x)|||\Big)\mathrm{d}s$ exists and is finite. The set of real matrices being a Lie algebra, the notion of commutator of two matrices will be of interest: for $(M,N) \in \big(\mathbb{R}^{(N+1)\times(N+1)}\big)^2$, we define their commutator as $[M,N] = MN - NM$ and consider the operator $\mathrm{ad}_M : \mathbb{R}^{(N+1)\times(N+1)} \to \mathbb{R}^{(N+1)\times(N+1)}, N \mapsto [M,N]$. We can now state the moment representation result.

Remark 18. *The right superscript* T *for a matrix denotes its transpose: transpose*$(M) = M^T$.

Theorem 3.23. *We assume that (i)* $\mathbb{E}[\|Y\|^{2n}] < \infty$ *and (ii)* $\int_s^t |||A_s|||\mathrm{d}s < \pi$ *for any* $0 \leq s \leq t \leq T$*. Then for any* $p \in \mathcal{P}_n(\mathbb{R}^d)$ *associated with vector representation* $\overrightarrow{p}$*, we have*

$$\mathbb{E}[p(X_t)|\mathcal{F}_s] = B_N(X_s) \cdot \exp\left(M^{A^T}(s,t)\right)^T \overrightarrow{p}, \tag{3.33}$$

for any $0 \leq s \leq t \leq T$*, with* $M(s,t)$ *being a matrix that can be expressed as the sum of an absolutely convergent power series*

$$M^{A^T}(s,t) = \sum_{k=1}^{+\infty} M_k^{A^T}(s,t),$$

whose coefficients depend on s, t *and* $(A_u)_{u\in[0,T]}$*. For any time dependent matrix* $(N_t)_{0\leq t\leq T}$*, the matrices* M_k^N *can be recursively expressed as*

$$M_1^N(s,t) = \int_s^t N_u \mathrm{d}u,$$
$$M_k^N(s,t) = \sum_{j=1}^{k-1} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=n-1 \\ k_1\geq 1,\ldots,k_j\geq 1}} \int_s^t \mathrm{ad}_{M_{k_1}^N(s,u)} \circ \cdots \circ \mathrm{ad}_{M_{k_j}^N(s,u)}(N_u)\mathrm{d}u, \ k \geq 2,$$

where $(B_n)_{n\in\mathbb{N}}$ *are the Bernoulli numbers.*

The proof relies on the use of Itô's formula as depicted in [FL16] in a time-independent framework. We adapt their proof for taken into account time dependency that involves Magnus series (exponential of matrix integrals) as explained in [dCAH18].

Proof. The polynomial property allows to show a well-known results for diffusion with Lipschitz coefficients.

Lemma 3.24. *Let* $(X_t)_{0\leq t\leq T}$ *be a time-inhomogeneous polynomial diffusion associated with family of infinitesimal generators* $(\mathcal{G}_t)_{0\leq t\leq T}$ *and starting from* $X_0 \overset{a.s.}{=} Y$*. Then, for any* $k \in \mathbb{N}$ *such that* $\mathbb{E}[\|Y\|^{2k}] < \infty$*, there exists a finite constant* $C > 0$ *such that*

$$\mathbb{E}[1 + \|X_t\|^{2k}] \leq (1 + \mathbb{E}[\|Y\|^{2k}])e^{Ct}, \ t \geq 0.$$

Proof of lemma. Let τ_m be the first exit time of the ball of radius m: $\tau_m = \inf\{t \geq 0 : \|X_t\| \geq m\}$ for $m \geq 0$. The polynomial property of $(X_t)_{t\leq T}$ implies that for $f(x) = 1 + \|x\|^{2k} \in \mathcal{P}_n(\mathbb{R}^d)$, there exists a positive constant C such that $|\mathcal{G}_t f(x)| \leq Cf(x)$ for any time t. Furthermore, Itô's lemma gives that

$$\mathbb{E}[f(X_{t\wedge\tau_m})] = \mathbb{E}[f(Y)] + \mathbb{E}\left[\int_0^{t\wedge\tau_m} G_s f(X_s)\mathrm{d}s + \int_0^{t\wedge\tau_m} \nabla f(X_s) \cdot \left(\sigma(s, X_s)\mathrm{d}W_s\right)\right]$$

By locating the process in the ball $\{x \in \mathbb{R}^d : \|x\| \leq m\}$, we ensure that the Itô's term is of

expectation zero. Then, with the previous bounding we pursue the computations to get that:

$$\begin{aligned}\mathbb{E}[f(X_{t\wedge\tau_m})] &= \mathbb{E}[f(Y)] + \mathbb{E}\left[\int_0^{t\wedge\tau_m} A_s f(X_s)\mathrm{d}s\right] \\ &\leq \mathbb{E}[f(Y)] + C\mathbb{E}\left[\int_0^{t\wedge\tau_m} f(X_s)\mathrm{d}s\right] \\ &\leq \mathbb{E}[f(Y)] + C\mathbb{E}\left[\int_0^{t} f(X_{s\wedge\tau_m})\mathrm{d}s\right].\end{aligned}$$

Gronwall's lemma provides that $\mathbb{E}[f(X_{t\wedge\tau_m})] \leq \mathbb{E}[f(Y)]e^{Ct}$ where the right hand side does not depend on m: sending it to infinity yields the result. □

We come back to the proof of the Theorem. Itô's formula yields

$$p(X_t) = p(X_s) + \int_s^t \mathcal{A}_u p(X_u)\mathrm{d}u + \int_s^t \nabla p(X_u)\cdot\big(\sigma(u,X_u)\mathrm{d}W_u\big). \tag{3.34}$$

The quadratic variation of the Itô's integral satisfies $\int_s^t \big(\nabla p(X_u)\big(\sigma(u,X_u)\big)\big(\nabla p(X_u)\big(\sigma(u,X_u)\big)^T\mathrm{d}u \leq C\int_s^t(1+\|X_u\|^{2n})\mathrm{d}u$ for some positive constant C. Previous lemma proves that it has finite expectation and thus Itô's integral in (3.34) is of expectation zero. Vector representation $\overrightarrow{p}$ of p in the basis B_N and matrix one A_s of $\mathcal{A}_s$, Fubini's theorem and the linearity of the expectation yield

$$\begin{aligned}\mathbb{E}[B_N(X_t)|\mathcal{F}_s]\cdot\overrightarrow{p} &= \mathbb{E}[B_N(X_s)|\mathcal{F}_s]\cdot\overrightarrow{p} + \int_s^t \mathbb{E}\left[B_N(X_u)\cdot\big(A_u\overrightarrow{p}\big)|\mathcal{F}_s\right]\mathrm{d}u \\ &= B_N(X_s)\cdot\overrightarrow{p} + \int_s^t \mathbb{E}\left[B_N(X_u)|\mathcal{F}_s\right]\cdot\big(A_u\overrightarrow{p}\big)\mathrm{d}u \\ &= B_N(X_s)\cdot\overrightarrow{p} + \int_s^t A_u^T\mathbb{E}\left[B_N(X_u)|\mathcal{F}_s\right]\mathrm{d}u\cdot\overrightarrow{p}.\end{aligned}$$

The vector function $F(u) = \mathbb{E}[B_N(X_u)|\mathcal{F}_s]$ satisfies thus a linear system of equations whose coefficients are time dependent. Its unique solution is given by (see Theorem 9 in [BCOR09] for instance):

$$F(u) = \exp\big(M^{A^T}(s,u)\big)B_N(X_s).$$

Then,
$\mathbb{E}[p(X_t)|\mathcal{F}_s] = \mathbb{E}[B_N(X_t)|\mathcal{F}_s]\cdot\overrightarrow{p} = \exp\big(M^{A^T}(s,u)\big)B_N(X_s)\cdot\overrightarrow{p} = B_N(X_s)\cdot\exp\big(M^{A^T}(s,u)\big)^T\overrightarrow{p}.$
□

Remark 19. *Observe that in the previous theorem, the integrability condition (ii) is used in the Theorem 9 of [BCOR09] to ensure the convergence of matrix exponential series. Note that it is a sufficient condition for convergence.*

Remark 20. *The required regularity conditions with respect to time of the coefficients are mild.*

Remark 21. *The previous theorem is hardly exploitable for practical uses. Some particular cases should be mentioned.*

- *If the family of generator is commuting meaning that for any couples of dates $(s,t) \in [0,T]^2$, $A_sA_t = A_tA_s$, then $[A_t, A_s] = \mathbf{0}_d$ and all the terms of the Magnus series are zero but the first. Combined with the fact that transpose of an exponential of matrix is the*

exponential of the transpose matrix, the Equation (3.33) writes now, for any $0 \leq s \leq t \leq T$*:*

$$\mathbb{E}[p(X_t)|\mathcal{F}_s] = B_N(X_s) \cdot \exp\left(\int_s^t A_u^T \mathrm{d}u\right)^T \overrightarrow{p} = B_N(X_s) \cdot \exp\left(\int_s^t A_u \mathrm{d}u\right) \overrightarrow{p}.$$

- *In the time homogeneous framework* $A_s \equiv A$ *and we recover the expression in Theorem 3.1 of [FL16]:*

$$\mathbb{E}[p(X_t)|\mathcal{F}_s] = B_N(X_s) \cdot \exp\left((t-s)A^T\right)^T \overrightarrow{p} = B_N(X_s) \cdot \exp\left((t-s)A\right) \overrightarrow{p}.$$

  Note that is such setting, there is no need of an integrability condition to ensure the convergence of the matrix exponential.

- *In a piecewise constant setting (also discussed in [BCOR09]), an equivalent representation but more understandable can be proposed. It has been demonstrated in more details in Proposition 1 of [AMLB20] using successive conditionings. We briefly recall it. Let us consider a time structure* $0 \leq t_1 \leq \cdots \leq t_n \leq T$ *such that the infinitesimal generator is piecewise constant on this grid:* $\mathcal{A}_t f(x) = \mathcal{A}_{t_k} f(x)$ *for all* $t \in [t_k, t_{k+1})$*, for* $k = 0, \ldots, n$*, with* $t_0 = 0$ *and* t_{n+1}*. In addition for* $0 \leq t \leq T$*, define* $\bar{k}(t) := \max\{k : t_k \leq t\}$ *and* $\underline{k}(t) := \min\{k : t_k \geq t\}$*. The alternative polynomial moments formula yields for any* $0 \leq s \leq t \leq T$*:*

$$\mathbb{E}\left[p(X_t)\middle|\mathcal{F}_s\right] = B_N(X_s) \cdot e^{(t_{\underline{k}(s)}-s)A_{t_{\underline{k}(s)-1}}} \left(\prod_{k=\underline{k}(s)}^{\bar{k}(t)-1} e^{(t_{k+1}-t_k)A_{t_k}}\right) e^{(t-t_{\bar{k}(t)})A_{t_{\bar{k}(t)}}} \overrightarrow{p}, \quad (3.35)$$

  where we emphasize that the order in the matrix product matters as matrices $(A_{t_k})_{k=1,\ldots,n}$ *do not commute in general. We stress that this representation is equivalent to the one proposed here in Theorem 3.23: to prove it, the Baker–Campbell–Hausdorff formula (see for instance [Hal15]) can be employed. Again, the convergence of the matrix exponential is obtained without additional constraint on each time interval.*

We end this section with some illustrative examples partly drawn from [dCAH18].

Brownian motion with drift
The process

$$X_t = \int_0^t b(u)\mathrm{d}u + W_t, \ t \leq T,$$

where b is a (piecewise) continuous function is a time-inhomogeneous polynomial diffusion. It is straight to get that the family of its infinitesimal generators commutes and some further computations are made in [dCAH18].

Time-dependent CIR
The time dependent Cox-Ingersoll-Ross process

$$X_t^x = x + \int_0^t \kappa\big(\theta(s) - \xi(s)X_s^x\big)\mathrm{d}s + \int_0^t \epsilon(s)\sqrt{X_s^x}\mathrm{d}W_s, \ t \leq T,$$

as defined in [BGM10] forms a time-inhomogeneous polynomial diffusion. Its infinitesimal generator at time t is given by

$$\mathcal{A}_t f(x) = \kappa(\theta(t) - \xi(t)x)\frac{\mathrm{d}f}{\mathrm{d}x} + \frac{1}{2}\epsilon^2(t)x\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}.$$

Taking in particular $f(x) = x$, for any (s,t) we get:

$$\begin{aligned}\mathcal{A}_t\big(\mathcal{A}_s f(x)\big) &= \kappa\big(\theta(t) - \kappa\xi(t)(\theta(s) - \xi(s)x)\big)\\ \mathcal{A}_s\big(\mathcal{A}_t f(x)\big) &= \kappa\big(\theta(s) - \kappa\xi(s)(\theta(t) - \xi(t)x)\big)\end{aligned}$$

showing that the associated family of matrices will not commute.

3.2 The problem of moments

In this thesis, we are particularly interested in characterizing distributions based on their moments. It is well known that different distributions can share the same sequence of moments. The problem of identifying a distribution using the sequence of its moments is known in mathematics as the «moment problem». This question dates back to 1894-1895 and the work [Sti94] in which Thomas Joannes Stieltjes introduced[3] «*le problème des moments*». The context is the following: let $I \subset \mathbb{R}$ be associated with a σ-algebra $\sigma(I)$ and consider a given sequence of real numbers $(\lambda_n)_{n\in\mathbb{N}} \in I^{\mathbb{N}}$. The problem corresponds to the finding of conditions ensuring existence and/or uniqueness of a measure μ defined over $\sigma(I)$ being such that

$$\int_I x^n \mu(\mathrm{d}x) = \lambda_n,\ n \in \mathbb{N}.$$

When focusing on probability measures (i.e. distributions), one must additionally require $\lambda_0 = 1$. If uniqueness is ensured, the moments problem is said to be *determined*: knowing the moments or knowing the measure is equivalent. Differently said: a measure determinate by its moments is the only one having such moments. Conversely, when the moments problem is not determined or undetermined, at least two distinct measures have the same sequence of moments (as long as existence is provided). Different conditions and solutions have been derived depending on the set I. In the literature, the problem of moments is said to be (i) of Hausdorff if $I = [0,1]$ –or more generally, if I is a compact interval– named after Felix Hausdorff (see [Hau23]); (ii) of Hamburger if $I = \mathbb{R}$ named after Hans Hamburger (see for instance [Ham20] or [Ham44]); (iii) of Stieltjes if $I = \mathbb{R}_+$ named after Thomas Joannes Stieltjes (as originally introduced in [Sti94]).
Note that the moments problem is well known and studied in the literature in one-dimensional space as illustrated by the above quoted works. The multiple dimensional case has been later on studied and some questions are still open (see [BL12] and references therein).

We refer the interested reader to [Sch17] for a late thorough survey on the moment problem. [ST43] and references therein can also be consulted. We also refer to Section 11 of [Sto13] for instructive examples –some are presented below– on the topic. We present now some of the main results on the one-dimensional moments problem.

Determinacy problem The first result we provide is a well known sufficient condition for the determinacy of moment problem proved in [Pat95] (Theorem 30.1).

[3] To our knowledge, he was the first to formulate the problem in those terms.

Theorem 3.25 ([Pat95]). *Let X be a random variable taking values in $\mathbb{R}$ and whose probability measure is μ is of finite moments of any order: $\forall n \in \mathbb{N}$,*

$$m_n = \mathbb{E}\left[X^n\right] = \int_{\mathbb{R}} x^n \mu(\mathrm{d}x) < +\infty.$$

If the power series (of Laplace)

$$\sum_{n\geq 0} \frac{m_n}{n!} z^n$$

has a positive radius of convergence, then μ is determined by its moments.

This result is quite intuitive since the involved power series is nothing more than the series expansion of the moment generating function of the random variable X: $\mathbb{E}\left[e^{zX}\right] = \sum_{n\geq 0} \frac{m_n}{n!} z^n$ whenever this expression is well defined. The sequence $(m_n)_{n\geq 0}$ characterizes thus the moment generating function of X. However, it is well known (see for instance [Fel], Theorem 1 in Chapter XIII) that the moment generating function characterizes a probability distribution.

Remark 22. *Let X be a random variable taking values over a bounded interval of $\mathbb{R}$: $X \sim \mu(\mathrm{d}x)$ where μ is defined over $\sigma\big((a,b)\big)$ where $-\infty < a < b < \infty$. Then $|m_n| \leq \max(|a|^n, |b|^n)$ for all $n \in \mathbb{N}$ and thus the Laplace series has a positive radius of convergence: a random variable with bounded support is always determined by its moments. In other words, the Hausdorff formulation is always determined.*

Example 1.

i. Let $X \sim \mathcal{N}(0, \sigma^2)$ with $\sigma > 0$. Then $\sum_{n\geq 0} \frac{\mathbb{E}[X^n]}{n!} z^n = \sum_{n\geq 0} \frac{\sigma^{2n} z^{2n}}{n!} = e^{\sigma^2 z^2/2}$ has an infinite radius of convergence: the Gaussian distribution is determined by its moments.

ii. Let $X \sim \Gamma(a,k)$ with $(a,k) \in (\mathbb{R}_+^*)^2$. Then $\mathbb{E}\left[e^{zX}\right] = \left(1 - \frac{z}{k}\right)^{-a}$ for $0 < z < k$: the gamma distribution is determined by its moments.

iii. Let $Y \sim \mathcal{N}(0,1)$ and set $Y = X^3$. Then, the density function of Y writes $f(x) = \frac{1}{3\sqrt{2\pi}} |x|^{-2/3} \exp\left(-\frac{1}{2}|x|^{2/3}\right)$ and thus the Laplace series has a null convergence radius: we can not conclude on the determinacy of Y by its moments.

As long as the moments are known, the following result provides a famous practical criterion, often named Carleman's condition after the work of Torsten Carleman in [Car]. A proof can be found in [Akh20] or in [ST43].

Theorem 3.26 ([Car]). *Denote the moment of the studied measure μ by $m_n = \int_{\mathbb{R}} x^n \mu(\mathrm{d}x)$ for all $n \in \mathbb{N}$. Assume that for all $n \in \mathbb{N}$, $m_n < \infty$.*

i. (Hamburger) Suppose that μ is defined over $\sigma(\mathbb{R})$. If

$$\sum_{n\geq 1} m_{2n}^{-1/2n} = +\infty,$$

then μ is determined by its moments.

ii. (Stieltjes) Suppose that μ is defined over $\sigma(\mathbb{R}_+)$. If

$$\sum_{n\geq 1} m_n^{-1/2n} = +\infty,$$

then μ is determined by its moments.

We now focus on case of measures having Radon–Nikodym densities with respect to the Lebesgue measure. The following result is a consequence of previous Theorem 3.26 and shows in particular that measures having exponential moments are determined by their moments. It can be viewed as an extension of Theorem 3.25. A proof can be found in [ST43].

Corollary 1 ([ST43]). *Let be μ a measure whose density with respect to Lebesgue measure is m: $\mu(\mathrm{d}x) = m(x)\mathrm{d}x$.*

i. (Hamburger) If there exists $p \geq 1$ and $\eta > 0$ such that

$$\int_{\mathbb{R}} m(x)^p e^{\eta|x|}\mathrm{d}x < +\infty,$$

then μ is determined by its moments.

ii. (Stieltjes) If there exists $p \geq 1$ and $\eta > 0$ such that

$$\int_{\mathbb{R}_+} m(x)^p e^{\eta\sqrt{|x|}}\mathrm{d}x < +\infty,$$

then μ is determined by its moments.

Finally, we provide some determinacy conditions of the moments problem when considering positive measure. Whether some integrability condition is satisfied by the logarithm of the density function –named Krein's condition after the work of Mark Grigorievich Krein [Kre44]– or not, the determinacy of the problem can be addressed. For the following statements, we refer to [S+00] and [Ped98] and references therein for more details.

Theorem 3.27. *Let μ be a measure having a positive density m with respect to the Lebesgue measure: $\mu(\mathrm{d}x) = m(x)\mathrm{d}x$, and $m(x) > 0$ for all $x \geq 0$.*

i. (Hamburger) If

$$\int_{\mathbb{R}} -\frac{\ln\big(m(u)\big)}{1+u^2}\mathrm{d}u < +\infty,$$

then μ is not determined *by its moments.*

ii. (Hamburger) If

$$\int_{\mathbb{R}} -\frac{\ln\big(m(u)\big)}{1+u^2}\mathrm{d}u = +\infty,$$

and if in addition, m is: symmetric, differentiable and for some $x_0 > 0$ and $x \geq x_0$, $-\frac{xm'(x)}{m(x)} \nearrow +\infty$ as $x \to +\infty$, then μ is determined *by its moments.*

iii. (Stieltjes) If

$$\int_{\mathbb{R}_+} -\frac{\ln\big(m(u^2)\big)}{1+u^2}\mathrm{d}u < +\infty,$$

then μ is not determined *by its moments.*

iv. (Stieltjes) If

$$\int_{\mathbb{R}_+} -\frac{\ln\big(m(u^2)\big)}{1+u^2}\mathrm{d}u = +\infty,$$

and if in addition, m is differentiable and for some $x_0 > 0$ and $x \geq x_0$, $-\frac{xm'(x)}{m(x)} \nearrow +\infty$ as $x \to +\infty$, then μ is determined *by its moments.*

Example 2. *The log-normal distribution is not determined by its moments. Let $Y \sim \mathcal{N}(0,1)$ so that $X = e^Y$ is log-normally distributed. Denote by $m(x) = \frac{e^{-\ln(x)^2/2}}{x\sqrt{2\pi}}\Big(1+\sin\big(\pi\ln(x)\big)\Big)$: one can check that $\mu(\mathrm{d}x) = m(x)\mathrm{d}x$ is a probability measure and that if $\tilde{X}$ is distributed according to this distribution, then for any $n \in \mathbb{N}$, $\mathbb{E}\big[X^n\big] = \mathbb{E}\big[\tilde{X}^n\big]$. Yet $\mathbb{P}\big(X \in [1,e]\big) - \mathbb{P}\big(\tilde{X} \in [1,e]\big) = \frac{1}{\sqrt{2\pi}}\int_1^e \frac{e^{-\ln(x)^2/2}}{x}\sin\big(\pi\ln(x)\big)\mathrm{d}x \geq \frac{1}{\sqrt{2\pi}}\int_1^e \sin\big(\pi\ln(x)\big)\mathrm{d}x = \frac{(1+e)\pi}{(1+\pi^2)\sqrt{2\pi}} > 0$ so that X and $\tilde{X}$ do not have the same law.*

Existence problem While the determinacy of Hausdorff's formulation is straight to obtain (see proof of Theorem 3.25), the existence result is «*a deep and powerful result*» according to William Feller (p.226 in [Fel]). The corresponding theorem is due to F. Hausdorff and a proof can be found in [Fel] pp.255-227.

Theorem 3.28. *Let $(m_n)_{n\in\mathbb{N}}$ be a completely monotone sequence*[4] *of real numbers such that $m_0 = 1$. Then there exists a unique probability measure μ defined over $\sigma([0,1])$ with moments $(m_n)_{n\in\mathbb{N}}$.*

Regarding the Hamburger and Stieltjes formulations we provide the following result which proof can be found in [RS75].

Theorem 3.29. *Let $(m_n)_{n\in\mathbb{N}}$ be a sequence of real numbers.*

1. *(Hamburger) There exists a non-negative measure μ defined over $\sigma(\mathbb{R})$ whose moments are $(m_n)_{n\in\mathbb{N}}$ if and only if, for all $n \in \mathbb{N}$, all $(z_0,\dots,z_n) \in \mathbb{C}^{n+1}$,*

$$\sum_{p,k=0}^{n} z_k\bar{z}_p m_{k+p} \geq 0.$$

2. *(Stieltjes) There exists a non-negative measure μ defined over $\sigma(\mathbb{R}_+)$ whose moments are $(m_n)_{n\in\mathbb{N}}$ if and only if, for all $n \in \mathbb{N}$, all $(z_0,\dots,z_n) \in \mathbb{C}^{n+1}$,*

$$\sum_{p,k=0}^{n} z_k\bar{z}_p m_{k+p} \geq 0 \text{ and } \sum_{p,k=0}^{n} z_k\bar{z}_p m_{k+p+1} \geq 0.$$

In this manuscript, we will approximate unknown probability measures using polynomial expansions whose coefficients write as linear combinations of the moments of the approximated distribution. When approximating a density function that is not determined by its moments, the mentioned expansions may diverge at best or approximate an improper density function at worst. Those polynomial expansions are presented and illustrated in next section.

3.3 Series expansion and orthogonal polynomials: a statistical point of view

The polynomial expansions we will consider involve the Gaussian distribution. In the early of the 18th, Abraham De Moivre was the first to deal with the normal density without genuinely knowing it. His book [DM56] is one of the oldest work on games of chance. By modelling

[4] Let $a = (a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers and define the difference operator recursively as $\delta^{(0)}(a) = a$, $\delta(a) = (a_{n+1} - a_n)_{n\in\mathbb{N}}$ and $\delta^{(p+1)}(a) = \delta\big(\delta^{(p)}(a)\big)$ for $p \in \mathbb{N}^*$. The sequence a is said to be completely monotone if $\forall (k,p) \in \mathbb{N}^2$ $(-1)^p\delta^{(p)}(a)_k \geq 0$.

gamble games, he was led to study the binomial distribution. He observed that increasing the number of Bernoulli realizations results in the convergence of the binomial distribution towards a symmetric and continuous one with thin tailed. In that respect De Moivre was the first to establish a particular form of the Central Limit Theorem (CLT) which general statement was provided later by Pierre-Simon de Laplace in [Lap10]. The asymptotic distribution was later rigorously introduced as the *normal density* independently by Robert Adrain in [Adr14] and Carl Friedrich Gauss in [Gau09]. They both worked on the Least Square Method (introduced in 1805 by Adrien-Marie Legendre in [Leg05]) and were led to somehow describe the distribution of the residues: that is the birth of the normal distribution –also often referred to as the Gaussian distribution.

The expansion series we are interested in precisely are the Gram-Charlier and the Edgeworth expansions. They are highly related to the normal distribution since they are designed to approximate density functions that are close (in some sense) to the density function of the normal distribution ϕ. Both approximating series can be introduced in a unified framework, as it has historically been the case. In the next subsection we roughly recap the history of Gram-Charlier and Edgeworth approximating series following the authoritative work of [Hal00]. According to [Hal00], polynomials series approximations appear independently in three different research fields: first one is relative to the generalization of the CLT; second one is about the fitting of polynomial regressions using least square method; third one relates to the establishment of new distributions defined as perturbations the normal one.
Let us also mention the authoritative work of Harald Cramér often quoted on the history of approximating polynomial series (see in particular [Cra28], [Cra46] or [Cra04]). Expansions series have been atfirst severely criticised by mathematicians because of their lack of rigorous theoretical justifications: as pointed out by Cramér in [Cra28] (p. 22), the «*mathematical foundations hitherto laid down seem to be rather weak*». It has become progressively accepted by academics during the first half of the 19th after numerous works of mathematicians (including that of H. Cramér and other) to establish a rigorous and precise framework. In Subsections 3.3.2 and 3.3.3, we go into technical details of the building of such approximating series.

3.3.1 History and general idea

This paragraph is widely inspired by [Hal00].

3.3.1.1 Around the extension of the Central Limit Theorem

The CLT claims that the normalized empirical sum of any independent identically distributed (i.i.d.) square integrable random variables converges towards a Gaussian distribution as the number of realizations increases. It additionally stipulates that the distribution convergence occurs with a speed rate proportional to the square root of the number of events. The mentioned extension of the CLT concerns this last point. The characteristic function of a sum of i.i.d. variates writes as the product of characteristic functions of the variables appearing in the sum. When sending the number of term to the infinity, the leading term of the product converges towards the characteristic function of the Gaussian distribution as the number of terms in the sum goes to infinity. Yet, additional terms in the Taylor expansion of the product characteristic function can be taken into account to specify the behaviour of the remainder. One can then apply the Fourier transform to the obtained Taylor series to recover an approximation of the density function of the distribution of the normalized sum of the variates. It turns out that the later expresses as the density function of the standard normal distribution plus additional corrective terms. The development of this approach took place through successive works in the

first half of the 19th century led by Pierre-Simon de Laplace, Siméon Denis Poisson, Irénée-Jules Bienaymé or also Friedrich Wilhelm Bessel.

Let us briefly give some insights on the described extension, before giving more details in Section 3.3.3. Note that the central idea here is that of P.-S. Laplace in its proof of the Central Limit Theorem. Let $(X_i)_{1\leq i\leq n}$ be i.i.d. random variables with mean μ and standard deviation σ. Let us consider the standardized sum $S_n = \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^n X_i - n\mu\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$ where we set $Y_i = (X_i - \mu)/\sigma$. With ψ_{Y_1} and ψ_{S_n} being the respective characteristic functions of the identically distributed variates $(Y_i)_{1\leq i\leq n}$ and S_n, it is straightforward to get:

$$\psi_{S_n}(t) = \psi_{Y_1}\left(t/\sqrt{n}\right)^n. \tag{3.36}$$

Let now $(\kappa_n)_{n\in\mathbb{N}^*}$ be the sequence of the cumulants of the $(Y_i)_{1\leq i\leq n}$. They are defined as the coefficient appearing in the series expression of the log-characteristic function of Y_1:

$$\ln\psi_{Y_1}(t) = \sum_{n=1}^{+\infty}\frac{\kappa_n(it)^n}{n!}. \tag{3.37}$$

Y_1 being standardized, $\kappa_1 = 0$ and $\kappa_2 = 1$. Combining then (3.36) and exponential form of (3.37), we get

$$\psi_{S_n}(t) = \exp\left(-\frac{t^2}{2}\right)\exp\left(-\frac{i\kappa_3 t^3}{6\sqrt{n}} + \frac{\kappa_4 t^4}{24n} + \frac{i\kappa_5 t^5}{120n^{3/2}} + \dots\right).$$

A Taylor expansion of the exponential function as n goes to infinity along with the application of an inversion theorem allows to identify the density function of S_n as

$$f_{S_n}(x) = \frac{\phi(x)}{\sqrt{n}}\left(1 + \frac{\kappa_3 H_3(x)}{3!\sqrt{n}} + \frac{\kappa_4 H_4(x)}{4!n} + \frac{\kappa_5 H_5(x)}{5!n^{3/2}} + \frac{1}{6!}\left(\frac{\kappa_6}{n^2} + 10\frac{\kappa_3^2}{n}\right)H_6(x) + \dots\right). \tag{3.38}$$

where the $\big(x \mapsto H_n(x)\big)_{n\in\mathbb{N}}$ are polynomial functions known as the Hermite polynomials (see Section 3.3.2). This is the Gram-Charlier expansion of f_{S_n}. When alternatively ordering the terms with respect to increasing powers of $1/n$ one obtains the *so-called* Edgeworth expansion of f_{S_n}.

We refer the interested reader to [Cra46] or in [Hal13] and references therein for more details on the topic. [Cra46] provides a thorough survey of mathematical background used in probability and pays thus a special attention to the normal distribution and its extension; [Hal13] works from a statistical point of view to introduce the Edgeworth expansion.

3.3.1.2 Around the least square polynomial fitting

We now move onto the second point of view indicated in [Hal00]: polynomial fitting via least square method. The development of this approach is essentially due to Augustin Louis Cauchy (between 1835 and 1853; see [Hal98] for a survey on the work of Cauchy) who work on the polynomial regression as extension of the linear model and, independently, to Pafnuty Tchebychev (1855 and beyond, see for instance [Che58]) who apply the least square approach to the polynomial fitting. Let $n+1$ observations $f(x_0), \dots, f(x_n)$ be given. Tchebychev proposed to model this data by fitting a polynomial of degree $m < n$:

$$f(x_i) = a_0 h_0(x_i) + a_1 h_1(x_i) + \dots + a_n h_m(x_i) + \epsilon_i,\ \ i = 1, \dots, n$$

where $(\epsilon_i)_{1\leq i\leq n}$ is a sequence of random variables (white noise) and $(h_i)_{0\leq i\leq n}$ is a sequence of appropriate polynomials with h_j being of degree j ($\deg(h_j) = j$). Let us denote by $\omega(x_i)$ the inverse of the variance associated to the observation $f(x_i)$. In Tchebychev's paper [Che58] on the continued fractions, he get the existence of a set of orthogonal polynomials $(\tilde{h}_i)_{0\leq i\leq n}$ with $\deg(\tilde{h}_j) = j$, satisfying $\sum_{i=0}^{n} \tilde{h}_k(x_i)\tilde{h}_m(x_i)\omega(x_i) = 0,\ k = 1, \ldots, m-1,\ h_0 \equiv 1$ and so that the solution of least square approximation can write as:

$$f_m(x) = \sum_{k=0}^{m} a_k h_k(x),$$

where the coefficients express as

$$a_k = \frac{\sum_{i=0}^{n} f(x_i)h_k(x_i)\omega(x_i)}{\sum_{i=0}^{n}\sum_{i=0}^{n} h_k^2(x_i)\omega(x_i)}.$$

Later in [Che59], Tchebychev extends these results to the continuous case and replaced, formally speaking, all sums by integrals. The choice of the weighting function $\mathbb{R} \ni x \mapsto \omega(x)$ is key. All the standard series representation commonly used in analysis (Fourier, Legendre, Laguerre, Hermite or MacLaurin-Taylor) can be recovered for an appropriate choice of the weighting function. When representing the distribution of the errors on measure by a Gaussian one, i.e. choosing $\omega = \phi$, it yields Gram-Charlier (and Edgeworth) series expansions (see for instance [Che59]). For a function f satisfying some integrable conditions, Tchebychev writes

$$f(x) = \sum_{n=0}^{+\infty} \frac{1}{n!} H_n(x) \int_{\mathbb{R}} H_n(u)\phi(u)f(u)\mathrm{d}u.$$

where the $(H_n)_{n\in\mathbb{N}}$ are the Hermite polynomials previously mentioned. For a probability measure whose density function writes as a product $g(x) = \phi(x)f(x)$, one eventually gets

$$g(x) = \phi(x)\left(\sum_{n=0}^{+\infty} \frac{1}{n!}\mathbb{E}\left[H_n(X)\right] H_n(x)\right) \tag{3.39}$$

where X is a random variable distributed according to $g(x)\mathrm{d}x$. The series (3.39) is the Gram-Charlier expansion of g. Note that according to A. Hald, Tchebychev did not notice that the expression he gets coincides with Gram-Charlier expansion at first.

3.3.1.3 As a disturbed probability distribution

The normal distribution is a «central» one, adjective first employed by George Polya in [Pól20] that motivates the name still currently given to Laplace's theorem. When it comes to approximate an unknown distribution function of some variate, as it is done when fitting some observed data, the normal distribution is a natural choice. Yet obviously the Gaussian distribution does not allow to fit every observed data distributions (economic, demographic, etc.). According to [Hal00], Ludvig Henrik Ferdinand Oppermann first had the idea to fit data using a perturbation of the normal density by multiplying it by a polynomial function and to determine the coefficients of the polynomial using the classic method of moments. He did not publish on the topic himself but its colleagues Jørgen Pedersen Gram and notably Thorvald Nicolai Thiele did.

In [Thi73], Thiele writes the density he proposed to model n given observations writes

$$g(x) = \phi(x) + c_1\phi'(x) + \cdots + c_n\phi^{(n)} + \cdots = \phi(x)\sum_{n=0}^{+\infty}\frac{c_k}{k!}H_k(x).$$

with the $(H_k)_{k\in\mathbb{N}}$ still denoting the Hermite polynomials. His work essentially consists in expressing the coefficients c_n in terms of the moments of the variable of which realization have been observed.

To this end, Thiele introduced the standardized variable $\frac{x-\mu_n}{\sigma_n}$ with μ_n and σ_n the respective empirical mean and empirical standard deviation of the data set of size n . If $(g_i)_{1\leq i\leq n}$ denotes the n observations, Thiele's model consists in representing those data as

$$g_i \approx g\left(\frac{x_i-\mu_n}{\sigma_n}\right) = \phi\left(\frac{x_i-\mu_n}{\sigma_n}\right)\sum_{k=0}^{+\infty}\frac{c_k}{k!}H_k\left(\frac{x_i-\mu_n}{\sigma_n}\right).$$

Using the least square method, he estimated the coefficients of the series as $c_m = \sum_{1\leq i\leq n} H_m\left(\frac{x_i-\mu_n}{\sigma_n}\right)g_i \approx \int H_m(\frac{x-\mu_n}{\sigma_n})g(\frac{x-\mu_n}{\sigma_n})\mathrm{d}x = \mathbb{E}\left[H_m\left(\frac{X-\mu}{\sigma}\right)\right]$. Using then the closed-form expression of the Hermite polynomials H_n and the relationship linking cumulants and moments, he finally derived in [Thi89] the first simple recurrence formula for the computation of the coefficients of the Gram-Charlier series as:

$$c_k + (k-1)c_{k-2} = \sum_{j=0}^{k-1}\binom{k-1}{j}c_{k-1-j}\kappa_{j+1},\ k\in\mathbb{N}^*$$

where the $(\kappa_i)_{i\in\mathbb{N}}$ are the cumulants of the observations that express in term of the moments (and vice-versa).

3.3.2 Gram-Charlier type A expansions

Let us now define quantities and concepts more precisely beginning with the Gram-Charlier expansion named after the mathematician J. P. Gram and the astronomer Carl Charlier. It is an expansion method designed for approximating density functions that are somehow close to the standard normal density. The approach considered in this paragraph is sometimes referred to as the Gram-Charlier *type A* expansion to discriminate from the little known *type B* series Charlier did also consider (see comments p. 23 in [Cra28]). In this manuscript, we only consider *type A* expansion and we will not systematically mention it.

In the following, f will denote an unknown density function that is aim at being approximate through expansion method; X will be a random variable whose density function (with respect to the Lebesgue measure) is f: for all $x\in\mathbb{R}$, we have $\mathbb{P}(X\leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$. Recall that ϕ denotes the standard normal density function and that Φ is the associated cumulative function: $\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ and $\Phi' = \phi$. Let us introduce more precisely the Hermite polynomials. Those are polynomial functions named after the mathematician Charles Hermite and well studied by the P. Tchebychev explaining why they are sometimes also named Tchebychev-Hermite polynomials[5]. Let us denote them by $(H_n)_{n\in\mathbb{N}}$: the n-th Hermite polynomial is of degree n. They can be

[5] According to [Mol30] and [Usp37], it is actually Laplace who was the first to operate on these polynomials.

defined using the Rodrigues[6] formulation as satisfying

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2} = (-1)^n H_n(x) e^{-x^2/2}, \ x \in \mathbb{R}, \ n \in \mathbb{N}^*, \ H_0 \equiv 1. \tag{3.40}$$

The coefficient $(-1)^n$ is sometimes included in the definition of the Hermite polynomials. Definition (3.40) shows that the Hermite polynomials are closely linked to the Gaussian distribution. For practical purposes, their explicit expressions are often used:

$$H_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k n!}{2^k k!(n-2k)!} x^{n-2k}. \tag{3.41}$$

Various properties of these polynomials are gathered in [AS64], Chapter 22. The Gram-Charlier expansion of f is based on these polynomial functions.

Definition 3. *Let the sequence of coefficients c_n be defined as: $c_0 = 1$,*

$$c_n = \frac{1}{n!} \int_{\mathbb{R}} H_n(x) f(x) \mathrm{d}x = \frac{1}{n!} \mathbb{E}\big[H_n(X)\big], \ n \in \mathbb{N}^*.$$

The Gram-Charlier series expansion $\tilde{f}$ of the approximated density f is defined as

$$\tilde{f}(x) = \phi(x) \sum_{n=0}^{\infty} c_n H_n(x), \ x \in \mathbb{R}. \tag{3.42}$$

For practical uses such as pricing of financial derivatives, the quantity of interest can be the cumulative distribution function rather than the density one. The Gram-Charlier expansion can be defined on the cumulative distribution function of f by integrating (3.42). Using that $c_n H_n(x) = \phi^{(n)}(x)$, the expansion writes

$$\mathbb{P}(X \leq x) = \Phi(x) + \phi(x) \sum_{n=1}^{\infty} c_n H_{n-1}(x), \ x \in \mathbb{R}. \tag{3.43}$$

Note that the coefficients c_n are linear combinations of the moments of X. Although the density f is unknown, one needs to be able to compute the moments of X to perform such an expansion in practice. Keep in mind that the polynomials H_n are explicitly known so that c_n are also fully known once the moments X are determined. Moreover, since the series in (3.42) is characterized by the sequence of its coefficients $(c_n)_{n\in\mathbb{N}}$, all density functions having the same sequence of moments $\left(\int_{\mathbb{R}} x^n f(x)\mathrm{d}x\right)_{n\in\mathbb{N}}$ would have the same Gram-Charlier expansion. Consequently, Gram-Charlier approximation method is not appropriate for distributions for which the moments problem is not determined. The question of the convergence of (3.42) arises. In the particular case of distribution not determined by their moments, the series is likely to diverge. To ensure the convergence of (3.42) and thus guarantee a consistent approximation method, f should satisfy some conditions. The notion of functions of bounded variations will be of interest.

Definition 4. *Let f being a complex valued function defined over an interval $I \subset \mathbb{R}$. Let $\mathcal{P}$ be the set of all partition $\sigma = \{x_0, \ldots, x_n\}$ of I satisfying $x_i \leq x_{i+1}$, $i \in [\![1, n]\!]$. For a given $\sigma \in \mathcal{P}$, the associated variation $V(f, \sigma)$ of f is the quantity $V(f, \sigma) := \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)|$.*

[6] Such formulas has been introduced by Olinde Rodrigues in [Rod15].

f is then said to be of bounded variation over I if its total variation is finite:

$$V_T(f) = \sup_{\sigma} V(f, \sigma) < \infty.$$

We have already said that Gram-Charlier expansion is particularly designed for density functions close to the normal one. The next key result due to Cramér that can be found in [Cra28] for instance clarifies this claim.

Theorem 3.30 ([Cra28])**.** *Assume that the unknown density function f is of bounded variation over any subinterval of $\mathbb{R}$ and is such that*

$$\int_{\mathbb{R}} e^{x^2/4} f(x) \mathrm{d}s < \infty. \tag{3.44}$$

Then, for any $x \in \mathbb{R}$, the Gram-Charlier expansion $\tilde{f}$ of f defined in (3.42) satisfies

$$\tilde{f}(x) = \frac{1}{2}\left(f(x^+) + f(x^-)\right).$$

In particular, the Gram-Charlier expansion (3.42) is exact in every continuity point f. Moreover, the convergence is uniform over every finite interval of continuity.

Example 3. *A continuously differentiable function over $\mathbb{R}$ with bounded derivative is of bounded variation over any subinterval of $\mathbb{R}$ so that under integrability condition (3.44), $\tilde{f}(x) = f(x)$ for all $x \in \mathbb{R}$.*

The *sufficient* condition (3.44) is a very restrictive one: f must decrease asymptotically faster than a quadratic exponential function. Most of the common probability distributions do not satisfy it (in particular log-normal, exponential, gamma or chi-squared distributions); some non standard Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$ do satisfy it. Indeed, it is straightforward to check that the density $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-\mu)^2/(2\sigma^2)\right)$ met (3.44) when $\sigma^2 < 2$.

Note furthermore that thanks to Corollary 1 of Section 3.2, Cramér's condition also ensures that one is dealing with distributions determined by their moments. In other words, (3.42) do converge toward the relevant distribution as the number of term increases.

In practice, the series $\tilde{f}$ is approximated by a sum of $N \in \mathbb{N}$ terms and the question of the truncation error naturally arises. There exist few results about it: the following result proved in [Mil29] is generally mentioned. Let us denote the N-th order approximating density by

$$\tilde{f}^N(x) = \phi(x) \sum_{n=0}^{N} c_n H_n(x), \ x \in \mathbb{R}. \tag{3.45}$$

Theorem 3.31 ([Mil29])**.** *Assume that the unknown density f is k times differentiable, has a continuous k-th derivative of bounded variation and that the following functions*

$$x \mapsto x e^{x^2/4} f^{(k)}(x), x \mapsto x^2 e^{x^2/4} f^{(k-1)}(x), \ldots, x \mapsto x^{k+1} e^{x^2/4} f(x),$$

are of bounded variation over $\mathbb{R}$. Then there exists some constant $C_k \in \mathbb{R}_+^*$ such that for all $N \in \mathbb{N}$, all $x \in \mathbb{R}$,

$$\left|f(x) - \tilde{f}^N(x)\right| \leq \frac{C_k}{N^{k/2}} (1 + x^2)^{1/6} e^{-x^2/4}.$$

In the literature, Gram-Charlier expansions are usually performed on variates which has been centred and reduced beforehand. Denote by $\bar{X} = \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}}$ the standardized variate in

present context. On one hand, by doing so we get that the first two moments of the standardized variate are the same of a standard Gaussian variable: $\mathbb{E}[\bar{X}] = 0$ and $\mathrm{Var}(\bar{X}) = 1$. On the other hand, successive integrations by part show that if $G \sim \mathcal{N}(\eta, 1)$ for some $\eta \in \mathbb{R}$, $\mathbb{E}[H_n(G)] = \eta^n$. Consequently, the coefficients c_1 and c_2 appearing in the Gram-Charlier expansion of the density of $\bar{X}$ are zero. More generally, if $\bar{X}$ and G have common moments up to order $p \in \mathbb{N}$ it results in fact that $c_1 = c_2 = \cdots = c_p = 0$. It will be discussed in a different setting in Section 3.4.

In next paragraph, we take a look at a forgotten (at least in recent works) polynomial expansion series similar to the Gram-Charlier approach but requires much less restrictive condition to ensure convergence.

A word on the Uspensky's framework

The question of the convergence of the Gram-Charlier series has been studied by several authors. We mentioned Cramér but many other could be quoted. Cramér himself quoted in [Cra28] the works of [Hil26], [Sze26] and [Usp37]. The first one worked on Hermitian series and provided associated convergence results; the second one studied Laguerre series. We take a closer look at the work of James Victor Uspensky [Usp37] less quoted in recent works. He obtained a convergence of a Gram-Charlier type series under much less restrictive conditions than (3.44).

Uspensky worked with so-called «physicists'» Hermite polynomials $(\tilde{H}_n)_{n\in\mathbb{N}}$ defined through the Rodrigues formulation as

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2} = (-1)^n \tilde{H}_n(x) e^{-x^2}, \ \tilde{H}_0 \equiv 1.$$

Those new polynomials relate to the density measure $g(x)\mathrm{d}x = \frac{e^{-x^2}}{\sqrt{\pi}}\mathrm{d}x$. Note that they can be linked to the previously introduced «probabilists'» Hermite polynomial in (3.40) via $\tilde{H}_n(x) = 2^{n/2} H_n(\sqrt{2}x)$. He then considered the polynomial functions

$$\hat{H}_n(x) = \frac{1}{\sqrt{2^n n! \sqrt{\pi}}} \tilde{H}_n(x),$$

that are normalized polynomials in the sense that $\int_{\mathbb{R}} e^{-x^2} (\hat{H}_n(x))^2 \mathrm{d}x = 1$. He then considered the «Fourier coefficients» of f with respect to these orthonormal basis of polynomials defined as $\alpha_n = \int_{\mathbb{R}} e^{-x^2} f(x) \hat{H}_n(x) \mathrm{d}x, \ n \in \mathbb{N}$. Uspensky considered and studied the following series

$$\sum_{n=0}^{+\infty} \alpha_n \hat{H}_n(x), \ x \in \mathbb{R}, \tag{3.46}$$

and obtained the following result.

Theorem 3.32 ([Usp37]). *Assume that f is absolutely integrable over any finite interval of $\mathbb{R}$, is of bounded variation over a certain interval $[x - m, x + m]$ for some $x \in \mathbb{R}$ and $m > 0$, and is such that the following holds for some $A \in \mathbb{R}$:*

$$\max\left(\int_{-\infty}^{-A} e^{-u^2} f(u)^2 \mathrm{d}u, \int_{A}^{+\infty} e^{-u^2} f(u)^2 \mathrm{d}u \right) < \infty.$$

Then,

$$\frac{1}{2}\left(f(x^+) + f(x^-)\right) = \sum_{n=0}^{+\infty} \alpha_n \hat{H}_n(x).$$

In particular, the expansion is exact at every continuity point of f.

While Theorem 3.30 a is very restrictive result, Theorem 3.32 is much more global. Yet in Gram-Charlier's framework is widely more used in practice essentially for the following reason. Indeed, coefficients c_n in standard Gram-Charlier series (3.42) express as linear combinations of the moments of X. In Uspensky's framework, those coefficients are more complex. They express as linear combinations of $\mathbb{E}[e^{-X^2}X^n]$ and it turns out that the class of stochastic variable offering closed-form formulas for such quantities is (much) smaller than those offering analytical formulas for moments. Moreover and contrary to Theorem 3.30, Theorem 3.32 does not provided any detail on the mode of convergence of the series (3.46).

Remark 23. *The Uspensky's approach is also referred to as a Gauss-Hermite expansion. [BM98] and references therein provide alternative convergence results of the expansion (3.46). All provided sufficient conditions that are still much less restrictive than (3.44). For instance, it is enough for the approximated unknown density f to be such that $x^3 f(x) \xrightarrow[x\to\infty]{} 0$ along with finiteness and continuity of f'' to get the convergence of (3.46).*

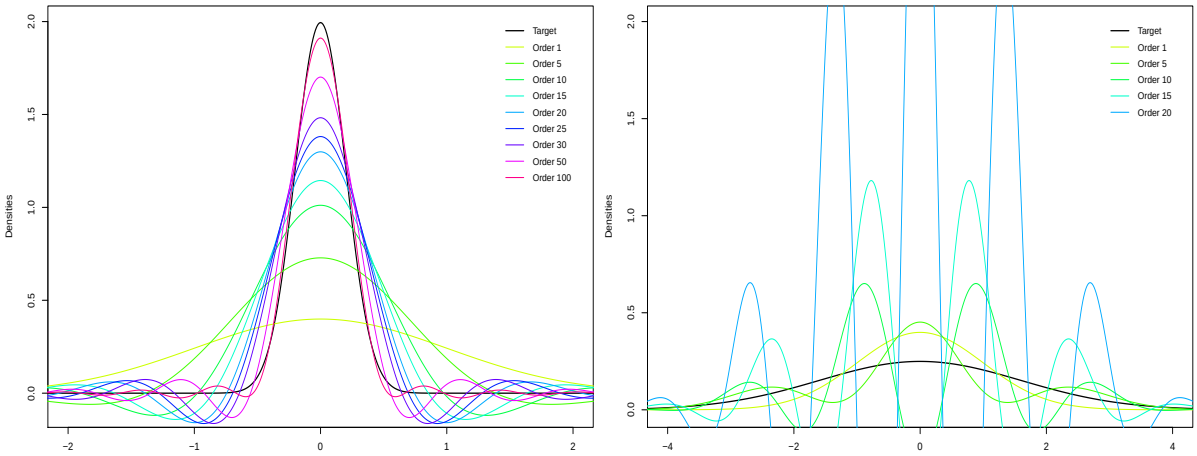
Illustrations

We now provide some examples and illustraions of the presented series expansions.

Gaussian distributions As already mentioned, Gram-Charlier expansion are designed for density close enough to the Gaussian one. It is natural to first check if the method applied thus to non-standard Gaussian distributions. We retake the previous examples of centred Gaussian distribution: if $G \sim \mathcal{N}(0,\sigma^2)$ then (3.44) is satisfied if (and only if) $\sigma^2 < 2$. The distinct behaviour of the Gram-Charlier series regarding the value of the variance parameter is depicted in Figure 3.3. Namely in Figure 3.3a, $\sigma^2 < 2$ but we observe that the convergence is quite slow: at order 100, the approximating density and the target one can still be clearly distinguished. Approximating densities are indeed pseudo-densities since taking negative values. In Figure 3.3b $\sigma^2 > 2$ and the divergence of the successive approximating series appears quickly. Observe that the successive approximating densities $(\tilde{f}^N)_{N\in\mathbb{N}}$ always integrate to one but are non necessarily non-negative: they are sometimes classified as *pseudo-densities*.

Instructive examples Later in this thesis, Gram-Charlier expansions will be performed to approximate density of stochastic volatility type diffusions. To anticipate our future discussions, we examine now one of the simplest –if not the simplest– stochastic volatility type model: it is a static model without dynamics. Let G be a normally distributed random variable $\mathcal{N}(0,\sigma^2)$ and S be a chi-squared distributed variate with 4 degrees of freedom $\chi^2(4)$. S and G are supposed to be independent. Consider the variable $X = \sqrt{S}\times G$: its density function f will be approximated by Gram-Charlier expansions. We first check the possibility to perform such an expansion on it: does f satisfies (3.44)? To answer it we first derive a semi-analytical expression for the function f. To do so consider h a measurable function defined over $\mathbb{R}$ so that $\mathbb{E}\big[|h(X)|\big] < \infty$. We obtain

$$\begin{aligned}\mathbb{E}[h(X)] &= \int_{\mathbb{R}_+^*\times\mathbb{R}} h(\sqrt{s}g)\frac{se^{-s/2}}{4}\frac{e^{-g^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}\mathrm{d}s\mathrm{d}g \\ &= \int_{\mathbb{R}} h(z)\left(\int_{\mathbb{R}_+^*}\frac{\sqrt{t}e^{-t/2}}{4}\frac{e^{-z^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}}\mathrm{d}t\right)\mathrm{d}z,\end{aligned}$$



(a) G.-C. series expansion of a centred Gaussian distribution with standard deviation $\sigma = 0.2$; up to order 100.

(b) G.-C. series expansion of a centred Gaussian distribution with standard deviation $\sigma = 1.6$; up to order 20.

Figure 3.3: Gram-Charlier expansions performed on centred Gaussian distributions.

where we applied the Fubini theorem and used the change of variables $\mathbb{R}_+^* \times \mathbb{R} \ni (s, g) \mapsto (\sqrt{s}g, s)$. We are able to identify f as

$$f(x) = \int_{\mathbb{R}_+^*} \frac{\sqrt{t}e^{-t/2}}{4} \frac{e^{-x^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t, \ x \in \mathbb{R}.$$

Let now $\eta > 0$ and write

$$\mathbb{E}[e^{\eta X^2}] = \int_{\mathbb{R}} e^{\eta x^2} \int_{\mathbb{R}_+^*} \frac{\sqrt{t}e^{-t/2}}{4} \frac{e^{-x^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t\mathrm{d}x.$$

A sufficient condition for this expectation to be finite is that for any $t \geq 0$, the function $\mathbb{R} \ni x \mapsto \frac{\sqrt{t}e^{-t/2}}{4} \frac{e^{-x^2(1/(2t\sigma^2)-\eta)}}{\sqrt{2\pi\sigma^2}}$ is integrable. This can not be ensured and thus $\mathbb{E}[e^{\eta X^2}] = \infty$. Consequently, the sufficient condition (3.44) ensuring the convergence of the Gram-Charlier expansion of f is not guaranteed. Though it is not rigorously equivalent, we have not empirically observe example of density not satisfying (3.44) but whose Gram-Charlier series still converge. We illustrate it below in Figure 3.4. Note that to perform the expansion, the moments of X have been derived in closed-form formula: for all $n \in \mathbb{N}$,

- $\mathbb{E}[X^{2n+1}] = \mathbb{E}[S^{n+1/2}]\mathbb{E}[G^{2n+1}] = 0$;
- $\mathbb{E}[X^{2n}] = \mathbb{E}[S^n]\mathbb{E}[G^{2n}] = 2^n(n+1)! \times \frac{(2n)!}{2^n n!}\sigma^{2n} = (n+1)\sigma^{2n}(2n)!$.

To approximate the density of X anyway and overcome the observed divergence, the following strategy is implemented: approximate X and its density using an auxiliary variate $X^{(m)}$ parametrized by $m \in \mathbb{R}_+$ and whose distribution density has tails thin enough to ensure the Gram-Charlier series to converge. The approximating random variable $X^{(m)}$ should be chosen so that the distance between X and $X^{(m)}$ can be assessed and minimized somehow. The series expansion is then performed on the density f_m of the approximating variable $X^{(m)}$ hopefully

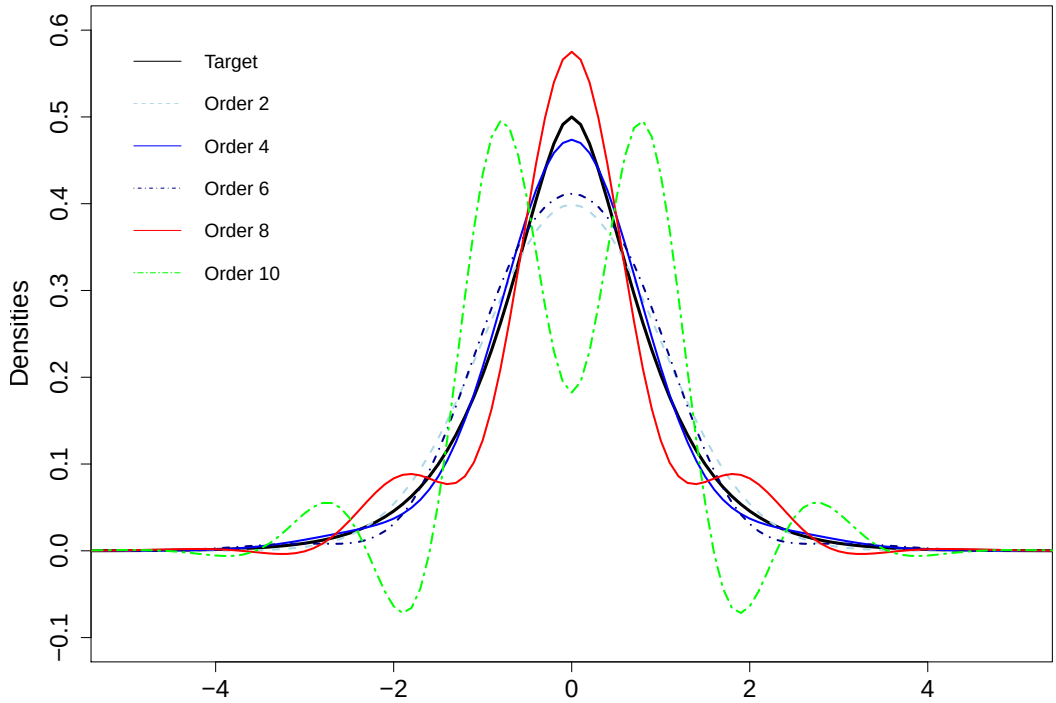


Figure 3.4: G.-C. series expansion of the density of $X = \sqrt{S}G$ up to order 10. As expected, the expansion diverges. Variance parameter: $\sigma^2 = 0.25$.

close to f.
Let $m \in \mathbb{R}_+$ and consider the approximating variable $X^{(m)} = \sqrt{S \wedge m}G$: m allows to truncate the density of S so that tails of the density of X are reduced. We first assess the $\mathbb{L}^2$ distance between X and $X^{(m)}$:

$$\begin{aligned}\mathbb{E}\left[(X - X^{(m)})^2\right] &= \mathbb{E}\left[(\sqrt{S} - \sqrt{S \wedge m})^2\right]\mathbb{E}\left[G^2\right] \\ &= \mathbb{E}\left[(\sqrt{S} - \sqrt{m})^2 \mathbb{1}_{S \geq m}\right]\mathbb{E}\left[G^2\right] \\ &\leq \mathbb{E}\left[|S - m|\,\mathbb{1}_{S \geq m}\right]\sigma^2,\end{aligned} \tag{3.47}$$

using that $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$ for $(x, y) \in (\mathbb{R}_+)^2$. Furthermore, we can easily check that (3.44) is satisfy for $X^{(m)}$ when m is not too large: for $m < 2/\sigma^2$,

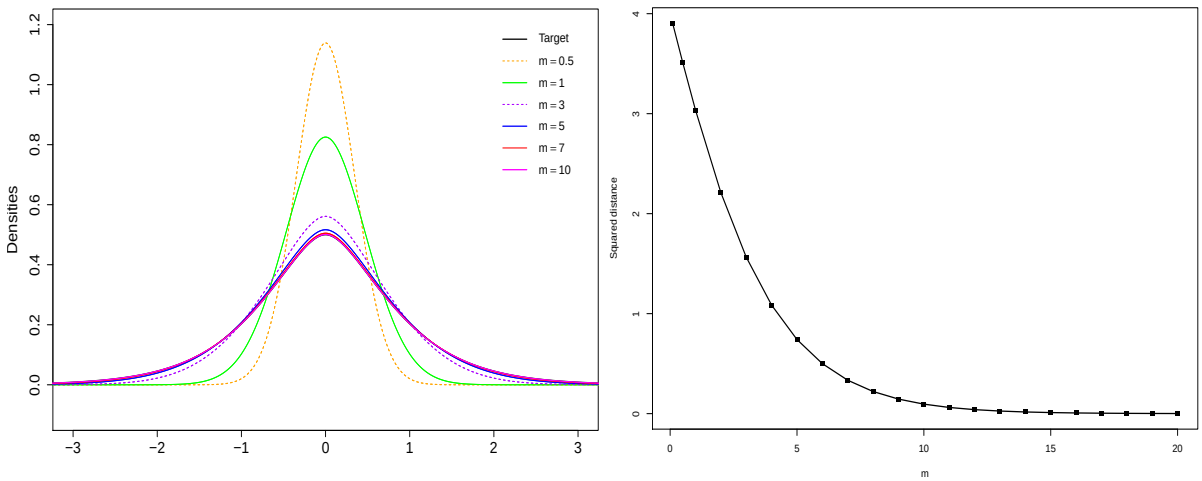
$$\mathbb{E}\left[e^{(X^{(m)})^2/4}\right] \leq \mathbb{E}\left[e^{mG^2/4}\right] < \infty.$$

and (3.44) is satisfied. Note that a trade-off should be made here on the truncating parameter m: on the one hand, to better approximate the density f, m should be taken as large as possible but on the other hand, the convergence condition requires $m < 2/\sigma^2$.
Closed-form formula for density and the moments of $X^{(m)}$ can be derived. The density f_m is identified following the previous method. For a measurable function h such that $\mathbb{E}\big[|h(X^{(m)})|\big] <$

∞, we write

$$\begin{aligned}
\mathbb{E}[h(X^{(m)})] &= \int_{\mathbb{R}_+^* \times \mathbb{R}} h(\sqrt{s \wedge m} g) \frac{s e^{-s/2}}{4} \frac{e^{-g^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}s\mathrm{d}g \\
&= \int_{]0,m] \times \mathbb{R}} h(z) \frac{\sqrt{t} e^{-t/2}}{4} \frac{e^{-z^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t\mathrm{d}z + \int_{]m,\infty[\times \mathbb{R}} h(z) \frac{t e^{-t/2}}{4\sqrt{m}} \frac{e^{-z^2/(2m\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t\mathrm{d}z \\
&= \int_{\mathbb{R}} h(z) \left(\int_{]0,m]} \frac{\sqrt{t} e^{-t/2}}{4} \frac{e^{-z^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t + \int_{]m,\infty[} \frac{t e^{-t/2}}{4\sqrt{m}} \frac{e^{-z^2/(2m\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathrm{d}t \right) \mathrm{d}z.
\end{aligned}$$

so that $f_m(x) = \int_{\mathbb{R}_+} \left(\frac{\sqrt{t} e^{-t/2}}{4} \frac{e^{-x^2/(2t\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathbb{1}_{t \in (0,m]} + \frac{t e^{-t/2}}{4\sqrt{m}} \frac{e^{-x^2/(2m\sigma^2)}}{\sqrt{2\pi\sigma^2}} \mathbb{1}_{t \in (m,\infty)} \right) \mathrm{d}t$. Before performing Gram-Charlier expansion on the density $X^{(m)}$, we illustrate the distance between $X^{(m)}$ and X. In Figure 3.5a we provide the densities f (referred as the *target* density) and f_m for different values of m. Following (3.47) we display in Figure 3.5b the quantity $\mathbb{E}\big[|S - m| \mathbb{1}_{S \geq m}\big]$ as a function of m. In the following experiment, we fix $m = 4$ which corresponds to a reasonable error made when substituting f for f_m.



(a) Densities of f_m for different values of m compared to the target density f.

(b) Bound on $\mathbb{L}^2$ distance between $X^{(m)}$ and X ($\mathbb{E}\big[|S - m| \mathbb{1}_{S \geq m}\big]$ estimated on 10^6 realizations) as a function of m.

Figure 3.5: Gram-Charlier expansions performed on centred Gaussian distributions.

Before performing Gram-Charlier expansions on f_m, let us take a look at the moments of $X^{(m)}$ that express as: for any $n \in \mathbb{N}$,

- $\mathbb{E}\big[(X^{(m)})^{2n+1}\big] = \mathbb{E}\big[(\sqrt{S \wedge m})^{2n+1}\big] \times \mathbb{E}\big[G^{2n+1}\big] = 0$;
- $\mathbb{E}\big[(X^{(m)})^{2n}\big] = \mathbb{E}\big[(S \wedge m)^n\big] \mathbb{E}\big[G^{2n}\big] = \big(\mathbb{E}\big[S^n \mathbb{1}_{s \leq m}\big] + m^n \mathbb{P}(S \geq m)\big) \frac{(2n)!}{2^n n!} \sigma^{2n}$.

We specify this last expression: $\mathbb{P}(S \geq m) = 1 - F_S(m)$ where F_S is the cumulative distribution

function of the $\chi^2(4)$ distribution and writes $F_S(x) = \int_0^x \frac{se^{-s/2}}{4}\mathrm{d}s$ and

$$\begin{aligned}
\mathbb{E}[S^n \mathbb{1}_{s\leq m}] &= \int_0^m s^n \frac{se^{-s/2}}{4}\mathrm{d}s \\
&= \frac{n+1}{2}\int_0^m s^n e^{-s/2}\mathrm{d}s - \frac{1}{2}m^{n+1}e^{m/2} \text{ (Int. by part)} \\
&= \dots \text{ (Successive integrations by part)} \\
&= \frac{1}{2}\left(2^{n+1}(1-e^{-m/2})(n+1)! - e^{-m/2}\sum_{k=1}^{n} 2^k \frac{(n+1)!}{(n-k)!}m^{n-k+1} - m^{n+1}e^{m/2}\right).
\end{aligned}$$

We are thus able to perform the Gram-Charlier expansion on the density f_m of $X^{(m)}$ with a guaranteed convergence. Successive approximating densities of f_m are displayed in Figure 3.6: the convergence is clear now.

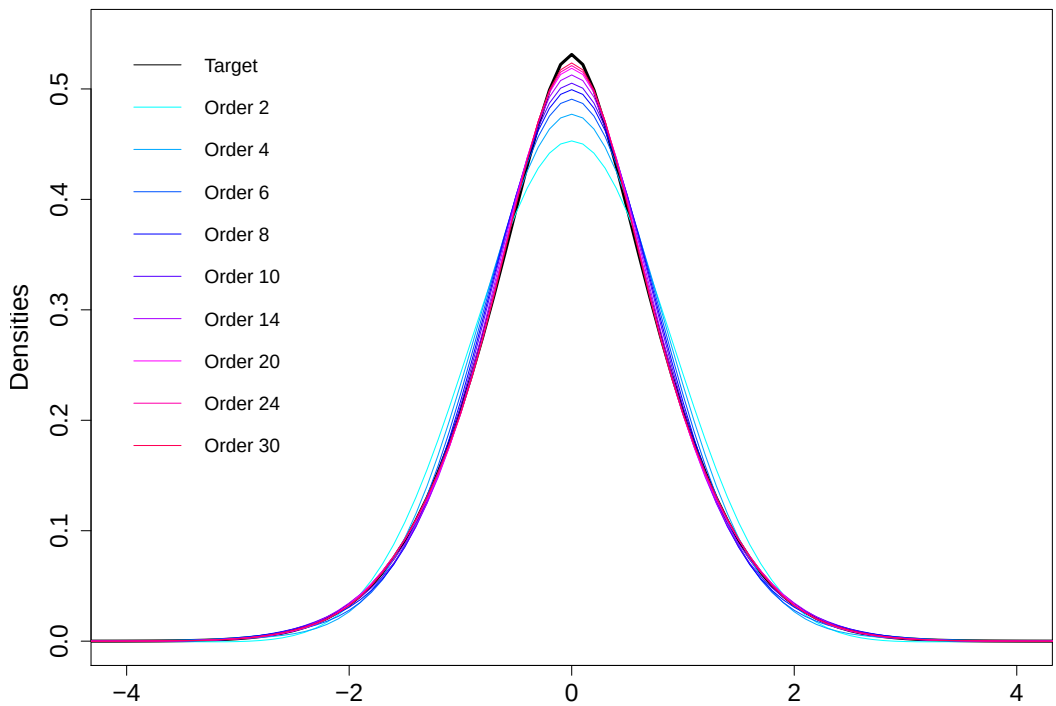


Figure 3.6: G.-C. series expansion of the density of $X^{(m)} = \sqrt{S \wedge m}G$ up to order 30. Parameters: $\sigma^2 = 0.25$ and $m = 4$.

Pathologic examples Let us now take a look at distributions whose supports are strict subsets of $\mathbb{R}$. They are associated with densities functions that write as a given expression multiply by the indicator function of the support. In view of the expression of truncated Gram-Charlier series (3.45), it is clear that the approximating series can not accurately approximate the target density function at least around the discontinuity points. Moreover, beyond boundaries of the target density support, truncated approximating density can not be zero. It would result in the existence of an infinity of zeros of the polynomial function $x \mapsto \sum_{0\leq n\leq N} c_n H_n(x)$. It is not pos-

sible except if all coefficients c_n are zero. That being said, distributions with bounded supports always satisfy (3.44). In practice, it is thus always worthy to try to perform Gram-Charlier expansions on such densities: it may locally result in quite good approximations.

We illustrate our comments below in Figure 3.7. First, a Gram-Charlier expansion is performed on a Gamma distribution $\Gamma(2,2)$ whose support is $\mathbb{R}_+$. It should be observed that the Gamma density does not satisfy the sufficient condition (3.44). The 5-th order expansion allows for a rough but reasonable approximation of the Gamma distribution around its mean 1. Expansion orders beyond 5 are not provided since a clear divergence is noticed.
In Figure 3.8 we perfom a Gram-Charlier expansion of the beta distribution $\beta(2,2)$ whose support is $[0,1]$. The beta distribution does thus satisfy (3.44). We observe a very good fitting over $[0,1]$ as the truncation order increases. Contrary to the previous Gamma distribution case, we are able to go to quite high orders –here up to 100– with good stability in the expansion.

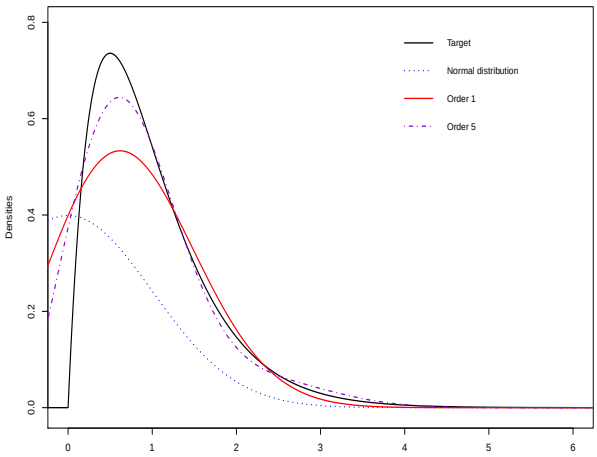


Figure 3.7: G.-C. series expansion of the gamma distribution $\Gamma(2,2)$. Expansion of orders 0, 1 and 5 are depicted.

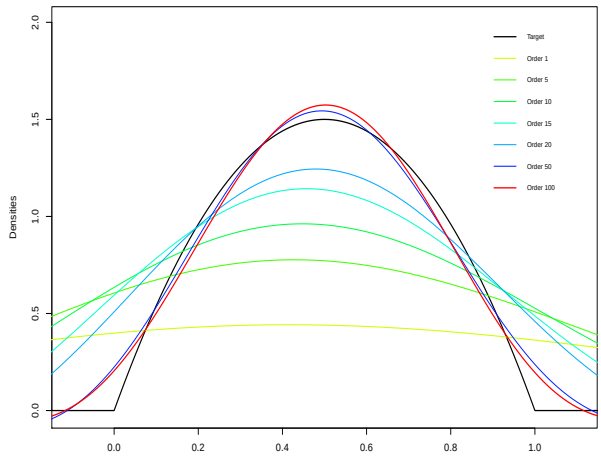


Figure 3.8: G.-C. series expansion of the beta distribution $\beta(2,2)$ up to order 100.

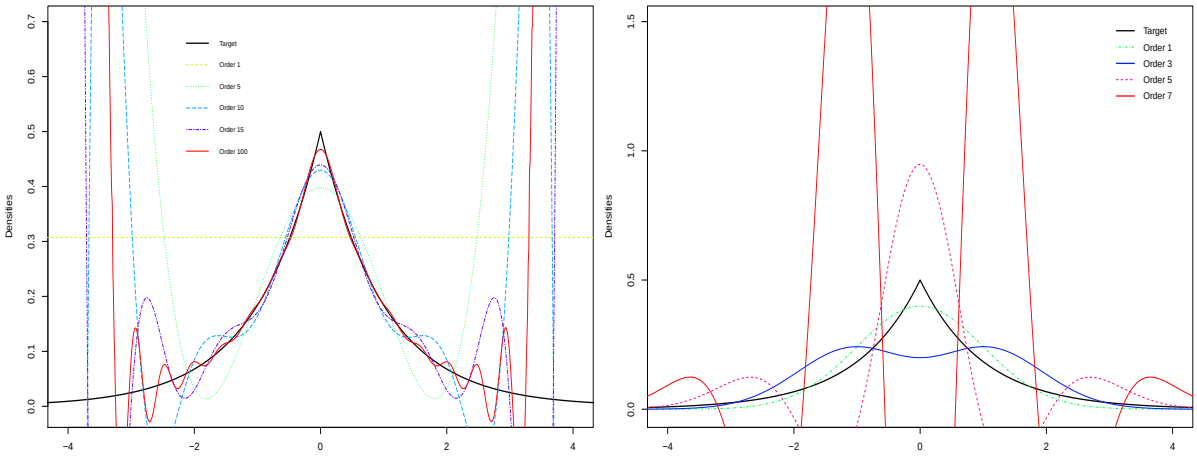
Figure 3.9: Gram-Charlier expansions of distributions of (semi-)bounded supports.

Uspensky framework We illustrate now the framework of [Usp37] resulting notably in (3.46). We apply its methodology to the standard Laplace distribution Laplace$(0,1)$. As discussed above, the difficulty here is to compute the sequence of coefficients $(\alpha_n)_{n\in\mathbb{N}}$: in this particular experiment, we can take advantage of semi-analytical formulas to compute it. Recall that they are defined in present example as $\alpha_n = \mathbb{E}\left[e^{-X^2}X^n\right]$ for $n \in \mathbb{N}$, where $X \sim$ Laplace$(0,1)$. The following identity is admitted and will be employed in numerical experiments:

$$\int_{\mathbb{R}} x^n e^{-x^2} \frac{e^{-|x|}}{2} \mathrm{d}x = \frac{n!}{2^{n+2}} \left(\frac{1}{2} U\left(1+\frac{n}{2}, \frac{3}{2}; \frac{1}{4}\right) + (-1)^n U\left(\frac{n+1}{2}, \frac{1}{2}; \frac{1}{4}\right) \right)$$

where $U(\alpha,\beta,z)$ is the hypergeometric function of the second kind (introduced in [Tri47]; see also Chapter 13 in [AS64]).
Figure 3.10a displays Uspensky approximations up to order 100: we observe that the higher the expansion order is, the better the approximation despite the singularity at the origin. The

tails of the distribution are also better approximated however order 100 is not enough to get a reasonable approximation over the whole selected interval (here $[-4, 4]$). Again, the truncating densities are simply pseudo-densities. For comparison, a standard Gram-Charlier expansion is performed and the first orders of it are provided in Figure 3.10b: the divergence of the expansion is clear.



(a) Uspensky expansion of the standard Laplace distribution up to order 100.

(b) Standard Gram-Charlier expansion of the standard Laplace distribution up to order 7.

Figure 3.10: Approximation of Laplace density function thanks to various expansions approaches.

3.3.3 Edgeworth series

We provide in this section more details on the Edgeworth expansion introduced in first paragraph of Section 3.3.1 and following the presentation of Peter Hall in [Hal13]. Recall that Edgeworth and Gram-Charlier expansions are very similar as they are only distinguished by the order in which terms are ordered. Consequently, both expansion methods only differ from a practical point of view. We recall that the Edgeworth expansion has been introduced while attempting to specify further terms in the Central Limit Theorem. Below we describe in more details how this expansion is performed.

Let us retake the notations of Paragraph 3.3.1.1: let us consider a sequence of i.i.d. random variables $(X_i)_{i\in\mathbb{N}}$ with mean μ and variance σ^2 and their standardized counterparts $(Y_i)_{i\in\mathbb{N}}$. Let $n \in \mathbb{N}$ be given and denote the standardized sum by $S_n = \frac{1}{\sqrt{n\sigma^2}}(X_1 + \cdots + X_n - n\mu)$. Let us denote by ψ_{Y_1} the common characteristic function of the standardized variables $Y_k = (X_k - \mu)/\sigma$ and F their common cumulative distribution function. Let ψ_{S_n} be the characteristic function of S_n. We recall that it is expressed as the power of the characteristic function ψ_{Y_1} as provided in (3.36). Under square integrability assumptions of the variates Y_i, the Central Limit Theorem state that $\psi_{S_n}(u) \xrightarrow[n\to+\infty]{} \psi(u) = e^{-u^2/2}$, for all $u \in \mathbb{R}$. We now specify how additional terms can be obtained to specify this convergence.
The (common) sequence of cumulants of variates $(Y_i)_{1\leq i\leq n}$ denoted $(\kappa_j)_{j\in\mathbb{N}}$ is defined as satisfying (3.37). Using the series representation of the characteristic function, one can alternatively

write: for all $u \in \mathbb{R}$,

$$\begin{aligned}
\log \psi_{Y_1}(u) &= \log \left(1 + \sum_{j \geq 1} \frac{(iu)^j}{j!} \mathbb{E} \left[\left(\frac{X_1 - \mu}{\sigma} \right)^j \right] \right) \\
&= \sum_{p \geq 1} \frac{(-1)^{p+1}}{p} \left(\sum_{j \geq 1} \frac{(iu)^j}{j!} \mathbb{E} \left[\left(\frac{X_n - \mu}{\sigma} \right)^j \right] \right)^p \\
&= \sum_{p \geq 1} \frac{\kappa_p}{p!} (iu)^p.
\end{aligned}$$

Cumulants can thus be expressed as linear combinations of the moments of Y_1. In particular case of standardized variables as in present case one can proves that $\kappa_1 = 0$ and $\kappa_2 = 1$. Combining the latter identity and (3.36) yields

$$\psi_{S_n}(u) = \exp \left(\sum_{p \geq 2} \frac{\kappa_p}{p!} \frac{(iu)^p}{n^{p/2-1}} \right).$$

Since $\left| \psi_{S_n}(u) \right| < \infty$ implies the absolute convergence of all involved series, we obtain using the representation of the exponential function

$$\begin{aligned}
\psi_{S_n}(u) &= e^{-u^2/2} \exp \left(2 \sum_{p \geq 2} \frac{\kappa_p}{p!} \frac{(iu)^{p-2}}{n^{p/2-1}} \right) \\
&= e^{-u^2/2} \sum_{q \geq 0} \frac{1}{q!} \left(\sum_{p \geq 1} \frac{\kappa_p}{p!} \frac{(iu)^p}{n^{p/2-1}} \right)^q \\
&= e^{-u^2/2} \sum_{p \geq 0} \frac{r_p(iu)}{n^{p/2}}
\end{aligned} \tag{3.48}$$

where we introduced the sequence of polynomial functions $(r_p)_{p \geq 0}$. For each $p \in \mathbb{N}$, r_p is defined as being the polynomial function of argument iu in factor of $n^{-p/2}$: it is of degree $3p$ (in particular $r_0 \equiv 1$) and its coefficients depend on the cumulants $\kappa_3, \ldots, \kappa_{p+2}$ but not on n. Assuming in addition that ψ_{S_n} is integrable over $\mathbb{R}$ and applying the Fourier transform technique to it allows to recover the density function f_{S_n} of S_n:

$$f_{S_n}(x) = \int_{\mathbb{R}} e^{-iux} \psi_{S_n}(u) \mathrm{d}u = \int_{\mathbb{R}} e^{-iux - u^2/2} \sum_{p \geq 0} \frac{r_p(iu)}{n^{p/2}} \mathrm{d}u. \tag{3.49}$$

Since the r_j are polynomial functions, they can be written in term of Hermite polynomials: fix $p \in \mathbb{N}$ and let us introduce the sequence of coefficients $(\beta_k^p)_{k \leq 3p}$ such that $r_p(iu) = \sum_{k \leq 3p} \beta_k^p i^k H_k(x)$. Examining (3.48) we deduce that the coefficients β_k are linear combinations of the cumulants of Y_1 and *a fortiori* of the moments of Y_1.

Before going further, we recall that by definition of the Hermite polynomials, we also have

$$\begin{aligned} e^{-t^2/2} &= (-it)^{-1} \int_{\mathbb{R}} e^{itu} \phi(u) \mathrm{d}u \\ &= \ldots \text{(Successive integrations by part)} \\ &= (-it)^{-p} \int_{\mathbb{R}} e^{itu} H_p(u) \phi(u) \mathrm{d}u. \end{aligned} \tag{3.50}$$

Combining (3.49) and (3.50) finally allows to recover the density function of S_n written as a series of increasing exponents of $1/n$.

Definition 5. *The Edgeworth expansion of the density function of S_n denoted by $\hat{f}_{S_n}$ is defined as*

$$\hat{f}_{S_n}(x) = \phi(x) \sum_{p=0}^{\infty} \frac{1}{n^{p/2}} \sum_{k \leq 3p} \beta_k^p x^k, \ x \in \mathbb{R}. \tag{3.51}$$

As for the Gram-Charlier approach in (3.43), an expansion of the cumulated distribution function is sometimes used by integrating (3.51). Formally it writes

$$\mathbb{P}(S_n \leq x) = \Phi(x) + \sum_{j \geq 1} \frac{P_j(x)}{n^{j/2}}, \ x \in \mathbb{R},$$

where P_j a polynomial of degree $3j - 1$ for each $j \in \mathbb{N}^*$.

The truncated Edgeworth expansion (named after the economist and philosopher Francis Ysidro Edgeworth) can be defined similarly as in (3.45): for $N \in \mathbb{N}$,

$$\hat{f}_{S_n}^N(x) = \phi(x) \sum_{p=0}^{N} \frac{1}{n^{p/2}} \sum_{k \leq 3p} \beta_k^p x^k, \ x \in \mathbb{R}. \tag{3.52}$$

In practice, $\hat{f}_{S_n}^N$ is often expressed in term of Hermite polynomials. Some comments should be made at this stage. To our knowledge, there exists no simple analytical formulas for polynomial function $(r_j)_{j \in \mathbb{N}}$ nor recurrence relationship. However, [BM98] proposed a quite simple algorithm to implement for performing high order Edgeworth expansions. Secondly and contrary to the truncated Gram-Charlier series defined in (3.45), truncated Edgeworth expansion (3.52) does not necessarily integrate to one. Third, in practice the variate that is approximated usually reduces to one term: $S_1 = (X - \mu)/\sigma$. This amounts to set $n = 1$ in all previous formulas. Then observe that for a given truncation order $N \in \mathbb{N}$, $\hat{f}^N$ comprises an additional term compared to $\tilde{f}^N$. For instance for $N = 4$,

$$\hat{f}_{S_n}^4(x) = \phi(x) \left(1 + \frac{\kappa_3 H_3(x)}{\sqrt{n}} + \frac{1}{n} \left(\frac{\kappa_4}{24} H_4(x) + \frac{\kappa_3^2}{72} H_6(x) \right) + \frac{\epsilon(n,x)}{n} \right)$$

where the term $\epsilon(n,x) \xrightarrow[n \to \infty]{} 0$ for all x. The term $\frac{\kappa_3^2}{72} H_6(x)$ is an additional one when compared to $\tilde{f}_{S_n}^4$. When convergence is ensured, this additional term is expected to provide a better approximation. We illustrate the correction brought in by the Edgeworth expansion in Figure 3.11. In this example, we approximate the *skew Normal* distribution: it exhibits a strong asymmetric shape monitored by one parameter. Density of the skew Normal distribution writes as $f(x) = \frac{2}{v} \phi\left(\frac{x-\mu}{v}\right) \Phi\left(\alpha \frac{x-\mu}{v}\right)$ where we recall that $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ and $\Phi'(x) = \phi(x)$, $\alpha \in \mathbb{R}$ controls the shape of the distribution, $\mu \in \mathbb{R}$ is the position parameter and $v > 0$ is the scale

parameter. To visualize the impact of the extra term in the Edgeworth expansion, a fourth order Gram-Charlier expansion is performed and compared to a fourth order Edgeworth one. We observe that the asymmetry of the approximating Edgeworth density is quite pronounced and closer to that of the target density function than the one provided by Gram-Charlier density.

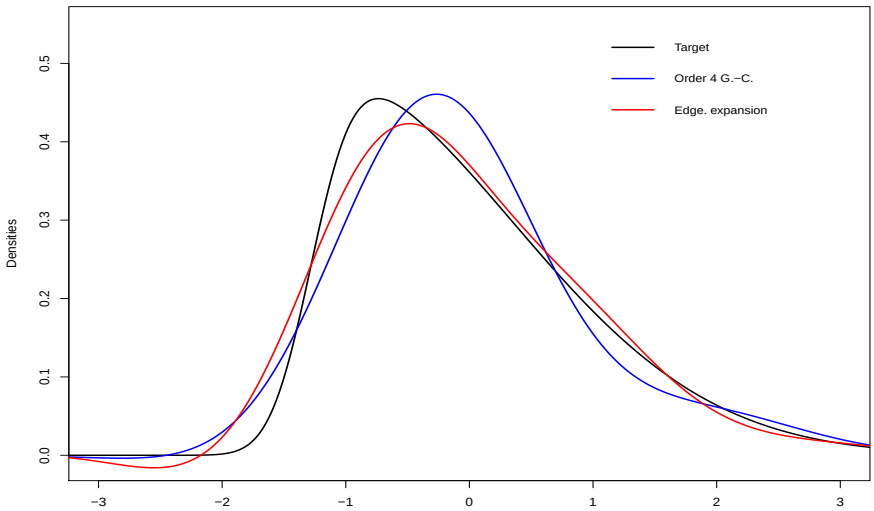


Figure 3.11: Gram-Charlier and Edgeworth expansions approximating a skew normal distribution with $\alpha = 6$, $\mu = v = 1$.

We conclude present section by discussing the convergence of the Edgeworth expansion. Two types of convergence should be discussed here: with respect to the truncation order N similar to the Gram-Charlier approach, and with respect to the number of variates n involved in the definition of S_n. Regarding the convergence of $\hat{f}^N_{S_n}$ towards the approximated density f, first recall that expansions (3.42) and (3.51) coincide up to the order in which terms are ordered. Second, Theorem 3.30 provides that the convergence of Gram-Charlier series in (3.42) is uniform over any bounded interval of continuity of the target density f. Consequently, under (3.44) one can reorder the terms and obtain the Edgeworth expansion. The following result due to Cramér is about the behaviour of Edgeworth approximation as n goes to infinity. Here after we denote by F_{S_n} the cumulative distribution function associated to approximated density function f_{S_n} and by $\mu_k = \int_{\mathbb{R}} |x|^k f_{S_n}(x)\mathrm{d}x$.

Theorem 3.33 ([Cra28])**.** *Assume that f_{S_n} has finite moments up to order 2. Then,*

$$F_{S_n}(x) \xrightarrow[n\to+\infty]{} \Phi(x), \forall x \in \mathbb{R} \text{ (Central Limit Theorem).}$$

If in addition f_{S_n} has a finite third absolute moment $\mu_3 < \infty$, then

$$|F_{S_n}(x) - \Phi(x)| \leq 3\frac{\mu_3}{\mu_2^{3/2}}\frac{\log(n)}{\sqrt{n}}.$$

Cramér extended this result to get the magnitude of the neglected terms when truncating the expansion of the cumulative distribution at arbitrary order and provide this way a result on simultaneous behaviour as $\min(N, n) \to \infty$.

Theorem 3.34. *Assume that f_{S_n} has a finite moment μ_{N+2} for some $N \in \mathbb{N}^*$ and that*

$\lim_{|t|\to+\infty} \sup |\psi_X(t)| < 1$[7]. *Then*

$$F_{S_n}(x) = \Phi(x) + \phi(x) \sum_{k=1}^{N} \frac{P_k(x)}{n^{k/2}} + R_{N,n}(x),$$

where the remainder term is that such there exists a constant M only depending on F and N and not on (x, n) satisfying $|R_{N,n}(x)| \leq Mn^{-(n+1)/2}$.

Remark 24. *Assuming that variables $(X_j)_{1\leq j\leq n}$ have finite absolute moments of order $k \in \mathbb{N}$, that is $\mathbb{E}\left[|X_j|^k\right] < \infty$, $j = 1, \ldots, n$ is enough to ensure that μ_k is finite.*

3.4 Density approximations in Hilbert space

Previous approaches are pointwise converging. We now turn into approximations of density functions in appropriate Lebesgue space of square-integrable densities. The idea is to apply the general theory of such spaces to expand a likelihood ratio within proper functional space along an orthonormal basis. We follow the presentation in Section 2 of [FMS13] to present the approach. note that contrary to preceding approaches we now place ourselves now in general multi-dimensional spaces.

Let E be a topological space associated with a σ-algebra Σ. Let μ be a measure defined over Σ and $\mathcal{L}^2(\mu)$ be the space of square integrable functions with respect to μ: $\mathcal{L}^2(\mu) = \left\{f : E \to E \text{ measurable: } \int_E f(x)^2\mu(\mathrm{d}x) < +\infty\right\}$. We focus on the particular case when $E \subset \mathbb{R}^n$, $n \in \mathbb{N}^*$, Σ is composed of Borel sets and μ has a density w with respect to the Lebesgue measure: $\mu(\mathrm{d}x) = w(x)\mathrm{d}x$. We denote in this particular case $\mathcal{L}^2(\mu) = \mathcal{L}^2_w$ the Lebesgue space of square integrable functions also called weighted space with weighting function w. For $(f, g) \in \left(\mathcal{L}^2_w\right)^2$, we define the scalar product $\langle f, g\rangle_{\mathcal{L}^2_w}$ by

$$\langle f, g\rangle_{\mathcal{L}^2_w} = \int_{\mathbb{R}^n} f(x)g(x)w(x)\mathrm{d}x, \tag{3.53}$$

and the induced $\mathcal{L}^2_w$-norm by $\|f\|_w = \sqrt{\langle f, f\rangle_{\mathcal{L}^2_w}}$.
We still denote by f the unknown density function that one wants to approximate. The only information available about f is the knowledge of its moments

$$\int_{\mathbb{R}^n} x^\alpha f(x)\mathrm{d}x$$

that are assumed to exist and are well defined for any multi-index[8] $\alpha \in \mathbb{N}^n$. Let us choose a reference density function w –thus, explicitly known. The key property here is that $\mathcal{L}^2_w$ equipped with the scalar product $\langle\cdot,\cdot\rangle_{\mathcal{L}^2_w}$ is a Hilbert space. being in a such space, the present technique consists in expanding the likelihood ratio f/w along an orthonormal basis composed of polynomial functions associated to the weighting function w. To do so, it should preliminarily be checked that the mentioned ratio does lie in $\mathcal{L}^2_w$ and that there exists such an orthonormal basis. The following results provide sufficient conditions for these requirements and are proved in [FMS13].

[7] This condition is often called the *Cramér condition*.
[8] For $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$, $x^\alpha = x_1^{\alpha_1} \times \cdots \times x_n^{\alpha_n}$.

The first requirement is equivalent to

$$\int_{\mathbb{R}^n} f(x)^2/w(x)\mathrm{d}x < \infty \tag{3.54}$$

but an alternative sufficient condition is given in the following lemma.

Lemma 3.35 ([FMS13])**.** *Suppose that f is bounded and admits exponential moments that is*

$$\int_{\mathbb{R}^n} e^{\eta\|u\|} f(u)\mathrm{d}u < +\infty,$$

for some $\eta > 0$. In addition, if w is such that

$$\sup_{x\in\mathbb{R}^n} \frac{e^{-\eta\|x\|}}{w(x)} < +\infty,$$

then $f/w \in \mathcal{L}^2_w$.

Remark 25. *Observe that in an expansion in the $\mathcal{L}^2_r$ space associated with a Gaussian density, the condition (3.54) is quite similar to that ensuring pointwise convergence (3.44).*

Below is given a sufficient condition for the existence of an orthonormal basis composed of polynomial functions.

Lemma 3.36 ([FMS13])**.** *Suppose that w admits an exponential moments that is*

$$\int_{\mathbb{R}^n} e^{\eta\|u\|} w(u)\mathrm{d}u < +\infty,$$

for some $\eta > 0$. Then, $\{p$ polynomials defined over $\mathbb{R}^n\}$ is dense in $\mathcal{L}^2_w$ and there exists an orthonormal basis of $\mathcal{L}^2_w$ composed of polynomials functions.

From now on, we assume that $f/g \in \mathcal{L}^2_w$ and that there exists an orthonormal basis of $\mathcal{L}^2_w$ composed of polynomials $\{P_j, j \in \mathbb{N}^n \mid \deg(P_j) = |j|\}$[9]. Note that the sequence of polynomials $(P_j)_{j\in\mathbb{N}}$ depends on the weighting function w. Truncated density functions can be defined as $f^{(N)}(x) = w(x)\big(1 + \sum_{|\alpha|=1}^{N} c_\alpha P_\alpha(x)\big)$ in a similar fashion as in (3.45) or (3.52), where the coefficients c_α are defined by

$$c_\alpha = \left\langle \frac{f}{w}, P_\alpha \right\rangle_{\mathcal{L}^2_w}.$$

They express thus as linear combination of the moments of f with coefficients being that of the polynomials P_α. Expanding f/w along this basis in $\mathcal{L}^2_w$ yields the following theorem.

Theorem 3.37 ([FMS13])**.** *Pseudo-density functions integrate to 1 and approximate f in the following sense: $\lim_{N\to+\infty} f^{(N)}/w = f/w$ in $\mathcal{L}^2_w$ meaning that*

$$\lim_{N\to+\infty} \int_{\mathbb{R}^n} \frac{1}{w(u)} \left| f^{(N)}(u) - f(u) \right|^2 \mathrm{d}u = 0.$$

Remark 26. *In the following of this thesis, by abuse of language, we will also speak of Gram-Charlier expansions for Hilbert expansions built around Gaussian density.*

[9] For $j \in \mathbb{N}^n$, $|j| = \sum_{i=1}^n j_i$.

Parseval-Bessel equality yields

$$\left\|\frac{f}{w}\right\|_{\mathcal{L}^2_w} = 1 + \sum_{|\alpha|\geq 1} c_\alpha^2 \text{ and } \left\|\frac{f^{(N)}}{w}\right\|_{\mathcal{L}^2_w} = 1 + \sum_{|\alpha|=1}^{N} c_\alpha^2,$$

so that the $\mathcal{L}^2_w$-norm of these (pseudo-)likelihood ratios is always greater than 1. In the particular case when $f = w$, it is clear that $\|f/w\|_{\mathcal{L}^2_w} = 1$ meaning that all coefficients $c_\alpha = 0$. Roughly speaking, the more f and w are close in some sense, the more there are null coefficients. The following result justifies this property: in $\mathcal{L}^2_w$, f and w can be deemed «close» by comparing their moments.

Lemma 3.38 ([FMS13]). *Suppose that for some $p \in \mathbb{N}$,*

$$\int_{\mathbb{R}^n} x^\alpha f(x)\mathrm{d}x = \int_{\mathbb{R}^n} x^\alpha w(x)\mathrm{d}x$$

for all $\alpha \in \mathbb{N}^n$ such that $|\alpha| \leq p$. Then, $c_\alpha = 0$ for $1 \leq |\alpha| \leq p$.

This echoes the comments we have made above following Theorem 3.31. On another note, observe that the above Parseval-Bessel identity shows that $c_\alpha \xrightarrow[\alpha\to+\infty]{} 0$.

Illustrations We end present section by giving some classical choices of reference density w along with the associated orthonormal basis when working over $\mathbb{R}$ coming from [AGL16]. These imply different families of orthogonal polynomials and for more properties on it, we refer to the standard book of [Sze39].

1. Normal distribution $\mathcal{N}(0,1)$: the density $w(x)\mathrm{d}x = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x$ is associated with normalized Hermite polynomials $(P_\alpha)_{\alpha\in\mathbb{N}}$ defined by $P_\alpha = \frac{1}{\|H_\alpha\|_{\mathcal{L}^2_w}} H_\alpha$ where H_α is given by (3.41). It should be observed that this framework is close to the above introduced Gram-Charlier expansion. Rigorously, they are not exactly the same since (i) Gram-Charlier type A expansion (3.42) does not involve normalized Hermite polynomials and (ii) Hilbert approximation of Theorem 3.37 does no allow to recover the cumulative distribution function.

2. Gamma distribution $\Gamma(k, 1/\theta)$: the density $w(x)\mathrm{d}x = \frac{e^{-x/\theta}x^{k-1}}{\theta^k\Gamma(k)}\mathbb{1}_{x\geq 0}\mathrm{d}x$ is associated with generalized Laguerre polynomials $(L_\alpha)_{\alpha\in\mathbb{N}}$ defined by $L_\alpha(x) = (-1)^\alpha \left(\frac{\Gamma(n+1)\Gamma(r)}{\Gamma(n+r)}\right)^{1/2} \times \sum_{i=0}^{\alpha} \binom{\alpha+k-1}{\alpha-i} \frac{(-x/\theta)^i}{i!}$[10].

3. Log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$: the density $w(x)\mathrm{d}x = \frac{e^{-(\ln(x)-\mu)^2/(2\sigma^2)}}{x\sqrt{2\pi\sigma^2}}\mathbb{1}_{x\geq 0}\mathrm{d}x$ is associated with following orthonormal polynomials $P_\alpha(x) = \frac{e^{-\alpha^2\sigma^2/2}}{\sqrt{[e^{-\sigma^2},e^{-\sigma^2}]_\alpha}} \sum_{j=0}^{\alpha}(-1)^{\alpha+j}e^{-j\mu-j^2\mu^2/2}e_{\alpha-j}\left(1,\ldots,e^{(\alpha-1)\sigma^2}\right)x^j$ where $e_i(x_1,\ldots,x_n) = \sum_{1\leq j_1<\cdots<j_i\leq n} x_{j_1}\ldots x_{j_n}$ if $i \leq n$, 0 otherwise and where $[x,y]_n = \prod_{k=0}^{n-1}(1-xy^k)$.

[10] For $z \in \mathbb{C}$, we may extend the binomial coefficient as $\binom{z}{n} = \frac{(z)_n}{n!} = \frac{(z)(z-1)\ldots(z-n+1)}{n!}$.

Chapter 4

Jacobi Stochastic Volatility factor for the Libor Market Model

This chapter comprises augmented version of [ABLM20].

4.1 Introduction

We now present the modelling framework we propose to efficiently price swap rates derivatives. In this chapter, we develop a method based on Gram-Charlier expansions introduced earlier to efficiently price swap rates derivatives under the Displaced Diffusion with Stochastic Volatility LIBOR Market Model (see Chapter 1).

Designed using the *freezing* technique (Chapter 2), the *so-called* DDSVLMM lies in the family of Heston like models. For those the Fast Fourier Transform (FTT) method described in [CM99] is generally employed to price derivatives over a grid of strikes. However, integrating the Heston characteristic function in the complex field can cause numerical issues as pointed out and solved by [AMST07]. Recent interest has been dedicated to the significant computational time cost of numerical integration the FTT requires in our context and the use of alternative more efficient pricing techniques. In particular, approximations based on Gram-Charlier and Edgeworth expansion techniques have been proposed. In [DABB17], the authors developed a 4-th order Gram-Charlier and Edgeworth expansions for pricing swap rates derivatives which proves to be competitive. They proposed an adjustment of a reference Gaussian distribution (Bachelier model) taking into account skewness and kurtosis and as a by-product derived a smile formula linking the volatility to the moneyness with interpretable parameters. These pricing formulas involve moments up to order four and are provided in analytical form by taking advantage of the explicit knowledge of the characteristic function in DDSVLMM. This technique is particularly interesting due to the tractability and the interpretability of the Bachelier model.

Variants of Gram-Charlier expansions have been proposed for Heston-type models: as pointed out in [FMS13], a bilateral Gamma density can be used as reference density in the Heston model at the price of computational difficulties relating to the orthonormal basis of polynomials, see for instance [AGL16]. Nevertheless, proving the convergence of Gram-Charlier series in the case of such affine dynamics is not fully solved. While [AFP17] justified the use of Gram-Charlier expansion in an equity context under the assumption of a bounded volatility, in this chapter we will see that the classical sufficient condition presented in [FMS13] used to establish the convergence of the expansion technique is not satisfied for most unbounded stochastic volatility models including the classical Heston model. Note also that the technique consisting in writing derivatives prices obtained under some complex models in terms of prices induced by simpler models (usually, Black-Scholes or Bachelier ones) plus additional corrective

terms has been used in various frameworks using PDEs techniques (see for instance [BGM10], [Bom13] or [GKT16]).

To provide a setup in which convergence of Gram-Charlier series is guaranteed, we will use a Jacobi dynamics (see Chapter 3). for modelling the volatility factor following [AFP17]. It is a bounded process by definition initially introduced in biology to study gene frequencies (see for instance [KT81]). Applications to mathematical finance have been studied more recently: [DS02] worked with an interest-rate model based on a Jacobi dynamics whereas [Ma09] studied a stochastic correlation adjustment modelled by a Jacobi process. In this work, Jacobi dynamics is viewed as a numerical tool to approximate the Cox-Ingersoll-Ross one. Computation of its moments is still tractable thanks to the polynomial property of this process. Polynomial processes have been introduced in [EP11] and [CKRT12] with further applications to financial modelling in [FL16] or in [Cuc19].

The new approximation of the DDSVLMM we propose - and referred to as the Jacobi version of the DDSVLMM - is built using the works of [AFP17] and [DABB17]. As in [AFP17], we will prove that the proposed model converges to the original Heston-like approximation of the DDSVLMM when volatility bounds vanish. Our approximating model has thus the advantages of allowing robust pricing based on Gram-Charlier expansion while being an approximation of the standard DDSVLMM. The proposed model should be viewed as an numerical efficient method for swaptions pricing. Note that we work with time-dependent coefficients (piecewise constant) which somewhat extends the previously mentioned works.

The quality of the approximation we propose can be measured through different means. We will be interested in by both strong and weak convergence of the Jacobi version of the DDSVLMM towards the reference approximated DDSVLMM as the upper volatility bound goes to infinity. [AFP17] proved the weak convergence of the Jacobi-based modelling towards the CIR-based one. In this work, we quantify it by obtaining a weak speed of convergence. Moreover, we have been able to get a strong convergence rate for the Jacobi approximation of the DDSVLMM in the $\mathbb{L}^1$ space. By exploiting it, we derive an estimation of the pricing error made when computing prices of swap derivatives when replacing the standard CIR stochastic volatility factor by a Jacobi one.

Swap rate model

We briefly recap the standard setting of the Displaced Diffusion LIBOR Market Model with Stochastic Volatility (DDSVLMM) we employ to price swaptions as described in Chapter 1. [JR03] proposed a twofold extension of the standard LMM: a stochastic volatility factor has been added to reproduce the observed implied smile of volatility while a displacement coefficient (also called *shift*) allows to generate negative interest-rates, which has became necessary in view of late market conditions. [WZ06] proposed a tractable version - without displacement factor - of the model by adding to the *forward rates* a stochastic volatility factor that is modelled by a Feller process (Cox-Ingersoll-Ross dynamics). Since the resulting dynamics of the swap rate process is complex, [WZ06] proposed to apply the so-called *freezing* approximation yielding a Heston-type dynamics for the swap rate process as introduced below.

Under the probability measure $\mathbb{P}^S$ (the *forward swap measure* named after [Jam97]) associated to the *numéraire* B^S defined in (A.21), the swap rate is a martingale. Its dynamics writes is given, for $t \leq T_m$:

$$\begin{aligned} \mathrm{d}S_t^{m,n} &= \sqrt{V_t}\Big(\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|\mathrm{d}W_t + \sqrt{1-\rho(t)^2}\boldsymbol{\lambda}^{m,n}(t)\cdot \mathrm{d}\boldsymbol{W}_t^{S,*}\Big), \\ \mathrm{d}V_t &= \kappa\big(\theta - \xi^0(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}W_t, \end{aligned} \tag{4.1}$$

where $(\boldsymbol{W}_t^{S,*})_{0\le t\le T_m}$ and $(W_t)_{0\le t\le T_m}$ are respectively d-dimensional and 1-dimensional Brownian motions under $\mathbb{P}^S$. Components of $\boldsymbol{W}^{S,*}$ are all independent one another and of W. The coefficients κ, θ, and ϵ are non-negative parameters. Under Feller condition $2\kappa\theta \ge \epsilon^2$ the process V remains non-negative almost surely as long as $V_0 > 0$. The time-dependent coefficients are all assumed to be bounded. In particular, ξ^0 is positive and bounded: $1 \le \xi^0_{min} \le \xi^0(t) \le \xi^0_{max}$. All time dependent quantities are assumed to be piecewise constant on time intervals $[T_i, T_{i+1}[, 1 \le i \le m-1$. The function ρ accounts for the correlation between the swap rate and its instantaneous volatility; d-dimensional vector function $\boldsymbol{\lambda}^{m,n}$ distorts the volatility structure over time; ξ^0 is a deterministic adjustment in the drift term of the instantaneous volatility due to the correlation structure between swap rate and instantaneous volatility. Detailed parametrization of the time-varying functions involved in (4.1) is given in Section 4.5.2. For more details on the derivation of the approximating dynamics (4.1), the interested reader is referred to [WZ06]. We also refer to [BGM10] for the treatment of the Heston model with time-dependent parameters.

Observe that (4.1) corresponds to a (logarithmic) Heston-type process. As an affine dynamics, (4.1) offers the ability of explicitly knowing the moment generating function of $S^{m,n}$ through Riccati equations. This has been developed and solved in [DABB17] or in [WZ06] for the swaptions pricing. Exploiting the analytical knowledge of the moment generating function allows to derive closed-form formulas for prices of swap rate derivatives (especially, for swaptions). In our work, we aim at pricing such derivatives using Gram-Charlier expansions.

4.2 Density approximation for stochastic volatility models

In present section, we prove a short result that is central for the motivation of our work. It demonstrates that density function generated by common stochastic volatility type models can not be accurately approximated by Gram-Charlier expansion technique. Said in other words, the sufficient condition (3.54) (or (3.44)) is not satisfied for most of stochastic volatility models. We retake the framework introduced in Section 3.4 of expansion series in Hilbert space. The reference distribution used to build the Hilbert space we will work in is a Gaussian one of mean $\mu_r \in \mathbb{R}$ and variance $\sigma_r^2 > 0$. We denote its density function by

$$g_r(x) = e^{-(x-\mu_r)^2/(2\sigma_r^2)}/\sqrt{2\pi\sigma_r^2} \tag{4.2}$$

and define the $\mathcal{L}_r^2$ Hilbert space as

$$\mathcal{L}_r^2 = \left\{ h : \mathbb{R} \to \mathbb{R} \text{ measurable such that } \|h\|_r^2 := \int_{\mathbb{R}} h(u)^2 g_r(u)\mathrm{d}u < \infty \right\}.$$

We denote by $(\check{H}_n)_{n\in\mathbb{N}}$ the orthonormal basis of $\mathcal{L}_r^2$. They express in term of standard Hermite polynomials defined in (3.41) as

$$\check{H}_n(x) = \frac{1}{\sqrt{n!}} H_n\Big(\frac{x-\mu_r}{\sigma_r}\Big). \tag{4.3}$$

We recall that, by abuse of language, we will still speak of Gram-Charlier expansions in this Hilbert space while it rigorously refers to pointwise expansions based on standard Hermite polynomials (see discussions Section 3.3).

The result proved below in Theorem 4.1 does not restrict to the particular case of swap rates models; this is why, we beforehand introduce some general modelling notations for the purpose of the present section. Let us fix a time horizon T. Under some probability measure $\mathbb{P}^*$, g_T will denote the density of S_T that is the risk factor we are modelling. Its evolution over $[0, T]$

is given by

$$\begin{aligned} \mathrm{d}S_t &= \mu(t,V_t)\mathrm{d}t + u(V_t)\Big(\rho(t)\|\boldsymbol{\lambda}(t)\|\mathrm{d}W_t + \sqrt{1-\rho(t)^2}\boldsymbol{\lambda}(t)\cdot\mathrm{d}\boldsymbol{W}_t^*\Big), \\ \mathrm{d}V_t &= b(t,V_t)\mathrm{d}t + h(t,V_t)\mathrm{d}W_t, \end{aligned} \tag{4.4}$$

where $\mu : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ is a regular function with linear growth, $\boldsymbol{\lambda}$ is a bounded vector function modelling the time dependent deformation of the volatility structure such that $0 < \lambda_{min} \leq \|\boldsymbol{\lambda}(t)\| \leq \lambda_{max} < \infty$ for any time $t \in [0,T]$, $u : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $b : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ and $h : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ are assumed to be regular enough to ensure existence and uniqueness of a solution to (4.4), ρ is such that $\sup_{0\leq t\leq T}\rho(t)^2 < 1$ and accounts for the instantaneous correlation between S and V, $(W_t)_{t\leq T}$ is a Brownian motion and $(\boldsymbol{W}_t^*)_{t\leq T}$ is a multidimensional Brownian motion whose components are all independent one another and independent from W.

Theorem 4.1. *Assume that the cumulated variance $\int_0^T u(V_s)\mathrm{d}s$ is positive almost-surely and unbounded in the sense that, for all $M > 0$,*

$$\mathbb{P}^*\left(\int_0^T u(V_s)\mathrm{d}s \geq M\right) > 0 \text{ and } \mathbb{P}^*\left(\int_0^T u(V_s)\mathrm{d}s = 0\right) = 0. \tag{4.5}$$

Then, the likelihood ratio g_T/g_r does not lie in $\mathcal{L}_r^2$ that is

$$\int_{\mathbb{R}} g_T(u)^2 e^{\frac{(u-\mu_r)^2}{2\sigma_r^2}}\mathrm{d}u = \infty.$$

An immediate consequence is that g_T/g_r can not be conveniently approximated by a Gram-Charlier expansion series in $\mathcal{L}_r^2$.

Proof. We first derive an analytical expression of the density g_T (the method to derive the expression of g_T is employed on multiple occasions in this thesis this is why a fuller discussion is made in Appendix C). From the dynamics (4.4), we directly get that

$$S_T \overset{a.s.}{=} S_0 + \int_0^T \mu(t,V_t)\mathrm{d}t + \int_0^T u(V_t)\rho(t)\|\lambda(t)\|\mathrm{d}W_t + \int_0^T u(V_t)\sqrt{1-\rho(t)^2}\boldsymbol{\lambda}(t)\cdot\mathrm{d}\boldsymbol{W}_t^*.$$

Then conditionally to $\mathcal{F}_T^W = \sigma(W_t, t \leq T)$, S_T is normally distributed with mean $\tilde{S}_T := S_0 + \int_0^T \mu(t,V_t)\mathrm{d}t + \int_0^T u(V_t)\rho(t)\|\lambda(t)\|\mathrm{d}W_t$ and variance $C_T^2 := \int_0^T \big(1-\rho(t)^2\big)u(V_t)^2\|\lambda(t)\|^2\mathrm{d}t$. By application of Jensen's inequality:

$$C_T^2 \geq \tfrac{\lambda_{min}^2}{T}\big(1-\sup_{t\leq T}\rho(t)^2\big)\left(\int_0^T u(V_s)\mathrm{d}s\right)^2 \text{ a.s.,} \tag{4.6}$$

and thus with Assumption (4.5), we deduce $\mathbb{P}\left(C_T^2 = 0\right) = 0$ (recall that we have assumed $1-\sup_{0\leq t\leq T}\rho(t)^2 > 0$). For any measurable function f such that $\mathbb{E}\big[|f(S_T)|\big] < \infty$, Fubini theorem yields

$$\mathbb{E}^*[f(S_T)] = \mathbb{E}^*\big[\mathbb{E}^*[f(S_T)|\mathcal{F}_T^W]\big] = \int_{\mathbb{R}} f(x)\mathbb{E}^*\Big[\big(2\pi C_T^2\big)^{-1/2}e^{-\frac{(x-\tilde{S}_T)^2}{2C_T^2}}\Big]\mathrm{d}x$$

which allows to identify the density of S_T as $g_T(x) = \mathbb{E}^*[G_T(x)]$ where $G_T(x) := (2\pi C_T^2)^{-1/2} \times e^{-\frac{(x-\tilde{S}_T)^2}{2C_T^2}}$ for all $x \in \mathbb{R}$. Equation (4.6) along with Assumption (4.5) shows that $\mathbb{P}^*\left(C_T^2 > 2\sigma_r^2\right) >$

0 for all $\sigma_r > 0$ and thus $Y_T := -\left(\frac{1}{2C_T^2} - \frac{1}{4\sigma_r^2}\right) > 0$ with positive probability. Now, we have

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{g_T^2(x)}{g_r(x)}\mathrm{d}x = \sqrt{2\pi\sigma_r^2}\int_{\mathbb{R}} \mathbb{E}^*[G_T(x)]^2 e^{(x-\mu_r)^2/(2\sigma_r^2)}\mathrm{d}x \\
\geq\ & \sqrt{2\pi\sigma_r^2}\int_{-a}^{a} \mathbb{E}^*\left[\mathbb{1}_{C_T^2>2\sigma_r^2} \times G_T(x) e^{(x-\mu_r)^2/(4\sigma_r^2)}\right]^2 \mathrm{d}x \\
\geq\ & \frac{\sqrt{2\pi\sigma_r^2}}{2a}\Big(\int_{-a}^{a} \mathbb{E}^*\big[\mathbb{1}_{C_T^2>2\sigma_r^2} \times G_T(x) e^{(x-\mu_r)^2/(4\sigma_r^2)}\big]\mathrm{d}x\Big)^2 \text{ (Jensen inequality)} \\
=\ & \frac{\sigma_r}{2a}\mathbb{E}^*\left[\frac{\mathbb{1}_{C_T^2>2\sigma_r^2}}{C_T}\int_{-a}^{a} e^{-\frac{(x-\tilde{S}_T)^2}{2C_T^2}+\frac{(x-\mu_r)^2}{4\sigma_r^2}}\mathrm{d}x\right]^2 \text{ (Fubini theorem)} \\
=\ & \frac{\sigma_r}{2a}\mathbb{E}^*\left[\frac{\mathbb{1}_{Y_T>0}}{C_T}\exp\left(\frac{\mu_r^2}{4\sigma^2} - \frac{\tilde{S}_T^2}{2C_T^2} - \frac{1}{4Y_T}\Big(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\Big)^2\right)\int_{-a}^{a} e^{\left(\sqrt{Y_T}x + \frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)\right)^2}\mathrm{d}x\right]^2 \\
=\ & \frac{\sigma_r}{2}\mathbb{E}^*\left[\frac{\mathbb{1}_{Y_T>0}}{\sqrt{Y_T}C_T}\exp\left(\frac{\mu_r^2}{4\sigma^2} - \frac{\tilde{S}_T^2}{2C_T^2} - \frac{1}{4Y_T}\Big(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\Big)^2\right)\frac{1}{a}\int_{-\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)}^{\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)} e^{y^2}\mathrm{d}y\right]^2.
\end{aligned}$$

By Fatou's lemma

$$\begin{aligned}
&\lim_{a+\infty}\mathbb{E}^*\left[\frac{\mathbb{1}_{Y_T>0}}{\sqrt{Y_T}C_T}\exp\left(\frac{\mu_r^2}{4\sigma^2} - \frac{\tilde{S}_T^2}{2C_T^2} - \frac{1}{4Y_T}\Big(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\Big)^2\right) \times \frac{1}{a}\int_{-\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)}^{\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)} e^{y^2}\mathrm{d}y\right] \\
&\geq \mathbb{E}^*\left[\frac{\mathbb{1}_{Y_T>0}}{\sqrt{Y_T}C_T}\exp\left(\frac{\mu_r^2}{4\sigma^2} - \frac{\tilde{S}_T^2}{2C_T^2} - \frac{1}{4Y_T}\Big(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\Big)^2\right)\lim_{a+\infty}\frac{1}{a}\int_{-\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)}^{\sqrt{Y_T}a+\frac{1}{2\sqrt{Y_T}}\left(\frac{\tilde{S}_T}{C_T^2} - \frac{\mu_r}{2\sigma^2}\right)} e^{y^2}\mathrm{d}y\right].
\end{aligned}$$

Finally, the lower bound can be proved to be infinite since for any $\lambda > 0$ and $\mu \in \mathbb{R}$,

$$\frac{1}{a}\int_{-\lambda a+\mu}^{\lambda a+\mu} e^{y^2}\mathrm{d}y \geq \frac{1}{a}\int_{\frac{\lambda a+\mu}{2}}^{\lambda a+\mu} e^{y^2}\mathrm{d}y \geq \frac{\lambda a+\mu}{2a}e^{\frac{(\lambda a+\mu)^2}{4}},$$

so that $\lim_{a+\infty}\frac{1}{a}\int_{-\lambda a+\mu}^{\lambda a+\mu} e^{y^2}\mathrm{d}y = \infty$. This concludes the proof. □

The standard DDSVLMM as introduced in (4.1) does satisfy the assumptions of Theorem 4.1 which motivates the approach we have followed to approximate swaptions prices.

4.3 Swaption pricing with Gram-Charlier expansion

4.3.1 Jacobi process in the DDSVLMM

We introduce the *Jacobi version* of the DDSVLMM which is an approximation of the model (4.1). In the proposed setting, the volatility factor is modelled by a $[v_{min}, v_{max}]$-valued Jacobi process. The swap rate modelled in our proposal is denoted by $S^{m,n,J}$ and is described over $[0, T_m]$ by

$$\begin{aligned}
\mathrm{d}S_t^{m,n,J} &= \sqrt{Q(V_t)}\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|\mathrm{d}W_t + \sqrt{V_t - \rho(t)^2 Q(V_t)}\boldsymbol{\lambda}^{m,n}(t)\cdot\mathrm{d}\boldsymbol{W}_t^{S,*}, \\
\mathrm{d}V_t &= \kappa\big(\theta - \xi^0(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{Q(V_t)}\mathrm{d}W_t,
\end{aligned} \tag{4.7}$$

$\boldsymbol{\lambda}^{m,n}$, ξ^0 and ρ are the same as in (4.1), Q is a bounding function defined by $Q(v) = \frac{(v_{max}-v)(v-v_{min})}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2}$, where $0 \leq v_{min} < v_{max} \leq \infty$. Observe that $Q(v) \leq v$ for any $v \in \mathbb{R}$ and that $Q(v) \geq 0$ for $v \in [v_{min}, v_{max}]$. We recall that all components of $\boldsymbol{W}^{S,*}$ are independent and are also all independent from W. The volatility factor $(V_t)_{t\geq 0}$ follows a Jacobi dynamics with additional time dependency in the drift. To be well defined, θ should lie in $[v_{min}, v_{max}]$. Some properties of this process are given in Section 3.1. For this dynamics, the Feller condition writes:

$$\frac{\epsilon^2(v_{max} - v_{min})}{(\sqrt{v_{max}} - \sqrt{v_{min}})^2} \leq 2\kappa \min\left(\xi^0_{min} v_{max} - \theta, \theta - \xi^0_{max} v_{min}\right). \tag{4.8}$$

It ensures the process V in (4.7) to remain bounded at any date: $\mathbb{P}\big(\forall t \in [0, T_m] : V_t \in (v_{min}, v_{max})\big) = 1$. In this setting, the coefficient $\rho(t)$ (precisely set below in (4.60)) is interpreted in dynamics (4.7) as a scaling factor of the instantaneous correlation between the swap rate and its volatility since the following holds:

$$\frac{\mathrm{d}\left\langle V_\cdot, S^{m,n,J}_\cdot\right\rangle_t}{\sqrt{\mathrm{d}\left\langle V_\cdot, V_\cdot\right\rangle_t}\sqrt{\mathrm{d}\left\langle S^{m,n,J}_\cdot, S^{m,n,J}_\cdot\right\rangle_t}} = \rho(t)\sqrt{\frac{Q(V_t)}{V_t}}.$$

Since $0 \leq Q(v) \leq v$ for $v \in [v_{min}, v_{max}]$, the instantaneous correlation is smaller than $\rho(t)$ at each time. The time-dependent infinitesimal generator of the diffusion (4.7) applied to a function $[v_{min}, v_{max}] \times \mathbb{R} \ni (v, s) \longmapsto f(v, s)$ is given by

$$\begin{aligned}\mathcal{A}_t f(v,s) = \kappa\big(\theta - \xi^0(t)v\big)\frac{\partial f}{\partial v}(v,s) + \frac{\epsilon^2}{2}Q(v)\frac{\partial^2 f}{\partial v^2}(v,s) + \frac{v}{2}\|\boldsymbol{\lambda}^{m,n}(t)\|^2\frac{\partial^2 f}{\partial s^2}f(v,s)\\ + \epsilon Q(v)\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|\frac{\partial^2 f}{\partial s \partial v}f(v,s),\ f \in \mathrm{Dom}(\mathcal{A}_t),\ t \leq T_m.\end{aligned} \tag{4.9}$$

Note that for all $k \in \mathbb{N}$, at all date $t \in [0, T]$, $\mathcal{A}_t(\mathcal{P}_k(\mathbb{R}^2)) \subset \mathcal{P}_k(\mathbb{R}^2)$. Then, the dynamics (4.7) is a polynomial diffusion in the terminology of [FL16]; in addition it is a Markov process and thus (4.7) is a also 2-dimensional polynomial model in the terminology of [CKRT12]. Marginal moments of $S^{m,n,J}$ solution of (4.7) can thus be computed by matrix exponentials following method presented in Section 3.1.9.

Setting $v_{min} = 0$ and $v_{max} = \infty$ in (4.7), which amounts to set $Q(v) = v$, allows to recover the reference approximation of the DDSVLMM (4.1). The proposed process (4.7) converges towards the standard one (4.1) in the path space of processes as $v_{min} \to 0$ and $v_{max} \to \infty$, explaining why (4.7) can be viewed as an approximation of the reference model (4.1). This result is proved in [AFP17], Theorem 2.3. In Section 4.4.1 below, we will be able to further prove a $\mathbb{L}^1$ convergence in Theorem 4.5 coming along with a convergence rate.

4.3.2 Gram-Charlier expansion

We now present results showing that the Gram-Charlier expansion can be performed under Jacobi-based dynamics (4.7) and extend the previous results of [AFP17]. Let $T \leq T_m$ be a fixed time horizon. The Gram-Charlier expansion technique performed on the density function f_T of $S^{m,n,J}_T$ is justified now. We assume that the two following conditions hold:

Assumption 1. $4\kappa\theta > \epsilon^2$,

Assumption 2. $\sup_{t\in[0,T]} |\rho(t)| < 1$.

Note that when $v_{min} = 0$, Feller condition (4.8) implies that $4\kappa\theta > \epsilon^2$ meaning that Assumption 1 is stronger than (4.8).

4.3.2.1 Convergence result

Each coordinate of the $d-$dimensional function $t \in [0,T] \longmapsto \boldsymbol{\lambda}^{m,n}(t)$ is piecewise constant and so is its norm $t \in [0,T] \longmapsto \|\boldsymbol{\lambda}^{m,n}(t)\|$. We recall that its bounds are denoted by λ_{min} and λ_{max} and satisfy $0 < \lambda_{min} \leq \|\boldsymbol{\lambda}^{m,n}(t)\| \leq \lambda_{min} < \infty$. We define the cumulated volatility process

$$\Xi_t := \int_0^t \|\boldsymbol{\lambda}^{m,n}(s)\|^2 \big(V_s - \rho(s)^2 Q(V_s)\big)\mathrm{d}s, \ t \leq T. \tag{4.10}$$

Following the notations of Section 4.2, f_T will be approximated using a reference Gaussian density denoted by g_r. In some numerical experiments presented in Section 4.5, g_r will be chosen as a Gaussian mixture which will allow for matching the moments of f_T. Both choices allow to build a Hilbert space $\mathcal{L}_r^2$ having an orthonormal basis composed of polynomial functions (see Lemma 3.36).

Theorem 4.2. *We suppose that Assumption 1 and Assumption 2 hold, $v_{min} \geq 0$ and $v_{max} < \infty$. Consider now a centered Gaussian density g_r of variance σ_r^2 satisfying*

$$\sigma_r^2 > \frac{T v_{max}}{2} \lambda_{max}^2. \tag{4.11}$$

Then, a Gram-Charlier expansion can be performed on the density f_T of (4.7) using g_r the reference density. In particular, the sufficient condition to the $\mathcal{L}_r^2-$convergence of the family of approximating densities is satisfied; that is

$$\int_{\mathbb{R}} \frac{f_T(u)^2}{g_r(u)} \mathrm{d}u < \infty.$$

Remark 27. *[AFP17] proved similar claims as those in Theorem 4.2 in a time-independent framework by assuming that $\rho^2 < 1$ and $v_{min} > 0$. Here we still constrain the correlation coefficient but Assumption 1 which is stronger than the Feller condition in some sense allows to prove the statement also in the case $v_{min} = 0$.*

Remark 28. *When using a non-centered Gaussian density g_r with mean $\mu_r \in \mathbb{R}$, a slight modification of the proof of Theorem 4.2 using Jensen's inequality shows that the convergence of the Gram-Charlier expansion is still ensured for $\sigma_r^2 > v_{max}\lambda_{max}^2 T$.*

Proof. As for Theorem 4.1, we will prove the claim in a more general framework to account for general option pricing models. We retake the notations introduced for the dynamics (4.4) and we indicate the differences below. Under appropriate probability measure $\mathbb{P}^*$, f_T will denote the density of S_T whose evolution over $[0,T]$ is given by the following dynamics

$$\begin{aligned} \mathrm{d}S_t &= \mu(t,V_t)\mathrm{d}t + \sqrt{Q(V_t)}\rho(t)\|\boldsymbol{\lambda}(t)\|\mathrm{d}W_t + \sqrt{V_t - \rho(t)^2 Q(V_t)}\boldsymbol{\lambda}(t)\cdot \mathrm{d}\boldsymbol{W}_t^*, \\ \mathrm{d}V_t &= \kappa\big(\theta - \xi(t)V_t\big)\mathrm{d}t + \epsilon\sqrt{Q(V_t)}\mathrm{d}W_t, \end{aligned} \tag{4.12}$$

where we recall that $\mu : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ is a regular function such that $|\mu(t,x)| \leq c(1+|x|)$ for some positive constant c, $[0,T] \ni t \mapsto \rho(t) \in [-1,1]$ accounts for a scaling factor of the instantaneous correlation between S and V, Q is the bounding function defined in Section 4.3.1, ξ is a bounded function, $\kappa > 0$, $\theta \in [\xi_{max}v_{min}, \xi_{min}v_{max}]$, $\epsilon > 0$. The cumulated variance process

$(\Xi_t)_{0\le t\le T}$ is defined as in (4.10) by replacing $\boldsymbol{\lambda}^{m,n}$ with a d-dimensional function $\boldsymbol{\lambda}$.

The proof of the theorem relies on the following lemma.

Lemma 4.3. *For $\eta < 1/(2Tv_{max}\lambda_{max}^2)$, f_T is such that*

$$\int_{\mathbb{R}} e^{\eta u^2} f_T(u)\mathrm{d}u < \infty.$$

In addition, if

$$\mathbb{E}^*\left[\Xi_T^{-1/2}\right] < \infty \tag{4.13}$$

holds, then then $u \longmapsto f_T(u)$ and $u \longmapsto e^{\eta u^2} f_T(u)$ are uniformly bounded over $\mathbb{R}$.

Proof of Lemma 4.3. The proof is an an adaptation of Theorem 3.1 in [AFP17] and an extension of it in the sense that it accounts for time dependent coefficient. Let us denote by $\tilde{S}_T := S_0 + \int_0^T \mu(u, V_u)\mathrm{d}u + \frac{1}{\epsilon}\int_0^T \rho(u)\|\boldsymbol{\lambda}(u)\|\mathrm{d}V_u - \frac{\kappa}{\epsilon}\int_0^T \rho(u)\|\boldsymbol{\lambda}(u)\|\big(\theta - \xi(u)V_u\big)\mathrm{d}u$. We have from Equation (4.12) the distributional equality

$$S_T \overset{a.s.}{=} \tilde{S}_T + \int_0^T \sqrt{V_u - \rho(u)^2 Q(V_u)}\boldsymbol{\lambda}(u)\cdot \mathrm{d}\boldsymbol{W}_u^*.$$

Since $\boldsymbol{W}^*$ and V are independent, S_T is a Gaussian variable of mean $\tilde{S}_T$ and variance Ξ_T conditionally to $\sigma(V_t, t\le T)$. Since $\lambda_{min} > 0$ and $\sup_{t\in[0,T]}|\rho(t)|^2 < 1$, note that $\Xi_T > 0$ almost surely. The conditional density of S_T writes

$$F_T(u) = \frac{1}{\sqrt{2\pi\Xi_T}}\exp\left(-\frac{(u-\tilde{S}_T)^2}{2\Xi_T}\right).$$

Fubini's theorem allows to identify $f_T(u) = \mathbb{E}^*\left[F_T(u)\right]$. Hence, f_T is uniformly bounded as long as $\mathbb{E}^*\left[\frac{1}{\sqrt{\Xi_T}}\right] < \infty$. The maps $[0,T]\ni t \longmapsto \|\boldsymbol{\lambda}(t)\|$ and $[0,T]\ni t\longmapsto \rho(t)$ are piecewise constant, so their derivatives are almost everywhere zero. They also have a finite number of discontinuities over $[0,T]$ occurring at times $(T_i)_{1\le i\le n(T)}$ where $n(T) = \max\{i : T_i \le T\}$. Integrating with respect to the volatility process V over interval $[0,T]$ gives $\tilde{S}_T \overset{a.s.}{=} S_0 + \int_0^T \mu(u,V_u)\mathrm{d}u + \frac{1}{\epsilon}\sum_{i\le n(T)}\rho(T_i)\|\boldsymbol{\lambda}(T_i)\|\big(V_{T_{i+1}} - V_{T_i}\big) - \frac{\kappa}{\epsilon}\int_0^T \rho(u)\|\boldsymbol{\lambda}(u)\|\big(\theta - \xi(u)V_u\big)\mathrm{d}u$. As a consequence, we have almost surely: $|\tilde{S}_T| \le S_0 + cT(1+v_{max}) + \frac{2Tv_{max}}{\epsilon}\lambda_{max} + \frac{T\kappa}{\epsilon}\lambda_{max}\big(\theta + \sup_{0\le u\le T}|\xi(u)|v_{max}\big) =: \Lambda < \infty$. Moreover, since $|\Xi_T| \le Tv_{max}\lambda_{max}^2$, we set

$$\nu := 1 - 2\eta T v_{max}\lambda_{max}^2 \le 1 - 2\eta\Xi_T,$$

where $\nu > 0$ since we assumed $\eta < 1/(2Tv_{max}\lambda_{max}^2)$. Furthermore, one can obtained by completing the squares

$$e^{\eta u^2}F_T(u) = \frac{1}{\sqrt{2\pi\Xi_T}}\exp\left(-\frac{1-2\eta\Xi_T}{2\Xi_T}\Big(u - \frac{\tilde{S}_T}{1-2\eta\Xi_T}\Big)^2 + \frac{\eta\tilde{S}_T^2}{1-2\eta\Xi_T}\right). \tag{4.14}$$

Integration of (4.14) gives $\int_{\mathbb{R}} e^{\eta u^2}F_T(u)\mathrm{d}u = \frac{1}{\sqrt{1-2\eta\Xi_T}}\exp\left(\frac{\eta\tilde{S}_T^2}{1-2\eta\Xi_T}\right) \le \frac{e^{\eta\Lambda^2/\nu}}{\sqrt{\nu}} < \infty$ almost surely. In addition, taking expectation in (4.14) leads to $e^{\eta u^2}f_T(u) = \mathbb{E}^*\left[e^{\eta u^2}F_T(u)\right] \le$

$\mathbb{E}^*\left[\frac{1}{\sqrt{\Xi_T}}\right]\frac{e^{\eta\Lambda^2/\nu}}{\sqrt{2\pi}}$. Thus, $e^{\eta u^2}f_T(u)$ is uniformly bounded as soon as (4.13) holds and the Lemma is proved. □

We can now prove Theorem 4.2. We proceed in two steps. First, we show that (4.13) holds under Assumption 1 and Assumption 2, for any $0 \le v_{min}$ and $v_{max} < \infty$; secondly, we explain why (4.13) implies (3.54) when (4.11) is satisfied.

The treatment of the case $v_{min} > 0$ is done in [AFP17]. We focus here on case when $v_{min} = 0$. Almost surely, the following stands

$$\lambda_{min}^2 \int_0^T \left(V_s - \rho(s)^2 Q(V_s)\right)\mathrm{d}s \le \Xi_T \le \lambda_{max}^2 \int_0^T \left(V_s - \rho(s)^2 Q(V_s)\right)\mathrm{d}s.$$

Since $\mathbb{R}_+^* \ni x \longmapsto \frac{1}{\sqrt{x}}$ is convex, Jensen inequality implies

$$\frac{1}{\sqrt{\frac{1}{T}\int_0^T \|\boldsymbol{\lambda}(s)\|^2 \left(V_s - \rho(s)^2 Q(V_s)\right)\mathrm{d}s}} \le \frac{1}{T\lambda_{min}} \int_0^T \frac{\mathrm{d}s}{\sqrt{V_s - \rho(s)^2 Q(V_s)}} \text{ a.s..}$$

Taking expectation and applying Fubini theorem lead to

$$\mathbb{E}^*\left[\Xi_T^{-1/2}\right] \le \frac{1}{\lambda_{min} T^{3/2}} \int_0^T \mathbb{E}^*\left[\frac{1}{\sqrt{V_s - \rho(s)^2 Q(V_s)}}\right]\mathrm{d}s$$

with possibly infinite values. Since $Q(v) \le v$, for any $t \ge 0$, $\sqrt{V_s(1-\rho(t)^2)} \le \sqrt{V_s - \rho(t)^2 Q(V_s)}$ a.s. and thus

$$\mathbb{E}^*\left[\Xi_T^{-1/2}\right] \le (1 - \sup_{t\in[0,T]} \rho(t)^2)^{-1/2} T^{-3/2} \lambda_{min}^{-1} \int_0^T \mathbb{E}^*\left[\frac{1}{\sqrt{V_s}}\right]\mathrm{d}s. \tag{4.15}$$

In order to exhibit a control of the right hand side in (4.15), we consider the stopped process $X_t^{\tau_n} := \sqrt{V_{t\wedge\tau_n}}$ where τ_n, $n \in \mathbb{N}^*$, is the first time when the volatility process goes below the threshold $1/n$: $\tau_n = \inf\{t \ge 0 : V_t \le \frac{1}{n}\}$. Itô's lemma applied to the stopped process gives

$$\begin{aligned} X_T^{\tau_n} \overset{a.s.}{=} \sqrt{V_0} + \left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\int_0^{T\wedge\tau_n} \frac{\mathrm{d}u}{\sqrt{V_u}} + \int_0^{T\wedge\tau_n} \left(\frac{\epsilon^2}{8v_{max}} - \frac{\kappa\xi(s)}{2}\right)\sqrt{V_u}\mathrm{d}u \\ + \frac{\epsilon}{2}\int_0^{T\wedge\tau_n} \sqrt{1 - \frac{V_u}{v_{max}}}\mathrm{d}W_u, \end{aligned} \tag{4.16}$$

where we used that $Q(v) = v - \frac{v^2}{v_{max}}$ when $v_{min} = 0$. Itô's integral is a martingale since $1 - \frac{v}{v_{max}} \le 1$ for $v \in [0, v_{max}]$ and thus vanishes when taking the expectation in (4.16):

$$\mathbb{E}^*\left[\sqrt{V_{T\wedge\tau_n}} - \sqrt{V_0}\right] - \mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\left(\frac{\epsilon^2}{8v_{max}} - \frac{\kappa\xi(s)}{2}\right)\sqrt{V_s}\mathrm{d}s\right] = \left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\frac{\mathrm{d}s}{\sqrt{V_s}}\right]. \tag{4.17}$$

First recall that $V_0 \in [0, v_{max}]$ (almost surely) and thus $\mathbb{E}\left[\sqrt{V_0}\right] < \infty$. Now, we aim at taking the limit as n goes to ∞ in the previous equality to show that the right-hand side is finite. Since $\tau_n \xrightarrow[n\to\infty]{\text{a.s.}} \tau_0 \overset{a.s.}{=} \infty$ where τ_0 is the first time when the volatility process hits 0, $T \wedge \tau_n \xrightarrow[n\to\infty]{\text{a.s.}} T$. $\sqrt{V_t} \le \sqrt{v_{max}}$ for any $t \ge 0$, and thus $\mathbb{E}^*\left[\sqrt{V_{T\wedge\tau_n}}\right] \le \sqrt{v_{max}}$ for all $n \in \mathbb{N}$. Lebesgue's dominated convergence theorem provides that $\mathbb{E}^*\left[\sqrt{V_T}\right] = \lim_n \mathbb{E}^*\left[\sqrt{V_{T\wedge\tau_n}}\right] \le \sqrt{v_{max}}$. Then,

we recall that ξ is bounded on $[0,T]$ which gives

$$\left|\mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\left(\frac{\epsilon^2}{8v_{max}}-\frac{\kappa\xi(s)}{2}\right)\sqrt{V_s}\mathrm{d}s\right]\right|\leq\left(\frac{\epsilon^2}{8v_{max}}+\frac{\kappa\sup_{t\leq T}|\xi(t)|}{2}\right)\sqrt{v_{max}}T=:K_1<+\infty.$$

Using monotone convergence theorem, we have that

$$\int_0^{T\wedge\tau_n}\sqrt{V_u}\mathrm{d}u\xrightarrow[n\to\infty]{\text{a.s.}}\int_0^T\sqrt{V_u}\mathrm{d}u\text{ and }\int_0^{T\wedge\tau_n}\xi(u)\sqrt{V_u}\mathrm{d}u\xrightarrow[n\to\infty]{\text{a.s.}}\int_0^T\xi(u)\sqrt{V_u}\mathrm{d}u.$$

This combined with dominated convergence theorem gives

$$\mathbb{E}^*\left[\int_0^T\left(\frac{\epsilon^2}{8v_{max}}-\frac{\kappa\xi(s)}{2}\right)\sqrt{V_s}\mathrm{d}s\right]=\lim_{n\to\infty}\mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\left(\frac{\epsilon^2}{8v_{max}}-\frac{\kappa\xi(s)}{2}\right)\sqrt{V_s}\mathrm{d}s\right]\leq K_1<\infty.$$

Observe that Assumption 1 implies $\left(\frac{\kappa\theta}{2}-\frac{\epsilon^2}{8}\right)>0$. Together with (4.17) gives the existence of a constant $C<\infty$ such that

$$\lim_{n\to\infty}\mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\frac{\mathrm{d}u}{\sqrt{V_u}}\right]\leq C.$$

Fatou's lemma provides

$$\mathbb{E}^*\left[\int_0^T\frac{\mathrm{d}u}{\sqrt{V_u}}\right]\leq\lim_{n\to\infty}\mathbb{E}^*\left[\int_0^{T\wedge\tau_n}\frac{\mathrm{d}u}{\sqrt{V_u}}\right]<\infty,$$

which allows to deduce the claimed result

$$\mathbb{E}^*\left[\Xi_T^{-1/2}\right]<\infty.$$

Now, following Lemma 4.3, as long as (4.13) holds, $x\in\mathbb{R}\longmapsto e^{\eta x^2}f_T(x)$ is uniformly bounded and integrable for $\eta<1/(2Tv_{max}\lambda_{max}^2)$: $e^{\eta x^2}f_T(x)\leq C'<\infty$ and $\int_{\mathbb{R}}e^{\eta x^2}f_T(x)\mathrm{d}x<\infty$. We apply this result with $\eta=1/(4\sigma_r^2)<1/(2Tv_{max}\lambda_{max}^2)\iff\sigma_r^2>v_{max}\lambda_{max}^2T/2$:

$$\int_{\mathbb{R}}\frac{f_T(x)^2}{g_r(x)}\mathrm{d}x\quad=\quad\sqrt{2\pi\sigma_r^2}\int_{\mathbb{R}}\left(e^{\frac{x^2}{4\sigma_r^2}}f_T(x)\right)^2\mathrm{d}x\leq C'\sqrt{2\pi\sigma_r^2}\int_{\mathbb{R}}e^{\frac{x^2}{4\sigma_r^2}}f_T(x)\mathrm{d}x<\infty.$$

The claim is proved. □

4.3.2.2 Application to pricing of swap rate derivatives

The convergence of approximating densities built with Gram-Charlier method is ensured in $\mathcal{L}_r^2$ when (4.11) holds. An application of the Cauchy-Schwarz inequality shows that the convergence of approximating prices can be deduced for square-integrable payoffs. Let us consider a (discounted) payoff $\varphi\in\mathcal{L}_r^2$ and f_T the density function of $S_T^{m,n,J}$ that is modelled with dynamics (4.7). The spot price of a European derivative $P_T(\varphi)$ expiring at time $T>0$ associated to the considered payoff can be computed thanks to the likelihood ratio $\bar{f}_T=f_T/g_r$:

$$P_T(\varphi)=\int_{\mathbb{R}}\varphi(x)f_T(x)\mathrm{d}x=\left\langle\varphi,\bar{f}_T\right\rangle_{\mathcal{L}_r^2}=\sum_{p\geq0}h_p\varphi_p,\tag{4.18}$$

where the $(\varphi_p)_{p\in\mathbb{N}}$ are coefficients (Fourier coefficients in $\mathcal{L}^2_r$) of φ given by

$$\varphi_p = \left\langle \varphi, \check{H}_p \right\rangle_{\mathcal{L}^2_r},$$

and the *Hermite moments* $(h_p)_{p\in\mathbb{N}}$ are defined by

$$h_p = \left\langle \bar{f}_T, \check{H}_p \right\rangle_{\mathcal{L}^2_r} = \int_{\mathbb{R}} \check{H}_p(x) f_T(x)\mathrm{d}x = \mathbb{E}^S\left[\check{H}_p\big(S_T^{m,n,J}\big)\right].$$

The $\left(\check{H}_p\right)_{p\in\mathbb{N}}$ denotes the Hermite polynomials defined in (4.3). The coefficients h_p are linear combinations of the moments of f_T. Since the polynomial functions H_p are analytically known (see 3.41), the only matter here is to be able to compute the moments of f_T. This can be achieved using the polynomial property of (4.7) discussed in Section 4.3.1. Once the Gram-Charlier convergence is ensured, approximating price can be built as

$$P_T^N(\varphi) = \sum_{p=0}^{N} h_p \varphi_p \tag{4.19}$$

and is so that $P_T^N(\varphi) \xrightarrow[N\to\infty]{} P_T(\varphi)$. Numerically, we have to truncate (4.18) at a given order $N \in \mathbb{N}$. The estimation of the speed of convergence of the series as N increases is discussed below in Section 4.4.2; note that [AFP17] provided a numerical estimation of the truncation error.

4.4 Rates of convergence

Dynamics (4.7) is use in our point of view as an approximating dynamics of (4.1). Derivatives prices induced by this dynamics are computed through Gram-Charlier expansion as explained above. In practice, one has to truncation the Gram-Charlier series. This way of proceeding induces two types of errors: the one coming from to the introduction of the Jacobi process instead of the CIR one, or in other words, the error due to the *bounding* of the volatility factor and the one due to the *truncation* of the Gram-Charlier series.

Let us consider the problem of pricing a swap rates derivatives of maturity $T \leq T_m$ and of payoff function φ. The price under (4.1) is assumed to be exactly known so that the global error induced by our methodology can be decomposed as

$$\begin{aligned}
\varepsilon &= \left|\mathbb{E}^S\left[\varphi(S_T^{m,n})\right] - P_T^N(\varphi)\right| \\
&\leq \left|\mathbb{E}^S\left[\varphi(S_T^{m,n})\right] - \mathbb{E}^S\left[\varphi(S_T^{m,n,J})\right]\right| + \left|\mathbb{E}^S\left[\varphi(S_T^{m,n,J})\right] - P_T^N(\varphi)\right| \\
&=: \varepsilon_1 + \varepsilon_2.
\end{aligned} \tag{4.20}$$

The purpose of the next sections will be to give some insights on the estimations of both errors.

4.4.1 Bounding error

We consider a sequence of bounds parameters $(v_{min}^{(p)}, v_{max}^{(p)})_{p\in\mathbb{N}} \in (\mathbb{R}_+^* \times \mathbb{R}_+^*)^{\mathbb{N}}$ such that $v_{min}^{(p)} \xrightarrow[p\to\infty]{} 0$ and $v_{max}^{(p)} \xrightarrow[p\to\infty]{} +\infty$. We recall that $Q(v) \to v$ as $(v_{min}^{(p)}, v_{max}^{(p)}) \to (0,\infty)$. We will denote by $v_{min}^{(p)} \equiv 0$ the particular case when for all $p \in \mathbb{N}$, $v_{min}^{(p)} = 0$ which yields that $Q(v) = v - \frac{v^2}{v_{max}^{(p)}}$; and by $v_{max}^{(p)} \equiv +\infty$ the case when for all $p \in \mathbb{N}$, $v_{max}^{(p)} = +\infty$ which corresponds to $Q(v) =$

$v - v_{min}^{(p)}$. $(S_t^{m,n,ref}, V_t^C)_{0\le t\le T}$ represents the solution of the reference dynamics (4.1) starting from $(S_0^{m,n,ref}, V_0^C) = (s_0^{m,n}, v_0) \in \mathbb{R}\times\mathbb{R}_+^*$. In particular, $(V_t^C)_{0\le t\le T}$ is a CIR process. For each $p\in\mathbb{N}$, let us denote by $(S_t^{m,n,J(p)}, V_t^{J(p)})_{0\le t\le T}$ the solution of (4.7) associated to the bound parameters $(v_{min}^{(p)}, v_{max}^{(p)})$ and starting from $(S^{m,n,J(p)}, V_0^{J(p)}) = (s_0^{m,n}, v_0)$. We require $\max_{p\in\mathbb{N}} v_{min}^{(p)} \le v_0 \le \inf_{p\in\mathbb{N}} v_{max}^{(p)}$ so that $(S^{m,n,J(p)}, V^{J(p)})$ is well defined for any $p\in\mathbb{N}$. The following result shows that the marginal moments of the Jacobi process can be bounded independently from the parameters $v_{max}^{(p)}$.

Lemma 4.4. *Assume either $v_{min}^{(p)} \equiv 0$ or $v_{max}^{(p)} \equiv +\infty$. Then for any $k\in\mathbb{N}$,*

$$\sup_{p\in\mathbb{N}}\sup_{0\le t\le T} \mathbb{E}^S\left[\left(V_t^{J(p)}\right)^k\right] \le C_k.$$

Proof. First consider case when $v_{min}^{(p)} \equiv 0$. Fix a $p\in\mathbb{N}$. We first observe that for any time $t\ge 0$, $\mathbb{E}^S[V_t^{J(p)}] = \mathbb{E}^S[V_t^C]$ does not depend on $v_{max}^{(p)}$ since the deterministic drifts in dynamics of $(V_t^{J(p)})_{t\le T}$ and $(V_t^C)_{t\le T}$ are the same. The claim is clear for $k=1$. Take $k=2$. Applying Itô's formula to the squared process $\left((V_t^{J(p)})^2\right)_{t\le T}$ and taking the expectation lead to

$$\begin{aligned}\mathbb{E}^S\left[\left(V_t^{J(p)}\right)^2\right] &= v_0^2 + \int_0^t \Big\{2\kappa\theta\mathbb{E}^S\left[V_s^C\right] - 2\kappa\xi(s)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^2\right] \\ &\qquad + \epsilon^2\Big(\mathbb{E}^S\left[V_s^C\right] - \mathbb{E}^S\left[\left(V_s^{J(p)}\right)^2\right]/v_{max}^{(p)}\Big)\Big\}\mathrm{d}s \\ &\le v_0^2 + \int_0^t \Big\{(2\kappa\theta+\epsilon^2)\mathbb{E}^S\left[V_s^C\right] - 2\kappa\xi(s)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^2\right]\Big\}\mathrm{d}s.\end{aligned}$$

Gronwall's lemma gives

$$\mathbb{E}^S\left[(V_t^{J(p)})^2\right] \le \Big\{v_0^2 + \int_0^t (2\kappa\theta+\epsilon^2)\mathbb{E}^S[V_s^C]\mathrm{d}s\Big\}\exp\Big(-2\kappa\int_0^t \xi(s)\mathrm{d}s\Big)$$

and the claim is true for the constant $C_2 = \Big(v_0^2 + \int_0^T (2\kappa\theta+\epsilon^2)\mathbb{E}^S[V_s^C]\mathrm{d}s\Big)\exp\Big(2\kappa\|\xi\|_\infty T\Big)$ that does not depend on $v_{max}^{(p)}$ nor on $t\le T$. Now assume that

$$\forall p\in\mathbb{N},\ \forall t\le T,\ \mathbb{E}^S\left[(V_t^{J(p)})^{k-1}\right] \le C_{k-1}$$

holds for a given $k\in\mathbb{N}\setminus\{0,1\}$. Applying Itô's lemma again to $\left((V_t^{J(p)})^k\right)_{t\ge 0}$ and taking expectation lead to, for $t\le T$,

$$\begin{aligned}\mathbb{E}^S[(V_t^{J(p)})^k] &= v_0^k + \int_0^t \Bigg\{\left(2k\kappa\theta + k(k-1)\frac{\epsilon^2}{2}\right)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^{k-1}\right] \\ &\qquad - \left(2\kappa\xi(s) + \frac{\epsilon^2 k(k-1)}{2v_{max}^{(p)}}\right)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^k\right]\Bigg\}\mathrm{d}s \\ &\le v_0^k + \int_0^t \Bigg\{\left(2k\kappa\theta + k(k-1)\frac{\epsilon^2}{2}\right)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^{k-1}\right] - 2\kappa\xi(s)\mathbb{E}^S\left[\left(V_s^{J(p)}\right)^k\right]\Bigg\}\mathrm{d}s.\end{aligned}$$

Gronwall's lemma again shows that

$$\mathbb{E}^S[(V_t^{J(p)})^k] \quad\le\quad \Big[(v_0)^k + t\big(2k\kappa\theta + k(k-1)\frac{\epsilon^2}{2}\big)C_{k-1}\Big]\exp\Big(-\int_0^t 2\kappa\xi(s)\mathrm{d}s\Big).$$

The result holds then for k-th order moment of the Jacobi process with bound $C_k = \Big(v_0^k + T\big(2k\kappa\theta + k(k-1)\frac{\epsilon^2}{2}\big)C_{k-1}\Big)\exp\Big(2\kappa\|\xi\|_\infty T\Big)$. The proof in case $v_{max}^{(p)} \equiv +\infty$ is similar. □

This result is useful to get a rate of convergence in the $\mathbb{L}^1$ space and in a weak sense of the Jacobi process towards the CIR one.

4.4.1.1 Strong convergence for the Jacobi process

We stress that the same Brownian motion $(W_t)_{0\le t\le T}$ is used to define the Cox-Ingersoll-Ross process in Equation (4.1) and Jacobi process in Equation (4.7). In this perfect coupling, we can now prove our result on the strong convergence of the Jacobi process towards the CIR one and derive a convergence rate.

Theorem 4.5. *In case when $v_{min}^{(p)} \equiv 0$, there exists finite constants $C', C'' \in \mathbb{R}_+^*$ such that for any $p \in \mathbb{N}$,*

$$\sup_{0\le t\le T} \mathbb{E}^S\big[|V_t^{J(p)} - V_t^C|\big] \le C' \big/ \log\big(v_{max}^{(p)}/v_0\big),$$

and

$$\mathbb{E}^S\Big[\sup_{0\le t\le T} |V_t^{J(p)} - V_t^C|\Big] \le C'' \Big/ \sqrt{\log\big(v_{max}^{(p)}/v_0\big)}.$$

In case when $v_{max}^{(p)} \equiv +\infty$, there exists finite constants $C^*, C^{**} \in \mathbb{R}_+^*$ such that for any $p \in \mathbb{N}$,

$$\sup_{0\le t\le T} \mathbb{E}^S\big[|V_t^{J(p)} - V_t^C|\big] \le C^* \big/ \log\big(v_0/v_{min}^{(p)}\big),$$

and

$$\mathbb{E}^S\Big[\sup_{0\le t\le T} |V_t^{J(p)} - V_t^C|\Big] \le C^{**} \Big/ \sqrt{\log\big(v_0/v_{min}^{(p)}\big)}.$$

Proof. In this proof, we draw from A. Alfonsi in [Alf06], Section 4.3, who studied discretization scheme for CIR process and proved strong convergence.

First consider the case $v_{min}^{(p)} \equiv 0$. Let us consider the family of Yamada functions, $\psi_{\eta,m}$, parametrized by two positive numbers η and m. Observe that for any $m \ge 1$, $\int_{\eta e^{-\epsilon^2 m}}^{\eta} \frac{1}{\epsilon^2 u}\mathrm{d}u = m$. Then, we can find a continuous function $\rho_{\eta,m}$, with compact support included in $]\eta e^{-\epsilon^2 m}, \eta[$, and such that $\rho_{\eta,m}(x) \le \frac{2}{\epsilon^2 x m}\mathbb{1}_{x\in]\eta e^{-\epsilon^2 m},\eta[}$ and $\int_{\eta e^{-\epsilon^2 m}}^{\eta} \rho_{\eta,m}(u)\mathrm{d}u = 1$. Then, $\psi_{\eta,m}$ is defined as

$$\psi_{\eta,m}(x) = \int_0^{|x|}\int_0^y \rho_{\eta,m}(u)\mathrm{d}u\mathrm{d}y,$$

so that these functions satisfy

$$|x| - \eta \le \psi_{\eta,m}(x) \le |x|,\ |\psi'_{\eta,m}(x)| \le 1,\ 0 \le \psi''_{\eta,m}(x) = \rho_{\eta,m}(|x|) \le \frac{2}{\epsilon^2|x|m}.$$

$\psi_{\eta,m}$ is a nice smooth (twice differentiable) and even function approximating the absolute value function. Take $p \in \mathbb{N}$. Since

$$|V_t^{J(p)} - V_t^C| \le \eta + \psi_{\eta,m}(V_t^{J(p)} - V_t^C), \tag{4.21}$$

we apply Itô's formula to the right-hand side to get

$$\begin{aligned}\psi_{\eta,m}(V_t^{J(p)} - V_t^C) \overset{a.s.}{=} &- \kappa \int_0^t (V_s^{J(p)} - V_s^C)\xi(s)\psi'_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \\ &+ \frac{\epsilon^2}{2}\int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \\ &+ \epsilon \int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)\psi'_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}W_s.\end{aligned} \tag{4.22}$$

Using that $\sup_{s\geq 0}\mathbb{E}^*[V_s^C] < \infty$, $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x-y|}$ (due to $\frac{1}{2}$-Hölder regularity of the square-root) and $|\psi'_{\eta,m}(x)| \leq 1$ lead to

$$\left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \psi'_{\eta,m}(V_s^{J(p)} - V_s^C) \leq \left|Q(V_s^{J(p)}) - V_s^C\right| \leq v_{max}^{(p)} + V_s^C \text{ a.s.}$$

which allows to deduce that the expectation of the Itô integral in (4.22) is zero. Using again $\frac{1}{2}$-Hölder regularity of the square-root, the second term in the right-hand side of (4.22) can be decomposed in the following way

$$\begin{aligned}&\int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \\ &\leq \int_0^t \left|Q(V_s^{J(p)}) - V_s^C\right| \psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \\ &\leq \int_0^t \left\{|Q(V_s^{J(p)}) - V_s^{J(p)}| + |V_s^{J(p)} - V_s^C|\right\}\psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s.\end{aligned} \tag{4.23}$$

The first term in this integral is handled using that $\|\psi''_{\eta,m}\|_\infty \leq \frac{2e^{\epsilon^2 m}}{\epsilon^2 \eta m}$ and

$$|Q(v) - v| \leq v^2/v_{max}^{(p)}. \tag{4.24}$$

The second term is bounded with $|x|\psi''_{\eta,m}(x) \leq \frac{2}{\epsilon^2 m}$. This leads to the almost sure inequality

$$\int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \leq \int_0^t \left\{\frac{(V_s^{J(p)})^2}{v_{max}^{(p)}}\frac{2e^{\epsilon^2 m}}{\epsilon^2 \eta m} + \frac{2}{\epsilon^2 m}\right\}\mathrm{d}s.$$

Finally, for the first integral in (4.22), we use that $\|\psi'_{\eta,m}\|_\infty \leq 1$ and thus

$$\left| - \kappa \int_0^t \xi(s)(V_s^{J(p)} - V_s^C)\psi'_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s\right| \quad \leq \quad \kappa \int_0^t |V_s^{J(p)} - V_s^C| \|\xi\|_\infty \mathrm{d}s \text{ a.s..}$$

Taking the expectation in (4.21) using the inequalities that have just been derived and using the fact that the moments of the Jacobi process are uniformly bounded with respect to $v_{max}^{(p)}$ (cf. Lemma 4.4), we get that

$$\mathbb{E}^S\left[|V_t^{J(p)} - V_t^C|\right] \quad \leq \quad \eta + \kappa\|\xi\|_\infty \int_0^t \mathbb{E}^S[|V_s^{J(p)} - V_s^C|]\mathrm{d}s + \frac{C'Te^{\epsilon^2 m}}{v_{max}^{(p)}\eta m} + \frac{T}{m}.$$

for some positive constant C'. Gronwall's lemma again shows that

$$\mathbb{E}^S\left[|V_t^{J(p)} - V_t^C|\right] \quad \leq \quad e^{\kappa\|\xi\|_\infty T}\left(\eta + \frac{C'Te^{\epsilon^2 m}}{v_{max}^{(p)}\eta m} + \frac{T}{m}\right).$$

Setting $\eta = \frac{1}{\log(v_{max}^{(p)}/v_0)}$ and $m = \frac{\log(v_{max}^{(p)}/v_0)}{2\epsilon^2}$, we get the existence of a constant $\tilde{C} \in \mathbb{R}_+$ satisfying

$$\sup_{0\leq t\leq T} \mathbb{E}^S\left[|V_t^{J(p)} - V_t^C|\right] \leq \tilde{C}/\log\left(v_{max}^{(p)}/v_0\right). \tag{4.25}$$

The Itô's integral in the right-hand side of (4.22) can be more accurately handled. We have

$$\begin{aligned}
&\mathbb{E}^S\left[\sup_{0\leq t\leq T}\int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)\psi'_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}W_s\right] \\
&\leq C_{BDG}\mathbb{E}^S\left[\left(\int_0^T \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \|\psi'_{\eta,m}\|_\infty^2 \mathrm{d}s\right)^{1/2}\right] \\
&\leq C_{BDG}\sqrt{\int_0^T \left(\mathbb{E}^S\left[|Q(V_s^{J(p)}) - V_s^{J(p)}|\right] + \mathbb{E}^S\left[|V_s^{J(p)} - V_s^C|\right]\right)\mathrm{d}s} \\
&\leq C_{BDG}\sqrt{\left(\frac{C_2}{v_{max}^{(p)}/v_0} + \frac{\tilde{C}}{\log(v_{max}^{(p)}/v_0)}\right)T}
\end{aligned} \tag{4.26}$$

where we used successively: Burkholder-Davis-Gundy inequality, $\|\psi'_{\eta,m}\|_\infty \leq 1$, Jensen's one, $\frac{1}{2}$-Hölder regularity of the square-root and triangle inequality. This allows to deduce (4.26) after using inequality (4.24) combined with Lemma 4.4 for the first term in the square-root and the previous result (4.25) for the second term. Combining this with previous inequalities used in this proof leads to the existence of a constant $C' \in \mathbb{R}$ such that

$$\mathbb{E}^S\left[\sup_{0\leq t\leq T}|V_t^{J(p)} - V_t^C|\right] \leq C'\Big/\sqrt{\log\left(v_{max}^{(p)}/v_0\right)}.$$

In a second time, consider that $v_{max}^{(p)} \equiv +\infty$. The reasoning is similar but needs to be adapted. In (4.23), one obtains now that

$$\begin{aligned}
&\int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C}\right)^2 \psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s \\
&\leq \int_0^t \left\{v_{min}^{(p)} + \left|V_s^{J(p)} - V_s^C\right|\right\}\psi''_{\eta,m}(V_s^{J(p)} - V_s^C)\mathrm{d}s,
\end{aligned}$$

which eventually leads to, for some positive constant $\tilde{C}$ and at any time $0 \leq t \leq T$:

$$\mathbb{E}^S[|V_t^{J(p)} - V_t^C|] \quad \leq \quad \tilde{C}e^{\kappa\|\xi\|_\infty T}\left(\eta + \frac{e^{\epsilon^2 m}}{\eta m}v_{min}^{(p)} + \frac{1}{m}\right).$$

Setting now the parameters of the Yamada function as $\eta = \frac{1}{\log(v_0/v_{min}^{(p)})}$ and $m = \frac{\log\left(v_0/v_{min}^{(p)}\right)}{\alpha\epsilon^2}$

for some $\alpha > 1$. Using that $(v_{min}^{(p)})^{1-1/\alpha} + \frac{1}{\log\left(v_0/v_{min}^{(p)}\right)} \underset{v_{min}^{(p)} \to 0^+}{\sim} \frac{1}{\log\left(v_0/v_{min}^{(p)}\right)}$ we get that for some positive constant C^*:

$$\sup_{0 \le t \le T} \mathbb{E}^S \left[|V_t^{J(p)} - V_t^C| \right] \le C^* / \log\left(v_0/v_{min}^{(p)}\right).$$

Regarding the second inequality, the use of Burkholder-Davis-Gundy inequality yields now

$$\mathbb{E}^S \left[\sup_{0 \le t \le T} \int_0^t \left(\sqrt{Q(V_s^{J(p)})} - \sqrt{V_s^C} \right) \psi'_{\eta,m}(V_s^{J(p)} - V_s^C) \mathrm{d}W_s \right]$$
$$\le C_{BDG} \sqrt{\left(v_{min}^{(p)} + C^* / \log\left(v_0/v_{min}^{(p)}\right) \right) T}$$

and the claim

$$\mathbb{E}^S \left[\sup_{0 \le t \le T} |V_t^{J(p)} - V_t^C| \right] \le C^{**} / \sqrt{\log\left(v_0/v_{min}^{(p)}\right)}$$

follows using that $\frac{1}{\log\left(v_0/v_{min}^{(p)}\right)} + \frac{1}{\sqrt{\log\left(v_0/v_{min}^{(p)}\right)}} \underset{v_{min}^{(p)} \to 0^+}{\sim} \frac{1}{\sqrt{\log\left(v_0/v_{min}^{(p)}\right)}}$. □

4.4.1.2 Weak convergence for the Jacobi process

In present subsection, we assume that $\xi^0(t) \equiv \xi^0$ is constant through time. We first recall some useful notations. Considering E a subset of $\mathbb{R}^d$. We will denote for $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and for a smooth function f defined over E by $\partial_\alpha f$ the differentiation operator

$$x = (x_1 \dots, x_d) \in E \mapsto \partial_\alpha f(x) = \partial_{x_1}^{\alpha_1} \dots \partial_{x_d}^{\alpha_d} f(x).$$

We then introduce the set of functions with derivatives of polynomial growth:

$$\mathcal{C}_{pol}^\infty(E) = \left\{ f : E \to \mathbb{R} : f \in \mathcal{C}^\infty, \forall \alpha \in \mathbb{N}^d, \exists C_\alpha > 0, e_\alpha \in \mathbb{N}^*, \forall x \in E, |\partial_\alpha f(x)| \le C_\alpha (1 + \|x\|^{e_\alpha}) \right\}.$$

The following result is key for deriving the weak convergence rate of the Jacobi process towards the CIR one. This result is stated in Proposition 3.3.1 of [Alf15].

Proposition 4.6. *Assume that $f \in \mathcal{C}_{pol}^\infty(\mathbb{R})$. The function $u^C(t,x) = \mathbb{E}[f(V_{T-t}^C) | V_0^C = x]$ is well defined over $[0,T] \times \mathbb{R}_+$, is $\mathcal{C}^\infty([0,T] \times \mathbb{R}_+)$ and satisfies the following backward Kolmogorov equation*

$$(P) : \begin{cases} \forall t \in [0,T], \forall x \in \mathbb{R}_+, \ \partial_t u^C + \kappa(\theta - \xi^0 x) \, \partial_x u^C + \frac{\epsilon^2}{2} x \, \partial_x^2 u^C = 0, \\ u^C(T,x) = f(x). \end{cases}$$

Furthermore,

$$\forall (l,m) \in \mathbb{N}^2, \exists (C_{l,m}, e_{l,m}) \in (\mathbb{R}_+^*)^2, \forall x \in \mathbb{R}_+, \forall t \in [0,T],$$
$$|\partial_t^l \partial_x^m u^C(t,x)| \le C_{l,m}(1 + x^{e_{l,m}}).$$

Remark 29. *Proposition 4.6 above is conjectured to hold for piecewise constant parameters ξ^0, θ and ϵ by applying a similar reasonning over each time period on which they take constant values.*

We can now derive our weak convergence result.

Theorem 4.7. *Let us consider $f \in \mathcal{C}^{\infty}_{pol}(\mathbb{R})$. Then there exists a constant $K > 0$ such that for any $p \in \mathbb{N}$,*

$$|\mathbb{E}[f(V_T^C)] - \mathbb{E}[f(V_T^{J(n)})]| \leq \frac{K}{v_{max}^{(p)}}.$$

Proof. Consider u^C as defined in Proposition 4.6; in particular, $u^C(T,x) = f(x)$ for all $x \in \mathbb{R}_+$. We have

$$\mathbb{E}[f(V_T^C)] - \mathbb{E}[f(V_T^{J(p)})] = \mathbb{E}[u^C(T, V_T^C)] - \mathbb{E}[u^C(T, V_T^{J(p)})].$$

Thanks to Proposition 4.6, we have $u^C(0, v_0) = \mathbb{E}\left[u^C(T, V_T^C)\right]$ and thus

$$\mathbb{E}[f(V_T^C)] - \mathbb{E}[f(V_T^{J(p)})] = -\left(\mathbb{E}[u^C(T, V_T^{J(p)}) - u^C(0, v_0)]\right).$$

Itô's formula gives:

$$\begin{aligned}
&u^C(T, V_T^{J(p)}) - u^C(0, v_0) \\
&\overset{a.s.}{=} \int_0^T \Big\{ \partial_t u^C(s, V_s^{J(p)}) + \kappa(\theta - \xi^0 V_s^{J(p)})\, \partial_x u^C(s, V_s^{J(p)}) \\
&\quad + \frac{\epsilon^2 Q(V_s^{J(p)})}{2}\, \partial_x^2 u^C(s, V_s^{J(p)}) \Big\} \mathrm{d}s + \epsilon \int_0^T \sqrt{Q(V_s^{J(p)})}\, \partial_x u^C(s, V_s^{J(p)}) \mathrm{d}W_s.
\end{aligned}$$

Using Proposition 4.6 and fact that moments of the Jacobi process are all finite, we deduce that the Itô's integral above is of null expectation. With Fubini's theorem, we get

$$\begin{aligned}
\mathbb{E}\big[u^C(T, V_T^{J(p)}) - u^C(0, v_0)\big] &= \int_0^T \Big\{ \partial_t u^C(s, V_s^{J(p)}) + \kappa(\theta - \xi^0 V_s^{J(p)})\, \partial_x u^C(s, V_s^{J(p)}) \\
&\quad + \frac{\epsilon^2 Q(V_s^{J(p)})}{2}\, \partial_x^2 u^C(s, V_s^{J(p)}) \Big\} \mathrm{d}s \\
&= \int_0^T \mathbb{E}\big[g^C(s, V_s^{J(p)})\big] \mathrm{d}s
\end{aligned}$$

where $g^C(t,x) = \partial_t u^C(t,x) + \kappa(\theta - \xi^0 x)\, \partial_x u^C(t,x) + \frac{\epsilon^2 Q(x)}{2}\, \partial_x^2 u^C(t,x)$ is defined for every $t \in [0,T]$ and $x \in \mathbb{R}_+$. Since u^C is solution of (P) of Proposition 4.6, we obtain that for all $t \geq 0$,

$$g^C(t, V_t^{J(p)}) = \frac{\epsilon^2}{2}\big(Q(V_t^{J(p)}) - V_t^{J(p)}\big)\, \partial_x^2 u^C(t, V_t^{J(p)}) = -\frac{\epsilon^2}{2 v_{max}^{(p)}} \big(V_t^{J(p)}\big)^2\, \partial_x^2 u^C(t, V_t^{J(p)}),$$

where the bounding function Q is defined by $Q(x) = x - \frac{x^2}{v_{max}^{(p)}}$ when $v_{min}^{(p)} \equiv 0$. Using the estimates on the growth of u^C from Proposition 4.6, we get that $\mathbb{E}\big[g^C(s, V_s^{J(p)})\big] \leq \frac{C'}{v_{max}^{(p)}}$ for some $C' > 0$ and finally

$$\mathbb{E}[f(V_T^C)] - \mathbb{E}[f(V_T^{J(p)})] \leq \frac{C}{v_{max}^{(p)}}.$$

□

Remark 30. *Numerical investigations suggest that this rate of convergence in weak sense can be extended to strong convergence. As discussed in the numerical analysis Section 4.5.4, we*

conjecture that a constant C'' (independent of $\left(v_{max}^{(p)}\right)_{p\geq 0}$) can be found such that

$$\mathbb{E}^S\Big[\sup_{0\leq t\leq T}\left|V_t^{J(p)}-V_t^C\right|\Big]\leq C''/v_{max}^{(p)}.$$

4.4.1.3 Strong convergence of the swap rate process

We are ready to state our estimation on the *bounding* error ε_1. It corresponds to the error made on prices of swap rates derivates when using Jacobi dynamics in place of the CIR one. The following proposition outlines the result for exotic options whose payoff depends on a path of the swap rate process. We consider a payoff φ that is Lipschitz in the following sense: given two continuous processes $(x(t))_{t\in[0,T]}$ and $(y(t))_{t\in[0,T]}$, there exists a constant $C_\varphi>0$ such that $\left|\varphi\left(\left(x(t)\right)_{t\in[0,T]}\right)-\varphi\left(\left(y(t)\right)_{t\in[0,T]}\right)\right|\leq C_\varphi\sup_{0\leq t\leq T}|x(t)-y(t)|$. Recall that the *bounding* error we aim at estimating here is defined for any $p\in\mathbb{N}$ as

$$\varepsilon_1(p):=\left|\mathbb{E}^S\left[\varphi\left(\left(S_t^{m,n,ref}\right)_{t\leq T}\right)\right]-\mathbb{E}^S\left[\varphi\left(\left(S_t^{m,n,J(p)}\right)_{t\leq T}\right)\right]\right|.$$

Lipschitz property implies directly that

$$\varepsilon_1(p)\leq C_\varphi\mathbb{E}^S\left[\sup_{0\leq t\leq T}|S_t^{m,n,ref}-S_t^{m,n,J}|\right]$$

and the derivation of the following result is obtained by bounding the right-hand side in the previous inequality; hence the strong convergence property.

Proposition 4.8. *Assume that $v_{min}^{(p)}\equiv 0$: there exists some constants $K_1\in\mathbb{R}_+^*$ and $K_2\in\mathbb{R}_+^*$ such that for any $p\in\mathbb{N}$*

$$\varepsilon_1(p)\leq\sqrt{\frac{K_1}{\log\left(v_{max}^{(p)}/v_0\right)}+\frac{K_2}{v_{max}^{(p)}/v_0}}.$$

Assume then that $v_{max}^{(p)}\equiv\infty$: there exists some constants $K_1'\in\mathbb{R}_+^*$ and $K_2'\in\mathbb{R}_+^*$ such that for any $p\in\mathbb{N}$

$$\varepsilon_1(p)\leq\sqrt{\frac{K_1'}{\log\left(v_0/v_{min}^{(p)}\right)}+K_2'v_{min}^{(p)}}.$$

Proof. We again prove the result in a more general framework by considering a non null drift function in the swap dynamics, as done in preceding proofs. We additionally assume that this drift function is Lipschitz in the space variable: $|\mu(t,x)-\mu(t,y)|\leq C_\mu|x-y|$. Let us first place in case when $v_{min}^{(p)}\equiv 0$. Similarly to the previous proof, we will successively use triangle inequality, the Burkholder-Davis-Gundy inequality, $\frac{1}{2}$-Hölder regularity of the square-root function ($|\sqrt{x}-\sqrt{y}|\leq\sqrt{|x-y|}$, for $x,y\geq 0$), the fact that for all $t\leq T$, $\rho(t)^2Q(v)\leq$

$\rho(t)^2 v \leq v$ since $\rho(t)^2 \leq 1$ and that the square-root function is concave to get the following:

$$
\begin{aligned}
\varepsilon_1(p) &\leq C_\varphi \mathbb{E}^S \left[\sup_{0\leq t\leq T} |S_t^{m,n,ref} - S_t^{m,n,J}| \right] \\
&= C_\varphi \mathbb{E}^S \Bigg[\sup_{0\leq t\leq T} \Big| \int_0^t \big(\mu(s,V_s^C) - \mu(s,V_s^J)\big)\mathrm{d}s + \int_0^t \rho(s)\Big(\sqrt{V_s^C} - \sqrt{Q(V_s^J)}\Big)\|\boldsymbol{\lambda}^{m,n}(s)\|\mathrm{d}W_s \\
&\quad + \int_0^t \Big(\sqrt{V_s^C - \rho(s)^2 V_s^C} - \sqrt{V_s^J - \rho(s)^2 Q(V_s^J)}\Big)\boldsymbol{\lambda}^{m,n}(s)\cdot \mathrm{d}\boldsymbol{W}_s^{S,*}\Big| \Bigg] \\
&\leq C_\varphi C_\mu \mathbb{E}^S \left[\int_0^T |V_s^C - V_s^J|\mathrm{d}s \right] + C_{\mathrm{Lip}} C_{BDG} \lambda_{max} \Bigg(\mathbb{E}^S \left[\sqrt{\int_0^T \rho(s)^2 |V_s^C - Q(V_s^J)|\mathrm{d}s} \right] \\
&\quad + \mathbb{E}^S \left[\sqrt{\int_0^T |V_s^C - V_s^J + \rho(s)^2\big(Q(V_s^J) - V_s^C\big)|\mathrm{d}s} \right] \Bigg) \\
&\leq C \Bigg(\int_0^T \mathbb{E}^S\Big[|V_s^C - V_s^J|\Big]\mathrm{d}s + \sqrt{\int_0^T \rho(s)^2 \mathbb{E}^S\left[|V_s^C - Q(V_s^J)|\right]\mathrm{d}s} \\
&\quad + \sqrt{\int_0^T \mathbb{E}^S\left[|V_s^C - V_s^J + \rho(s)^2\big(Q(V_s^J) - V_s^C\big)|\right]\mathrm{d}s} \Bigg) \\
&=: C(I_1 + I_2 + I_3)
\end{aligned}
$$

where C is a positive number. I_1 can be bounded using Theorem 4.5: $I_1 \leq T\tilde{C}/\log\left(v_{max}/v_0\right)$. Furthermore, observing that

$$
\begin{aligned}
I_2 &= \left(\int_0^T \rho(s)^2 \mathbb{E}^S\big[|V_s^C - V_s^J + V_s^J - Q(V_s^J)|\big]\mathrm{d}s \right)^{1/2} \\
&\leq \left(\int_0^T \big(\mathbb{E}^S\big[|V_s^C - V_s^J|\big] + \mathbb{E}^S\big[|V_s^J - Q(V_s^J)|\big]\big)\mathrm{d}s \right)^{1/2}
\end{aligned}
$$

and that

$$
\begin{aligned}
I_3 &= \left(\int_0^T \mathbb{E}^S\big[|V_s^C - V_s^J + \rho(s)^2\big(Q(V_s^J) - V_s^C\big)|\big]\mathrm{d}s \right)^{1/2} \\
&\leq \left(\int_0^T \big((1+\rho(s)^2)\mathbb{E}^S\big[|V_s^C - V_s^J|\big] + \rho(s)^2 \mathbb{E}^S\big[|Q(V_s^J) - V_s^J|\big]\big)\mathrm{d}s \right)^{1/2} \\
&\leq \left(\int_0^T \Big(2\mathbb{E}^S\big[|V_s^C - V_s^J|\big] + \mathbb{E}^S\big[|Q(V_s^J) - V_s^J|\big]\Big)\mathrm{d}s \right)^{1/2},
\end{aligned}
$$

we are led to study the quantities $\mathbb{E}^S\big[|V_s^C - V_s^J|\big]$ and $\mathbb{E}^S\big[|Q(V_s^J) - V_s^J|\big]$. Recall that $v - Q(v) = \frac{v^2}{v_{max}}$ for $v_{min} = 0$. Regarding the results of Lemma 4.4 and Theorem 4.5, we can find some positive constants c_1, c_2 such that

$$
I_2 \leq \sqrt{T}\sqrt{\frac{c_1}{\log(v_{max}^{(p)}/v_0)} + \frac{c_2}{v_{max}^{(p)}/v_0}} \text{ and } I_3 \leq \sqrt{T}\sqrt{\frac{2c_1}{\log(v_{max}^{(p)}/v_0)} + \frac{c_2}{v_{max}^{(p)}/v_0}}.
$$

Finally, for some positive constants K_0, K_1 and K_2,

$$\varepsilon_1(p) \leq C' \left(\frac{K_0}{\log(v_{max}^{(p)}/v_0)} + \sqrt{\frac{K_1}{\log(v_{max}^{(p)}/v_0)} + \frac{K_2}{v_{max}^{(p)}/v_0}} \right)$$

where $C' \geq 0$. In the statement of Theorem 4.8, $K_0 = 0$ as there is no drift in the dynamics (4.7).

Regarding the case $v_{max}^{(p)} \equiv +\infty$, the bounding of each term can be adapted so that: $I_1 \leq T\tilde{C}/\log\left(v_0/v_{min}^{(p)}\right)$; $I_2 \leq \left(\frac{C^*T}{\log\left(v_0/v_{min}^{(p)}\right)} + Tv_{min}^{(p)}\right)^{1/2}$ and $I_3 \leq \left(\frac{2C^*T}{\log\left(v_0/v_{min}^{(p)}\right)} + Tv_{min}^{(p)}\right)^{1/2}$. Hence the claim. □

Remark 31. *As already mentioned, the proof relies on the derivation of the strong convergence of* $\left(S_t^{m,n,J(p)}\right)_{0\leq t\leq T}$ *towards* $\left(S_t^{m,n,ref}\right)_{0\leq t\leq T}$ *which extends the result of Theorem 4.5 to swap rate processes. Observe also that if the conjecture phrased above in Remark 30 is proved to hold true, the result in Proposition 4.8 would be improved accordingly to get that*

$$\varepsilon_1(p) \leq K'/\sqrt{v_{max}^{(p)}/v_0}$$

for some positive constant K'.

4.4.1.4 Weak convergence of the swap rate process

As polynomial processes are always defined by their moments, it is interesting to know how fast moments of swap rate process defined in (4.1) will converge to that of (4.7).

Proposition 4.9. *Let us fix* $k \in \mathbb{N}$. *In the case when* $v_{min}^{(p)} \equiv 0$, *there exists some finite positive constant* C *such that:*

$$\left|\mathbb{E}\left[(S_T^{m,n})^k\right] - \mathbb{E}\left[(S_T^{m,n,J})^k\right]\right| \leq \frac{C}{v_{max}^{(p)}}.$$

In the case when $v_{max}^{(p)} \equiv +\infty$, *there exists some finite positive constant* C' *such that:*

$$\left|\mathbb{E}\left[(S_T^{m,n})^k\right] - \mathbb{E}\left[(S_T^{m,n,J})^k\right]\right| \leq C' v_{min}^{(p)}.$$

Proof. From the matrix representation of polynomial processes, we get that

$$\mathbb{E}\big[(S_T^{m,n})^k\big] = B_k(S_0^{m,n}, V_0) \cdot {}^t\exp\left(T^t A^{(k)}\right)\overrightarrow{p_k} = B_N(S_0^{m,n}, V_0) \cdot \exp\left(TA^{(k)}\right)\overrightarrow{p_k},$$

where $B_k(x,v)$ is a basis of $\mathcal{P}_k(\mathbb{R}^2)$, $\overrightarrow{p_k}$ is the vector representing the monomial x^k in the basis $B_k(x,v)$ and $A^{(k)} \in \mathbb{R}^{N\times N}$ is the matrix representation of the action of the generator associated to the diffusion (4.1) on this basis. The same stands for the moments of $S^{m,n,J}$ with adjusted notations:

$$\mathbb{E}\big[(S_T^{m,n,J})^k\big] = B_k(S_0^{m,n}, V_0) \cdot {}^t\exp\left(T^t A^{(k),J}\right)\overrightarrow{p_k} = B_N(S_0^{m,n}, V_0) \cdot \exp\left(TA^{(k),J}\right)\overrightarrow{p_k},$$

where $A^{(k),J} \in \mathbb{R}^{N\times N}$ is the matrix representation of the action of the generator associated to

the diffusion (4.7). If $|||\cdot|||$ is a matrix norm induced by a vector norm $\|\cdot\|$, we have that

$$\begin{aligned}
\left|\mathbb{E}\big[(S_T^{m,n})^k\big] - \mathbb{E}\big[(S_T^{m,n,J})^k\big]\right| &= \left|B_N(S_0^{m,n}, V_0)\cdot\Big(\exp\big(TA^{(k)}\big) - \exp\big(TA^{(k),J}\big)\Big)\overrightarrow{p_k}\right| \\
&\leq \|B_N(S_0^{m,n}, V_0)\|\left\|\Big(\exp\big(TA^{(k)}\big) - \exp\big(TA^{(k),J}\big)\Big)\overrightarrow{p_k}\right\| \\
&\leq \|B_N(S_0^{m,n}, V_0)\| \times \left|\left|\left|\exp\big(TA^{(k)}\big) - \exp\big(TA^{(k),J}\big)\right|\right|\right| \|\overrightarrow{p_k}\| \\
&\leq T\|B_N(S_0^{m,n}, V_0)\|\|\overrightarrow{p_k}\| \times \left|\left|\left|A^{(k)} - A^{(k),J}\right|\right|\right| e^{|||A^{(k)}|||} e^{|||A^{(k),J}|||}.
\end{aligned}$$

All matrix norms are equivalent in finite dimension and in the following we will focus on the infinite matrix norm: for $M \in \mathbb{R}^{N\times N}$, $|||M||| = \max_{1\leq i\leq N}|M_{i,j}|$. The infinitesimal generator of the diffusion (4.1) applied on a twice continuously differentiable bivariate function writes $\mathcal{A}f(v,s) = \kappa(\theta-\xi v)\partial_v f(v,s) + \frac{\epsilon^2}{2}v\partial_v^2 f(v,s) + \frac{v}{2}\|\lambda^{m,n}\|^2\partial_s^2 f(v,s) + \rho\epsilon\|\lambda^{m,n}\|v\partial_{vs}^2 f$ thus applied on monomials of the form $f(v,s) = v^l s^q$ for $(q,l)\in\mathbb{N}^2$ yields

$$\begin{aligned}
\mathcal{A}_t f(v,s) &= \kappa l(\theta-\xi v)s^q v^{l-1} + l(l-1)\frac{\epsilon^2}{2}vs^q v^{l-2} + \frac{v}{2}\|\lambda^{m,n}\|^2 q(q-1)s^{q-2}v^l + lq\rho\epsilon\|\lambda^{m,n}\|s^{q-1}v^l \\
&= \kappa l\theta s^q v^{l-1} - \kappa l\xi s^q v^l + \epsilon^2\frac{l(l-1)}{2}s^q v^{l-1} + \frac{q(q-1)}{2}\|\lambda^{m,n}\|^2 s^{q-2}v^{l+1} + lq\rho\epsilon\|\lambda^{m,n}\|s^{q-1}v^l.
\end{aligned}$$

Now f evaluated through $\mathcal{A}^J$ where $\mathcal{A}^J$ is defined in (4.9) gives

$$\begin{aligned}
\mathcal{A}_t^J f(v,s) &= \kappa l(\theta-\xi v)s^q v^{l-1} + l(l-1)\frac{\epsilon^2}{2}Q(v)s^q v^{l-2} + \frac{v}{2}\|\lambda^{m,n}\|^2 q(q-1)s^{q-2}v^l \\
&\qquad + lq\rho\epsilon\|\lambda^{m,n}\|Q(v)s^{q-1}v^{l-1} \\
&= \kappa l\theta s^q v^{l-1} - \kappa l\xi s^q v^l \\
&\quad + \epsilon^2\frac{l(l-1)}{2}\frac{s^q}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2}\big((v_{max}+v_{min})v^{l-1} - v^l - v_{min}v_{max}v^{l-2}\big) \\
&\quad + lq\rho\epsilon\|\lambda^{m,n}\|s^{q-1}\frac{(v_{max}+v_{min})v^l - v^{l+1} - v_{min}v_{max}v^{l-1}}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2}.
\end{aligned}$$

Without lack of generality, we can consider that B_k is the standard basis of $\mathcal{P}_k(\mathbb{R}^2)$ composed of monomial functions: $\big((v,s)\mapsto v^l s^q\big)_{l+q\leq k}$. Given an enumeration $\Lambda : E_k \to \mathbb{N}$ of the set of exponents $E_k = \big\{(l,q)\in\mathbb{N}^2 : l+q\leq k\big\}$, the non-zero elements of the matrix $A^{(k),J}$ representing

the action of $\mathcal{A}^J$ on $\mathcal{P}_k(\mathbb{R}^2)$ are given by:

$$\begin{aligned}
A^{(k),J}_{\Lambda(l,q),\Lambda(p,q)} &= -\left(\frac{l(l-1)\epsilon^2}{2(\sqrt{v_{max}}-\sqrt{v_{min}})^2} + l\kappa\xi^0(t)\right),\\
A^{(k),J}_{\Lambda(l-1,q),\Lambda(l,q)} &= \kappa l\theta + \frac{l(l-1)\epsilon^2(v_{min}+v_{max})}{2(\sqrt{v_{max}}-\sqrt{v_{min}})^2},\\
A^{(k),J}_{\Lambda(l-2,q),\Lambda(l,q)} &= -\frac{l(l-1)\epsilon^2 v_{min}v_{max}}{2(\sqrt{v_{max}}-\sqrt{v_{min}})^2},\\
A^{(k),J}_{\Lambda(l+1,q-2),\Lambda(l,q)} &= \frac{q(q-1)\|\boldsymbol{\lambda}^{m,n}(t)\|^2}{2},\\
A^{(k),J}_{\Lambda(l+1,q-1),\Lambda(l,q)} &= -\frac{lq\epsilon\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2},\\
A^{(k),J}_{\Lambda(l-1,q-1),\Lambda(l,q)} &= -\frac{lq\epsilon v_{min}v_{max}\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2},\\
A^{(k),J}_{\Lambda(l,q-1),\Lambda(l,q)} &= \frac{lq\epsilon\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|(v_{min}+v_{max})}{(\sqrt{v_{max}}-\sqrt{v_{min}})^2}.
\end{aligned}$$

The entries of the matrix $A^{(k)}$ representing the action of $\mathcal{A}$ on $\mathcal{P}_k(\mathbb{R}^2)$ are obtained by setting $v_{min}=0$ and $v_{max}=+\infty$ in previous equations. In particular, $A^{(k)}_{\Lambda(l-2,q),\Lambda(l,q)} = A^{(k)}_{\Lambda(l+1,q-1),\Lambda(l,q)} = A^{(k)}_{\Lambda(l-1,q-1),\Lambda(l,q)} = 0$.

Consider now the previously introduced sequences of bounds $\left(v^{(p)}_{min}, v^{(p)}_{max}\right)_{p\in\mathbb{N}}$ respectively converging towards 0 and $+\infty$. First in case when $v^{(p)}_{min}\equiv 0$, we have that

$$\begin{aligned}
\left|\!\left|\!\left|A^{(k),J}-A^{(k)}\right|\!\right|\!\right| &= \max_{i,j}\left|A^{(k),J}_{i,j}-A^{(k)}_{i,j}\right|\\
&= \max_{l+q\le k}\left\{\left|A^{(k),J}_{\Lambda(l,q),\Lambda(l,q)}-A^{(k)}_{\Lambda(l,q),\Lambda(l,q)}\right|, \left|A^{(k),J}_{\Lambda(l+1,q-1),\Lambda(l,q)}-A^{(k)}_{\Lambda(l+1,q-1),\Lambda(l,q)}\right|\right\}\\
&= \max_{l+q\le k}\left\{\frac{l(l-1)\epsilon^2}{2v^{(p)}_{max}}, \frac{lq\epsilon\rho\|\lambda^{m,n}\|}{v^{(p)}_{max}}\right\}\\
&\le \frac{C}{v^{(p)}_{max}},
\end{aligned}$$

for some finite positive constant C that can be written as $C=ck^2$ for some $c>0$. Conversely, in the case when $v^{(p)}_{max}\equiv+\infty$, we get

$$\begin{aligned}
\left|\!\left|\!\left|A^{(k),J}-A^{(k)}\right|\!\right|\!\right| &= \max_{i,j}\left|A^{(k),J}_{i,j}-A^{(k)}_{i,j}\right|\\
&= \max_{l+q\le k}\left\{\left|A^{(k),J}_{\Lambda(l-2,q),\Lambda(l,q)}-A^{(k)}_{\Lambda(l-2,q),\Lambda(l,q)}\right|, \left|A^{(k),J}_{\Lambda(l-1,q-1),\Lambda(l,q)}-A^{(k)}_{\Lambda(l-1,q-1),\Lambda(l,q)}\right|\right\}\\
&= \max_{l+q\le k}\left\{\frac{v^{(p)}_{min}l(l-1)\epsilon^2}{2}, lq\epsilon\rho\|\lambda^{m,n}\|v^{(p)}_{min}\right\}\\
&\le C' v^{(p)}_{min},
\end{aligned}$$

for some constant C' that can be written as $C'=c'k^2$ for some $c'>0$. Hence the result. □

Since dealing with processes characterized by their moments, it is quite natural to anticipate

that Proposition (4.9) can be expressed in terms of distribution convergence. In the following theorem, $\phi_J(z) = \mathbb{E}^S\left[e^{zS_T^{m,n,J}}\right]$ (respectively $\phi(z) = \mathbb{E}^S\left[e^{zS_T^{m,n}}\right]$) denotes the moment generating function associated with dynamics (4.7) (resp. (4.1)). They are defined over an open disk centred at the origin and we denote by $r > 0$ its radius denoted by $\mathcal{D}(0, r)$.

Theorem 4.10. *Let $z \in \mathcal{D}(0, r)$. Then*

$$|\phi_J(z) - \phi(z)| \leq C\frac{(|z|^2 + |z|)e^{|z|}}{v_{max}}$$

for some positive constant C.

Proof. We rely on Proposition 4.9 and its proof. Below, the notation C represent generic positive number that may change from line to line.

$$\begin{aligned}
|\phi_J(z) - \phi(z)| &= \Big|\sum_{k=0}^{+\infty}\frac{z^k}{k!}\big(\mathbb{E}^S[(S_T^{m,n,J})^k] - \mathbb{E}^S[(S_T^{m,n})^k]\big)\Big| \\
&\leq \frac{C}{v_{max}}\sum_{k=0}^{+\infty}\frac{k^2|z|^k}{k!} \\
&\leq \frac{C}{v_{max}}\sum_{k=0}^{+\infty}\frac{k(k-1)+k}{k!}|z|^k \\
&\leq C\frac{(|z|^2 + |z|)e^{|z|}}{v_{max}}.
\end{aligned}$$

□

We again restrict ourselves to the case of time homogeneous diffusions for (4.1) and (4.7). We state a similar result to that of Proposition 4.7 for the swap rate process giving thus an estimate on the error made when replacing Cox-Ingersoll-Ross process by Jacobi one.

Theorem 4.11. *Let us fix $v_{min} > 0$ and a real number $\beta > 1$. Let $f \in \mathcal{C}^2(\mathbb{R})$ be a bounded payoff function such that the functions $\mathbb{R} \ni x \mapsto (1 + x)f'(x)$ and $\mathbb{R} \ni x \mapsto (1 + x^2)f''(x)$ are bounded. Assume that the parameters of the model satisfy the relationship*

$$4\kappa\theta\xi^0 > 5\epsilon^2. \tag{4.27}$$

Then, there exists positive constants $C_{v_{min}}$, c_1 and c_2 such that

$$\left|\mathbb{E}^S\left[f(S_T^{m,n})\right] - \mathbb{E}^S\left[f(S_T^{m,n,J})\right]\right| \leq C_{v_{min}}\left(\frac{c_1}{v_{max}} + c_2\right)$$

for all $v_{max} > \beta v_{min}$.

Remark 32. *This result is rather theoretical than practical. Indeed in present result, the constant $C_{v_{min}}$ appearing in the statement of Theorem 4.11 explodes as $v_{min} \to 0$ due to the fact that, for any $\alpha > 0$, $\int_0^T \mathbb{E}^S\left[\frac{1}{(V_t^0)^\alpha}\right]\mathrm{d}t = +\infty$ where V^0 is the CIR process starting at zero. The estimate on the pricing error provided in the statement simply allows to get some intuition on the convergence as $v_{max} \to +\infty$ for a fixed $v_{min} > 0$. We conjecture that a similar result holds in the particular case when $v_{min} = 0$.*

Remark 33. *Theorem 4.11 provides a pricing error when comparing dynamics (4.1) and (4.7) for smooth payoff functions. We guess that a comparable result may be obtained for less regular payoff comprising standard ones that mainly writes as linear combination of indicator functions.*

Proof. We are interested in this proof in the characterization of distributions of $S^{m,n}$ and $S^{m,n,J}$. For simplicity we consider one dimensional SDE for modelling their respective dynamics. Namely, we focus on

$$\mathrm{d}S_t^{m,n} = \sqrt{V_t}\Big(\rho\|\boldsymbol{\lambda}^{m,n}\|\mathrm{d}W_t + \sqrt{1-\rho^2}\|\boldsymbol{\lambda}^{m,n}\|\mathrm{d}W_t^{S,*}\Big), \tag{4.28}$$

where $(W_t^{S,*})_{0\le t\le T}$ is a *one dimensional* Brownian motion –that replaces the multidimensional one in (4.1)– and where $(V_t)_{0\le t\le T}$ is a CIR process. Choosing V as a Jacobi process in right-hand side of (4.28) yields the one dimensional counterpart of (4.7) and allows to define $(S_t^{m,n,J})_{0\le t\le T}$. For simplicity, we will denote in this proof $X = S^{m,n}$ and $X^J = S^{m,n,J}$.

Throughout the proof of the Theorem 4.11 we will use several times the three following properties of the Cox-Ingersoll-Ross process presented in the Lemmas 4.12, 4.13 and 4.14 below. The first key property is about the regularity through time of the (negative) moments of the Cox-Ingersoll-Ross starting at zero.

Lemma 4.12. *Let $T > 0$ and let us consider the Cox-Ingersoll-Ross process starting at $x > 0$:*

$$V_t^x = x + \int_0^t \kappa(\theta - \xi^0 V_u^x)\mathrm{d}u + \epsilon\int_0^t \sqrt{V_u^x}\mathrm{d}B_u,\ 0\le t\le T, \tag{4.29}$$

where $(B_t)_{0\le t\le T}$ is a Brownian motion. Then there exists some constant $C' > 0$ such that

$$\forall t\in[0,T],\ \mathbb{E}\left[\frac{1}{(V_t^x)^\alpha}\right] \le \frac{C'}{x^\alpha}$$

for any $0 < \alpha < \frac{2\kappa\xi_0\theta}{\epsilon^2}$.

Proof of lemma 4.12. First observe that

$$V_t^x \overset{a.s.}{=} x + \int_0^t \tilde{\kappa}\Big(\frac{\theta}{\xi^0} - V_u^x\Big)\mathrm{d}u + \int_0^t \epsilon\sqrt{V_u^x}\mathrm{d}B_u$$

where $\tilde{\kappa} = \kappa\xi^0$. Now, since

$$\frac{1}{(V_t^x)^\alpha} \overset{a.s.}{=} \frac{1}{\Gamma(\alpha)}\int_0^{+\infty} y^{\alpha-1}e^{-yV_t^x}\mathrm{d}y$$

which follows from the definition of the Γ function. Using Fubini theorem and following Proposition 1.2.4 in [Alf15], we finally get that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(V_t^x)^\alpha}\right] &= \frac{1}{\Gamma(\alpha)}\int_0^{+\infty}\frac{y^{\alpha-1}}{\big(1+\frac{\epsilon^2}{2}y\xi_{\tilde{\kappa}}(t)\big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{ye^{-\tilde{\kappa}t}}{1+\frac{\epsilon^2}{2}y\xi_{\tilde{\kappa}}(t)}x\Big)\mathrm{d}y \\ &=: \frac{1}{\Gamma(\alpha)}\int_0^{+\infty} g(t,y)\mathrm{d}y\end{aligned} \tag{4.30}$$

where we set $\xi_{\tilde{\kappa}}(t) = \frac{1-e^{-\tilde{\kappa}t}}{\tilde{\kappa}}$. In identity (4.30), let us apply the change of variable $z = y\xi_{\tilde{\kappa}}(t)$

and set $u(t) = e^{-\tilde{\kappa}t}/\xi_{\tilde{\kappa}}(t)$ so that:

$$\begin{aligned}\mathbb{E}\Big[\frac{1}{(V_t^x)^\alpha}\Big] &= \frac{1}{\xi_{\tilde{\kappa}}(t)^\alpha\Gamma(\alpha)}\int_0^{+\infty}\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{xzu(t)}{1+\frac{\epsilon^2}{2}z}\Big)\mathrm{d}z \\ &=: \frac{1}{\xi_{\tilde{\kappa}}(t)^\alpha\Gamma(\alpha)}F\big(u(t)\big).\end{aligned} \tag{4.31}$$

Let us split the integral defining F into two integrals over $(0,1)$ and $(1,\infty)$ and study both of them. First, observe that

$$\begin{aligned}\int_0^1\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{xzu(t)}{1+\frac{\epsilon^2}{2}z}\Big)\mathrm{d}z &\le \int_0^1\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{xzu(t)}{1+\frac{\epsilon^2}{2}}\Big)\mathrm{d}z \\ &\le \int_0^1 z^{\alpha-1}\exp\Big(-\frac{xzu(t)}{1+\frac{\epsilon^2}{2}}\Big)\mathrm{d}z \\ &\le \Big(\frac{1+\epsilon^2/2}{xu(t)}\Big)^\alpha\int_0^{\frac{xu(t)}{1+\epsilon^2/2}}v^{\alpha-1}e^{-v}\mathrm{d}v \\ &\le \Big(\frac{1+\epsilon^2/2}{xu(t)}\Big)^\alpha\int_0^{+\infty}v^{\alpha-1}e^{-v}\mathrm{d}v \\ &= \Big(\frac{1+\epsilon^2/2}{xu(t)}\Big)^\alpha\Gamma(\alpha)\end{aligned} \tag{4.32}$$

where we set $v = \frac{xu(t)}{1+\epsilon^2/2}z$ in third inequality and used that $0 \le z \le 1 \Leftrightarrow 1 \le 1+\frac{\epsilon^2}{2}z \le 1+\frac{\epsilon^2}{2} \Leftrightarrow -xzu(t) \le -\frac{xzu(t)}{1+\frac{\epsilon^2}{2}z} \le -\frac{xzu(t)}{1+\frac{\epsilon^2}{2}}$ since $xu(t) \ge 0$. Second, the integral over $(1,\infty)$ is handled as follows:

$$\begin{aligned}\int_1^\infty\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{xzu(t)}{1+\frac{\epsilon^2}{2}z}\Big)\mathrm{d}z &= \int_1^\infty\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\exp\Big(-\frac{xu(t)}{1/z+\frac{\epsilon^2}{2}}\Big)\mathrm{d}z \\ &\le \exp\Big(-\frac{xu(t)}{1+\frac{\epsilon^2}{2}}\Big)\int_1^\infty\frac{z^{\alpha-1}}{\Big(1+\frac{\epsilon^2}{2}z\Big)^{2\tilde{\kappa}\theta/\epsilon^2}}\mathrm{d}z \\ &= C\exp\Big(-\frac{xu(t)}{1+\frac{\epsilon^2}{2}}\Big)\end{aligned} \tag{4.33}$$

where $C < \infty$ for $\alpha < 2\tilde{\kappa}\theta/\epsilon^2$ and we used that $1 \le z \Leftrightarrow \frac{1}{z}+\frac{\epsilon^2}{2} \le 1+\frac{\epsilon^2}{2} \Leftrightarrow -\frac{xu(t)}{\frac{1}{z}+\frac{\epsilon^2}{2}} \le -\frac{xu(t)}{1+\frac{\epsilon^2}{2}}$. We employ then (4.32) and (4.33) to get the following estimates in which the constant $C' > 0$ is only depending on α and the parameters of (4.29) but not on t; it may change from line to

line below. It is such that for all $t \in [0,T]$:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(V_t^x)^\alpha}\right] &\leq \frac{C'}{\xi_{\tilde\kappa}(t)^\alpha}\left(\frac{1}{\big(xu(t)\big)^\alpha} + \exp\left(-\frac{xu(t)}{1+\frac{\epsilon^2}{2}}\right)\right)\\ &\leq C'\left(\frac{1}{x^\alpha e^{-\alpha\tilde\kappa t}} + \frac{1}{\xi_{\tilde\kappa}(t)^\alpha}\exp\left(-\frac{xe^{-\tilde\kappa t}/\xi_{\tilde\kappa}(t)}{1+\frac{\epsilon^2}{2}}\right)\right)\\ &\leq C'\left(\frac{1}{x^\alpha} + \frac{1}{\xi_{\tilde\kappa}(t)^\alpha}\exp\left(-\frac{xe^{-\tilde\kappa t}/\xi_{\tilde\kappa}(t)}{1+\frac{\epsilon^2}{2}}\right)\right).\end{aligned} \tag{4.34}$$

Since $\xi_{\tilde\kappa}(t) \xrightarrow[t\to 0]{} 0$, (4.34) is studied below independently for large times t and small ones.

Set $T \geq \tau > 0$. For all $t \in [\tau, T]$, set $y = x/\xi_{\tilde\kappa}(t)$ and we have

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(V_t^x)^\alpha}\right] &\leq C'\left(\frac{1}{x^\alpha} + \frac{y^\alpha}{x^\alpha}\exp\left(-\frac{ye^{-\tilde\kappa T}}{1+\frac{\epsilon^2}{2}}\right)\right)\\ &\leq \frac{C'}{x^\alpha}\left(1 + \sup_{y\geq 0} y^\alpha \exp\left(-\frac{ye^{-\tilde\kappa T}}{1+\frac{\epsilon^2}{2}}\right)\right)\\ &\leq \frac{C'}{x^\alpha}.\end{aligned}$$

Previous estimate holds for all $\tau > 0$. Using for small t, that $\xi_{\tilde\kappa}(t) \sim t$ and that $\sup_{t\geq 0}\frac{1}{t^\alpha}\exp\left(-\frac{x/t}{1+\frac{\epsilon^2}{2}}\right) = \sup_{s\geq 0}\frac{s^\alpha}{x^\alpha}\exp\left(-\frac{s}{1+\frac{\epsilon^2}{2}}\right) = \frac{\tilde C}{x^\alpha}$ in (4.34), we finally get the following bounding for t small enough:

$$\mathbb{E}\left[\frac{1}{(V_t^0)^\alpha}\right] \leq \frac{C'}{x^\alpha}. \tag{4.35}$$

The claim is proved.

□

Note that under assumption of Theorem 4.11, the exponent $\alpha \in \mathbb{R}$ of Lemma 4.12 could be chosen in $(-\infty, 5/2]$.

The second important property is about the almost sure monotonicity of the Cox-Ingersoll-Ross process with respect to its initial value.

Lemma 4.13. *Denoting by $(V_t^v)_{0\leq t\leq T}$ the Cox-Ingersoll-Ross process starting at $v \in \mathbb{R}_+^*$, then for any time $t \in [0,T]$ the map $\mathbb{R}_+ \ni v \mapsto V_t^v$ is increasing almost surely. In particular, for any $0 < v \leq w$,*

$$\mathbb{P}\left(\forall t > 0 : \frac{1}{V_t^w} \leq \frac{1}{V_t^v} < \infty\right) = 1.$$

Proof of lemma 4.13. The first part of the claim is simply a consequence of the comparison theorem for solutions of stochastic differential equations: see for instance Proposition 2.18, Section 5.2.C in [KS91]. Consequently, $\mathbb{R}_+^* \ni v \mapsto 1/V_t^v$ is decreasing almost surely for any $t \in [0,T]$ and thus

$$\forall t \in [0,T], \text{ for any } w \geq v > 0, \ \frac{1}{V_t^w} \leq \frac{1}{V_t^v}.$$

Theorem 2.4 in [Duf01] provides that $\mathbb{E}\big[\big(V_t^v\big)^\gamma\big] < \infty$ whenever $\gamma > -2\kappa\theta\xi^0/\epsilon^2$. Consequently, as soon as $2\kappa\theta\xi^0 > \epsilon^2$, $\mathbb{E}^S[(V_t^w)^{-1}] < \infty$ for all $t \in [0,T]$. As being a positive random variable of finite expectation, $(V_t^w)^{-1}$ is necessarily finite almost surely for any $t \in [0,T]$. Hence the result. □

Third property we need is about the regularity of the *square-root* of the Cox-Ingersoll-Ross process with respect to its initial value. We denote by $\mathcal{V}^w = \sqrt{V^v}$ and $w = \sqrt{v}$; observe that by Itô's lemma we get

$$\mathcal{V}_t^w \overset{a.s.}{=} w + \int_0^t \Big((\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8})(\mathcal{V}_s^w)^{-1} - \frac{\kappa\xi^0}{2}\mathcal{V}_s^w \Big)\mathrm{d}s + \frac{\epsilon}{2}W_t,\ t \le T. \tag{4.36}$$

We denote by $(\partial\mathcal{V}_t^w)_{0\le t\le T}$ the first variation process of $(\mathcal{V}_t^w)_{0\le t\le T}$; it can be defined as the almost sure solution of the following ordinary differential equation

$$\partial\mathcal{V}_t^w = 1 - \int_0^t \Big(\Big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\Big)\big(\mathcal{V}_u^w\big)^{-2} + \frac{\kappa\xi^0}{2}\Big)\partial\mathcal{V}_u^w \mathrm{d}u,\ t \in [0,T]. \tag{4.37}$$

Remark 34. *The definition of $\partial\mathcal{V}^w$ can be formally obtained by simply deriving (4.36) with respect to w. In the literature, this is done and justified for SDEs with smooth coefficients; we refer to [Stu04] for more details. We extend those results in Lemma 4.14 to the particular case of the non bounded drift function of (4.36).*

Lemma 4.14. *Assume that $\kappa\theta > \epsilon^2$. Let us denote by*

$$\mathcal{W}_t^w := \mathcal{V}_t^w - \int_0^w \partial\mathcal{V}_t^y \mathrm{d}y$$

for all $t \in [0,T]$. Then, the first variation process defined in (4.37) is such that, at any time $t \in [0,T]$, $\mathcal{W}_t^w$ does not depend on w:

$$\mathcal{W}_t^w = \mathcal{W}_t^{w'}.$$

for any $w > 0, w' > 0$.

Proof of lemma 4.14. First, observe that the differential equation (4.37) satisfied by $(W_t^w)_{0\le t\le T}$ can be explicitly solved to obtain

$$\partial\mathcal{V}_t^w \overset{a.s.}{=} \exp\Big(- \int_0^t \Big(\Big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\Big)\big(\mathcal{V}_u^w\big)^{-2} - \frac{\kappa\xi^0}{2}\Big)\mathrm{d}u\Big),\ 0 \le t \le T, \tag{4.38}$$

proving the existence of $(\partial\mathcal{V}_t^w)_{0\le t\le T}$. Let us fix $w > 0$ and set $w' \ge w$.

$$\begin{aligned}
\mathcal{W}_t^w - \mathcal{W}_t^{w'} &= w - w' + \int_0^t \Big(\Big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\Big)\Big(\frac{1}{\mathcal{V}_u^w} - \frac{1}{\mathcal{V}_u^{w'}}\Big) - \frac{\kappa\xi^0}{2}\Big(\mathcal{V}_u^w - \mathcal{V}_u^{w'}\Big)\Big)\mathrm{d}u \\
&\quad + \int_w^{w'} \Big(1 - \int_0^t \Big((\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8})\frac{1}{(\mathcal{V}_u^y)^2} + \frac{\kappa\xi^0}{2}\Big)\partial\mathcal{V}_u^y \mathrm{d}u\Big)\mathrm{d}y \\
&= \int_0^t \Big((\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8})\Big(\frac{1}{\mathcal{V}_u^w} - \frac{1}{\mathcal{V}_u^{w'}} - \int_w^{w'} \frac{\partial\mathcal{V}_u^y}{(\mathcal{V}_u^y)^2}\mathrm{d}y\Big) - \frac{\kappa\xi^0}{2}(\mathcal{V}_u^w - \mathcal{V}_u^{w'} + \int_w^{w'} \partial\mathcal{V}_u^y \mathrm{d}y)\Big)\mathrm{d}u \\
&= \int_0^t \Big((\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8})\Big(\frac{\mathcal{V}_u^{w'} - \mathcal{V}_u^w}{\mathcal{V}_u^w\mathcal{V}_u^{w'}} - \int_w^{w'} \frac{\partial\mathcal{V}_u^y}{(\mathcal{V}_u^y)^2}\mathrm{d}y\Big) - \frac{\kappa\xi^0}{2}(\mathcal{W}_u^w - \mathcal{W}_u^{w'})\Big)\mathrm{d}u \\
&\le \int_0^t \frac{1}{(\mathcal{V}_u^w)^2}\Big(- (\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}) - \frac{\kappa\xi^0}{2}\Big)(\mathcal{W}_u^w - \mathcal{W}_u^{w'})\mathrm{d}u
\end{aligned}$$

where we used the monotonicity result of Proposition 4.13 to obtain the last inequality. Jensen's

inequality then implies for all $t \geq 0$,

$$|\mathcal{W}_t^w - \mathcal{W}_t^{w'}| \leq \Big(\big(\frac{\kappa\theta}{2} + \frac{\epsilon^2}{8}\big) + \frac{\kappa\xi^0}{2} \Big) \int_0^t \frac{1}{V_u^w} |\mathcal{W}_u^w - \mathcal{W}_u^{w'}| \mathrm{d}u.$$

Gronwall's lemma applied almost surely provides then

$$\mathbb{P}^S\big(\mathcal{W}_t^w = \mathcal{W}_t^{w'}\big) = 1.$$

The result follows as w and w' were set arbitrarily.

□

We are now ready to prove the statement of Theorem 4.11. Let us consider $f \in \mathcal{C}^2(\mathbb{R})$ as in the statement of the theorem. Let us define $u(t,x,v) = \mathbb{E}\big[f(X_{T-t})|(X_0,V_0) = (x,v)\big]$ and $\tilde{u}(t,x,v) = u(T-t,x,v)$. Proposition 3.2 in [ET10] (see Appendix D for the sake of completeness) justifies that $u \in \mathcal{C}^{1,2,2}\big([0,T) \times \mathbb{R} \times \mathbb{R}_+^*\big)$. Note in addition that $u(T,x,v) = f(x)$ for all $(x,v) \in \mathbb{R} \times \mathbb{R}_+^*$. The Feynman-Kac formula implies (see for instance [KS91]):

$$\partial_t u + \kappa(\theta - \xi^0 v)\partial_v u + \frac{\epsilon^2}{2} v \partial_v^2 u + \rho\epsilon v \partial_{xv}^2 u + \frac{v}{2}\partial_x^2 u = 0, \ (x,v) \in \mathbb{R} \times \mathbb{R}_+^*, t \in [0,T]. \tag{4.39}$$

Observe now that

$$\begin{aligned} \mathbb{E}^S[f(X_T^J)] - \mathbb{E}^S[f(X_T)] &= \mathbb{E}^S\left[u\big(T, X_T^J, V_T^J\big)\right] - \mathbb{E}^S\left[u(T, X_T, V_T)\right] \\ &= \mathbb{E}^S\left[u\big(T, X_T^J, V_T^J\big) - u(0,x,v)\right] \\ &= \mathbb{E}^S\left[\int_0^T g(s, X_s^J, V_s^J)\mathrm{d}s\right] \end{aligned} \tag{4.40}$$

where we set $g(t,x,v) = \partial_t u + \kappa(\theta - \xi^0 v)\partial_v u + \frac{\epsilon^2}{2} Q(v)\partial_v^2 u + \rho\epsilon Q(v)\partial_{xv}^2 u + \frac{v}{2}\partial_x^2 u$. Equation (4.39) yields that

$$g(t,x,v) = \frac{\epsilon^2}{2}(Q(v) - v)\partial_v^2 u(t,x,v) + \rho\epsilon(Q(v) - v)\partial_{xv}^2 u(t,x,v).$$

Let $v_{min} > 0$ be given. By definition, we have $v - Q(v) = v^2/\gamma + \frac{v_{min}v_{max}}{\gamma} - \big(\frac{v_{max}+v_{min}}{\gamma} - 1\big)v$ where $\gamma = (\sqrt{v_{max}} - \sqrt{v_{min}})^2$. Observe that it is enough to study $v - Q(v)$ for $v \in [v_{min}, v_{max}]$ following (4.40). First note that over $[v_{min}, v_{max}]$, $v - Q(v) \geq 0$. Secondly, since $\gamma \leq v_{max} + v_{min}$, we have $v - Q(v) \leq v^2/\gamma + \frac{v_{min}v_{max}}{\gamma}$. Third, pick $\beta > 1$ so that for $v_{min} \leq \frac{v_{max}}{\beta}$ we have on the one hand $\frac{v_{max}v_{min}}{\gamma} \leq C v_{min}$ and on the other hand $\frac{v^2}{\gamma} \leq C' \frac{v^2}{v_{max}}$ for some positive constants C, C'. Finally, there is some $\tilde{c} > 0$ so that for all $v \in [v_{min}, v_{max}]$, we get

$$|v - Q(v)| = v - Q(v) \leq \tilde{c}\Big(\frac{v^2}{v_{max}} + v_{min}\Big). \tag{4.41}$$

Using thus (4.41) and (4.40), it remains to show that $\max\Big(\mathbb{E}^S\Big[\int_0^T \big((V_s^J)^2 + 1\big)\big|\partial_{xv}^2 u(s, X_s^J, V_s^J)\big|\mathrm{d}s\Big], \mathbb{E}^S\Big[\int_0^T \big((V_s^J)^2 + 1\big)\big|\partial_v^2 u(s, X_s^J, V_s^J)\big|\mathrm{d}s\Big]\Big) < \infty$ to obtain the wanted result. To do so, we will bound $\mathbb{E}^S\left[\big((V_t^J)^2 + 1\big)\big|\partial_{xv}^2 u(t, X_t^J, V_t^J)\big|\right]$ and $\mathbb{E}^S\left[\big((V_t^J)^2 + 1\big)\big|\partial_v^2 u(t, X_t^J, V_t^J)\big|\right]$ independently of t. In the sequel, we will rather work with the square-root of the volatility process defined in (4.36) by observing that $\partial_v \tilde{u}(t,x,v) = \frac{1}{2w}\partial_w \tilde{u}(t,x,w)$ where $w = \sqrt{v}$.

Study of $\partial_w^2 u$. We first study the second derivative of u with respect to its third argument:

$$\partial_v^2 u(t,x,v) = \partial_v^2 \left(\mathbb{E}^S[f(X_{T-t})|(X_0,V_0)=(x,v)]\right) = \partial_v^2 \tilde{u}(T-t,x,v).$$

Based on the above relationship, we will alternatively study the quantity $\partial_w \tilde{u}$. As mentioned, we will rather consider the square-root of the CIR process in this proof; moreover, since we are studying the dependency with respect to initial condition, we emphasize it by denoting $(X^{x,w}, V^w)$ the process defined in (4.28) associated with CIR process such that $X_0^{x,w} = S_0^{m,n} = x$ and $V_0^w = V_0 = w^2$, similarly to the definition of the square-root process (4.36). With the definition (4.14), we can define the first variation process (with respect to w) of $X^{x,w}$ as

$$\partial_w X_t^{x,w} = \rho \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\| \int_0^t \partial \mathcal{V}_u^w \mathrm{d}W_u + \sqrt{1-\rho^2}\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\| \int_0^t \partial \mathcal{V}_u^w \mathrm{d}W_u^{S,*},\ 0 \le t \le T. \tag{4.42}$$

For any time $t \in [0,T]$, one can show that the quantity $X_t^{x,w} - \int_0^w \partial_w X_t^{x,y} \mathrm{d}y$ does not depend on w using Fubini's theorem for interchanging order between Lebesgue and Itô's integrals and Lemma 4.14.

From the hypothesis made on the payoff function f, we have at any time $t \in [0,T]$,

$$\begin{aligned} \left|\partial_w f(X_t^{x,w})\right| &= \left|f'(X_t^{x,w})\partial_w X_t^{x,w}\right| \\ &\le \|f'\|_\infty \left|\partial_w X_t^{x,w}\right|. \end{aligned} \tag{4.43}$$

We aim at bounding last term in (4.43) independently of (x,w) by an integrable random variable. By doing so, we could apply Lebesgue's theorem to get

$$\partial_w \tilde{u}(t,x,w) = \mathbb{E}_{x,w}^S \left[f'(X_t)\partial_w X_t^{x,w}\right]. \tag{4.44}$$

An integration by part in (4.42) yields

$$\begin{aligned} \partial_w X_t^{x,w} &\overset{a.s.}{=} \|\boldsymbol{\lambda}^{m,n}\| \int_0^t \partial \mathcal{V}_u^w \left(\rho \mathrm{d}W_u + \sqrt{1-\rho^2}\mathrm{d}W_u^{S,*}\right) \\ &= \|\boldsymbol{\lambda}^{m,n}\| \Big\{ \partial \mathcal{V}_t^w \left(\rho W_t + \sqrt{1-\rho^2} W_t^{S,*}\right) \\ &\quad + \int_0^t \left(\rho W_t + \sqrt{1-\rho^2} W_t^{S,*}\right) \Big[\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right) \frac{\partial \mathcal{V}_u^w}{(\mathcal{V}_u^w)^2} - \frac{\kappa \xi^0}{2} \partial \mathcal{V}_u^w \Big] \mathrm{d}u \Big\}. \end{aligned} \tag{4.45}$$

Using now Jensen's inequalities, fact that for any time $t \le T$, $|\partial \mathcal{V}_t^w| \le 1$ almost surely, Lemma 4.13 and that $|\rho| \le 1$, we get almost surely:

$$\begin{aligned} \left|\partial_w X_t^{x,w}\right| &\le 3\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\| \Big\{ \left|W_t + W_t^{S,*}\right| \\ &\quad + \Big| \int_0^t \left(\rho W_u + \sqrt{1-\rho^2} W_u^{S,*}\right) \Big[\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right) \frac{\partial \mathcal{V}_u^w}{(\mathcal{V}_u^w)^2} - \frac{\kappa \xi^0}{2} \partial \mathcal{V}_u^w \Big] \mathrm{d}u \Big| \Big\} \\ &\le 3\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\| \Big\{ \left|W_t\right| + \left|W_t^{S,*}\right| \\ &\quad + \int_0^t \left(|W_u| + |W_u^{S,*}|\right) \Big[\left(\frac{\kappa\theta}{2} + \frac{\epsilon^2}{8}\right) \left(\frac{1}{V_u^{v_{min}}}\right) + \frac{\kappa\xi^0}{2} \Big] \mathrm{d}u \Big\} \\ &=: Z_{1,t} \end{aligned} \tag{4.46}$$

where $Z_{1,t}$ does not depend on (w,x). We now justify its integrability. First note that $\mathbb{E}^S\left[|W_t|\right] = \mathbb{E}^S\left[|W_t^{S,*}|\right] = \sqrt{\frac{2t}{\pi}}$. Fubini theorem, Cauchy-Schwarz inequality, fact that $\mathbb{E}^S\left[(W_u)^2\right] =$

$\mathbb{E}^S\big[\big(W_u^{S,*}\big)^2\big] = u$ for all time $u \in [0,T]$ and Lemma 4.12 imply that

$$\begin{aligned}
&\mathbb{E}^S\Big[\int_0^t \big(|W_u| + |W_u^{S,*}|\big)\Big[\Big(\frac{\kappa\theta}{2} + \frac{\epsilon^2}{8}\Big)\Big(\frac{1}{V_u^{v_{min}}}\Big) + \frac{\kappa\xi^0}{2}\Big]\mathrm{d}u\Big] \\
&\leq \int_0^t \mathbb{E}^S\big[\big(|W_u| + |W_u^{S,*}|\big)^2\big]^{1/2}\mathbb{E}^S\Big[\Big(\frac{\kappa\theta}{2} + \frac{\epsilon^2}{8}\Big)^2\Big(\frac{1}{V_u^{v_{min}}}\Big)^2 + \frac{\kappa\xi^0}{4}\Big]^{1/2}\mathrm{d}u \\
&\leq Ct^{3/2}\Big(v_{min}^{-2} + 1\Big) < \infty
\end{aligned} \tag{4.47}$$

for some $C > 0$ depending on the volatility parameters κ, θ, ϵ. We finally get that $\mathbb{E}^S[Z_{1,t}] \leq C'\big(t^{1/2} + t^{3/2}\big) \leq C'\big(T^{1/2} + T^{3/2}\big)$. This justifies (4.44). Observe also that for all (t,x,w):

$$|\partial_w \tilde{u}(t,x,w)| \leq \|f'\|_\infty C' T^{1/2}\big(1+T\big) \tag{4.48}$$

where C' depends on v_{min} but not on v_{max}.

Let us move onto the second partial derivative with respect to initial variance value. From the hypothesis made on the payoff function again, we have

$$\begin{aligned}
\big|\partial_w^2 f(X_t^{x,w})\big| &= \big|f''(X_t^{x,w})\big(\partial_w X_t^{x,w}\big)^2 + f'(X_t^{x,w})\partial_w^2 X_t^{x,w}\big| \\
&\leq \Big(\|f''\|_\infty\big(\partial_w X_t^{x,w}\big)^2 + \|f'\|_\infty\big|\partial_w^2 X_t^{x,w}\big|\Big)
\end{aligned} \tag{4.49}$$

We are thus led to study $\big(\partial_w X_t^{x,w}\big)^2$ and $\big|\partial_w^2 X_t^{x,w}\big|$: as previously, we want to bound both of them by integrable random variables that do not depend on (w,x). Starting with (4.45), we follow a similar approach as in (4.46) to get:

$$\begin{aligned}
\big(\partial_w X_t^{x,w}\big)^2 &= \|\boldsymbol{\lambda}^{m,n}\|^2\Big\{\partial\mathcal{V}_t^w\big(\rho W_t + \sqrt{1-\rho^2}W_t^{S,*}\big) \\
&\qquad + \int_0^t \big(\rho W_t + \sqrt{1-\rho^2}W_t^{S,*}\big)\Big[\big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\big)\frac{\partial\mathcal{V}_u^w}{(\mathcal{V}_u^w)^2} - \frac{\kappa\xi^0}{2}\partial\mathcal{V}_u^w\Big]\mathrm{d}u\Big\}^2 \\
&\leq 3\|\boldsymbol{\lambda}^{\boldsymbol{m,n}}\|^2\Big\{W_t^2 + \big(W_t^{S,*}\big)^2 \\
&\qquad + 4t\int_0^t \big(|W_u|^2 + |W_u^{S,*}|^2\big)\Big[\Big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\Big)^2\Big(\frac{1}{V_u^{v_{min}}}\Big)^2 + \frac{\kappa^2\big(\xi^0\big)^2}{4}\Big]\mathrm{d}u\Big\} \\
&=: Z_{2,t}
\end{aligned} \tag{4.50}$$

where $Z_{2,t}$ does not depend on (x,w). Let us discuss its integrability. First, $\mathbb{E}^S\big[(W_t)^2\big] = \mathbb{E}^S\big[\big(W_t^{S,*}\big)^2\big] = t$. Furthermore, Hölder's inequality implies that for any $p > 1$ and any time $t > 0$, $\mathbb{E}^S\big[|W_t|^2(V_t^{v_{min}})^{-2}\big] \leq \mathbb{E}^S\big[|W_t|^{2p}\big]^{\frac{1}{p}}\mathbb{E}^S\big[(V_t^{v_{min}})^{-2q}\big]^{\frac{1}{q}}$ were $q = \frac{p}{p-1}$. For all $t \geq 0$, $\mathbb{E}^S\big[|W_t|^{2p}\big] < \infty$ and $[0,T] \ni t \mapsto \mathbb{E}^S\big[|W_t|^{2p}\big]$ is continuous thus it can be bounded independently of time t (moments of the absolute value of Brownian motions can be explicitly computed). Let us now fix $\alpha > 0$ so that $1 + \alpha < \kappa\theta\xi^0/\epsilon^2$ (exists following assumption (4.27)) and chose $p > 1$ large enough so that $q \leq 1 + \alpha$; Hölder's inequality again implies $\mathbb{E}^S\left[(V_t^{v_{min}})^{-2q}\right] \leq \mathbb{E}^S\left[(V_t^{v_{min}})^{-2(1+\alpha)}\right]^{\frac{q}{1+\alpha}} \leq \frac{C}{v_{min}^{2q}}$ following Lemma 4.12. Consequently, $\mathbb{E}^S\big[|W_t|^2(V_t^{v_{min}})^{-2}\big]$ can be bounded independently of time t. A similar discussion applies to $\mathbb{E}^S\big[|W_t^{S,*}|^2(V_t^{v_{min}})^{-2}\big]$. The finiteness of the expectation of $Z_{2,t}$ is deduced since

$$\mathbb{E}^S[Z_{2,t}] \leq C'(t + t^2) \leq C'(T + T^2) \tag{4.51}$$

where C' depends on v_{min} but not on v_{max}.

Let us now study $\partial_w^2 X_\cdot^{x,w}$. Using (4.38), we introduce the second derivative process defined at any time $t \geq 0$ by $\partial^2 \mathcal{V}_t^w = 2\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\frac{(\partial \mathcal{V}_t^w)^2}{(\mathcal{V}_t^w)^3}$. Following Lemma 4.13, we obtain the almost sure estimate

$$\left|\partial^2 \mathcal{V}_t^w\right| \leq 2\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\frac{1}{(\mathcal{V}_t^w)^3} \leq \left(\kappa\theta - \frac{\epsilon^2}{4}\right)\frac{1}{\left(\mathcal{V}_t^{\sqrt{v_{min}}}\right)^3}. \tag{4.52}$$

The right-hand side can be proved to be finite almost surely as a a positive random variable of finite expectation once $\frac{4}{3}\kappa\theta > \epsilon^2$ (Theorem 2.4 in [Duf01]). This is ensured under (4.27) that is a more restrictive condition. We can define then

$$\partial_w^2 X_t^{x,w} \overset{a.s.}{=} \rho\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\int_0^t \partial^2 \mathcal{V}_u^w \mathrm{d}W_u + \sqrt{1-\rho^2}\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\int_0^t \partial^2 \mathcal{V}_u^w \mathrm{d}W_u^{S,*},\ 0 \leq t \leq T,$$

and perform an integration by part to get for any $t \in [0,T]$:

$$\begin{aligned}
\partial_w^2 X_t^{x,w} &\overset{a.s.}{=} \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\int_0^t \partial^2 \mathcal{V}_u^w\left(\rho \mathrm{d}W_u + \sqrt{1-\rho^2}\mathrm{d}W_u^{S,*}\right) \\
&= \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\Bigg\{\partial^2 \mathcal{V}_t^w\left(\rho W_t + \sqrt{1-\rho^2}W_t^{S,*}\right) \\
&\qquad - \int_0^t \left(\rho W_u + \sqrt{1-\rho^2}W_u^{S,*}\right)\left(\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\left(2\frac{(\partial \mathcal{V}_u^w)^2}{(\mathcal{V}_u^w)^3} - \frac{\partial^2 \mathcal{V}_u^w}{(\mathcal{V}_u^w)^2}\right) + \frac{\kappa\xi^0}{2}\partial^2 \mathcal{V}_u^w\right)\mathrm{d}u\Bigg\}.
\end{aligned}$$

Previous estimate (4.52), Jensen's inequality and fact that $|\rho| \leq 1$, provide almost surely, for $t \in [0,T]$,

$$\begin{aligned}
\left|\partial_w^2 X_t^{x,w}\right| &\leq \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\Bigg\{\left|\partial^2 \mathcal{V}_t^w\right|\left|\rho W_t + \sqrt{1-\rho^2}W_t^{S,*}\right| \\
&\qquad + \left|\int_0^t \left(\rho W_u + \sqrt{1-\rho^2}W_u^{S,*}\right)\left(\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\left(\frac{2\left(\partial \mathcal{V}_u^w\right)^2}{(\mathcal{V}_u^w)^3} - \frac{\partial^2 \mathcal{V}_u^w}{\left(\mathcal{V}_u^w\right)^2}\right) + \frac{\kappa\xi^0}{2}\partial^2 \mathcal{V}_u^w\right)\mathrm{d}u\right|\Bigg\} \\
&\leq \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}\|\Bigg\{\left(\kappa\theta - \frac{\epsilon^2}{4}\right)\frac{\left|W_t\right| + \left|W_t^{S,*}\right|}{(\mathcal{V}_t^{\sqrt{v_{min}}})^3} \\
&\qquad + \int_0^t \left(\left|W_u\right| + \left|W_u^{S,*}\right|\right)\left(\left(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\right)\frac{2+\kappa\xi^0}{(\mathcal{V}_u^{\sqrt{v_{min}}})^3} + \frac{1}{2}\frac{\left(\kappa\theta - \epsilon^2/4\right)^2}{(\mathcal{V}_u^{\sqrt{v_{min}}})^5}\right)\mathrm{d}u\Bigg\} \\
&=: \tilde{Z}_t
\end{aligned}$$

where $\tilde{Z}_t$ does not depend on (w,x). To prove the integrability of $\tilde{Z}_t$, we first observe that for all $t \in [0,T]$, there is some $C' > 0$ such that

$$\mathbb{E}^S\left[\frac{\left|W_t\right|}{(\mathcal{V}_t^{\sqrt{v_{min}}})^3}\right] \leq \mathbb{E}^S\left[\frac{1}{(\mathcal{V}_t^{\sqrt{v_{min}}})^{3q}}\right]^{1/q}\mathbb{E}^S\left[\left|W_t\right|^p\right]^{1/p} \tag{4.53}$$

for all $(p,q) \in [1,\infty)^2$ such that $1/p + 1/q = 1$. For ally choice of $p \geq 1$, $[0,T] \ni t \mapsto \mathbb{E}^S\left[\left|W_t\right|^p\right]$ is continuous and thus bounded independently of t. As previously, we pick some $\alpha' > 0$ such that $1 + \alpha' < \frac{4\kappa\xi^0\theta}{3\epsilon^2}$ which is possible thanks to (4.27); secondly, we select p large enough so that $q =$

$\frac{p}{p-1} \leq 1+\alpha'$. Hölder's inequality implies $\mathbb{E}^S\Big[(\mathcal{V}_t^{\sqrt{v_{min}}})^{-3q}\Big] \leq \mathbb{E}^S\Big[(\mathcal{V}_t^{\sqrt{v_{min}}})^{-3(1+\alpha')}\Big]^{\frac{q}{1+\alpha'}} \leq \frac{C''}{v_{min}^{3q/2}}$ following Lemma 4.12. This way, $\mathbb{E}^S\Big[\big|W_t\big|(\mathcal{V}_t^{\sqrt{v_{min}}})^{-3}\Big]$ is proved to be bounded independently of t. Similar study can be led on $\mathbb{E}^S\Big[\big|W_t^{S,*}\big|(\mathcal{V}_t^{\sqrt{v_{min}}})^{-3}\Big]$. Now, Fubini theorem, Hölder's inequality with again convenient choice of conjugates p, q, Lemma 4.12 again and the condition on volatility parameters $2\kappa\theta > \frac{5}{2}\epsilon^2$ yield that for all $t \in [0,T]$,

$$\mathbb{E}^S\Big[\int_0^t \big(|W_u| + |W_u^{S,*}|\big)\Big(\Big(\frac{\kappa\theta}{2} - \frac{\epsilon^2}{8}\Big)\frac{2+\kappa\xi^0}{(\mathcal{V}_u^{\sqrt{v_{min}}})^3} + \frac{1}{2}\frac{\big(\kappa\theta - \epsilon^2/4\big)^2}{(\mathcal{V}_u^{\sqrt{v_{min}}})^5}\Big)\mathrm{d}u\Big] \leq C < \infty$$

for some $C > 0$ that does not depend on (t,x,w) and $\mathbb{E}^S\big[\tilde{Z}_t\big] \leq C'' < \infty$ is deduced for some $C'' > 0$. With (4.51), note also that

$$\begin{aligned}
\big|\partial_w^2\tilde{u}(t,x,w)\big| &= \big|\mathbb{E}^S\left[\partial_w^2 f(X_t^{x,w})\right]\big| = \Big|\mathbb{E}^S\Big[f''(X_t^{x,w})\big(\partial_w X_t^{x,w}\big)^2 + f'(X_t^{x,w})\partial_w^2 X_t^{x,w}\Big]\Big| \\
&\leq \|f''\|_\infty \mathbb{E}^S\Big[\big(\partial_w X_t^{x,w}\big)^2\Big] + \|f'\|_\infty \mathbb{E}^S\left[\partial_w^2 X_t^{x,w}\right] \\
&\leq \|f''\|_\infty \mathbb{E}^S\big[Z_{2,t}\big] + \|f'\|_\infty \mathbb{E}^S\big[\tilde{Z}_t\big] \\
&\leq \|f''\|_\infty C'(T+T^2) + \|f'\|_\infty C''.
\end{aligned} \tag{4.54}$$

Combining now (4.48) and (4.54) we eventually obtain, for all time $t \in [0,T]$:

$$\begin{aligned}
&\big|\mathbb{E}^S\left[\big((V_t^J)^2+1\big)\partial_v^2\tilde{u}(t,X_t^J,V_t^J)\right]\big| \\
&= \Big|\mathbb{E}^S\Big[-\frac{\sqrt{V_t^J}+1/(V_t^J)^{3/2}}{4}\partial_w\tilde{u}(t,X_t^J,V_t^J) + \big((V_t^J)^2+1\big)\partial_w^2\tilde{u}(t,X_t^J,V_t^J)\Big]\Big| \\
&\leq C\Big(\mathbb{E}^S\Big[\big(V_t^J\big)^{1/2} + \big(V_t^J\big)^{-3/2}\Big] + \mathbb{E}^S\big[(V_t^J)^2+1\big]\Big) \\
&\leq C\Big(\mathbb{E}^S\Big[\big(V_t^J\big)^{1/2}\Big] + \big(v_{min}\big)^{-3/2} + \mathbb{E}^S\big[(V_t^J)^2\big] + 1\Big) \\
&\leq C < \infty
\end{aligned}$$

where C is some positive constant that changed from a line to another: it does depend on v_{min} but not on (t,w,x) nor on v_{max} thanks to Lemma 4.4.

Study of $\partial_{wx}^2\tilde{u}$. We start again with (4.44). The justification that

$$\partial_{xw}^2\tilde{u}(t,x,w) = \mathbb{E}^S\left[f''(X_t^{x,w})\partial_x X_t^{x,w}\partial_w X_t^{x,w} + f'(X_t^{x,w})\partial_{xw}X_t^{x,w}\right]$$

is a straightforward consequence of what precedes since the first variation processes with respect to x are defined as $\partial_x X_t^{x,w} \overset{a.s.}{=} 1$ and $\partial_x\big(\partial_w X_t^{x,w}\big) \overset{a.s.}{=} 0$. Note also that the regularity of $\tilde{u}$ implies $\partial_{xw}^2\tilde{u} = \partial_{wx}^2\tilde{u}$. It follows

$$\begin{aligned}
\big|\partial_{xw}^2\tilde{u}(t,x,w)\big| &= \Big|\mathbb{E}^S\left[f''(X_t^{x,w})\partial_x X_t^{x,w}\partial_w X_t^{x,w} + f'(X_t^{x,w})\partial_{xw}X_t^{x,w}\right]\Big| \\
&\leq \|f''\|_\infty \mathbb{E}^S\big[\big|\partial_w X_t^{x,w}\big|\big]
\end{aligned} \tag{4.55}$$

It follows that, for all $t \in [0,T]$

$$\begin{aligned}\left|\mathbb{E}^S\left[\left((V_t^J)^2+1\right)\partial^2_{xv}\tilde{u}(t,X_t^J,V_t^J)\right]\right| &= \left|\mathbb{E}^S\Big[\frac{\left(V_t^J\right)^{3/2}+1/\sqrt{V_t^J}}{2}\partial^2_{xw}\tilde{u}(t,X_t^J,V_t^J)\Big]\right| \\ &\leq \tilde{C}\mathbb{E}^S\Big[\left(V_t^J\right)^{3/2}+\left(V_t^J\right)^{-1/2}\Big] \\ &\leq \tilde{C}\big(\mathbb{E}^S\big[\left(V_t^J\right)^{3/2}\big]+v_{min}^{-1/2}\big) \\ &\leq \tilde{C} < \infty\end{aligned}$$

where $\tilde{C}$ is a finite positive constant changing from line to line that does not depend on (t,w,x) nor on v_{max} (still using Lemma 4.4) but does depend on v_{min}.

This conclude the proof.

□

4.4.1.5 On the upper bound parameter

Before illustrating our results, we discuss how the upper bound parameter v_{max} can be set in practice. The numerical efficiency of swaptions pricing is due to the use of Gram-Charlier expansion. In order to guarantee its robustness, the upper bound parameter v_{max} should be chosen relatively small so that (4.11) holds.
The upper bound parameter v_{max} can be considered as a *meta-parameter* whose value is set beforehand. To do so, one can rely on the different convergence results derived in this section and on the numerical analysis made below. Indeed, one could use the above-mentioned Proposition 4.8 to chose an appropriate pair (v_{max}, v_0) so that the pricing error ε_1 is smaller than a prescribed threshold. Alternatively, one can resort on the numerical experiments of Section 4.5 below and the associated conjecture of Remark 30 that seem to indicate a faster rate of convergence. As already pointed in [AFP17], the approximating process (4.7) behaves very closely to the original one (4.1) even for quite small v_{max}.

Another choice would be to include this coefficient v_{max} in the set of parameters to calibrate as illustrated in Table 4.4. Of course, this increases the computational time of the calibration but still provide satisfactory results in terms of market consistency.

4.4.2 Truncation error

We aim now at estimating ε_2 and if possible derive its dependency with respect to N. We rely on the series representation of Equation (4.19). We first examine the asymptotic behaviour of the sequence of Hermite moments $(h_n)_{n\in\mathbb{N}}$ which does not depend on the payoff function φ of the option we aim at pricing.

4.4.2.1 Some examples of Hermite moments

We first recall that for any $p \in \mathbb{N}$, $h_p = \mathbb{E}^S[\check{H}_p(S_T^{m,n,J})]$ where $\check{H}_p$ is the normalized Hermite polynomials.

Theorem 4.15. *Assume that $\sigma_r^2 > \lambda_{max}^2 v_{max} T$. If $v_{min} > 0$ and Assumption 2 holds, there exists some real number $q \in (0,1)$ and some positive number C such that for all $p \in \mathbb{N}$,*

$$h_p^2 \leq Cq^p.$$

If $v_{min} = 0$ and Feller condition (4.8) holds, there exists some random variable $Q \in (0,1)$ almost

surely and some positive number C' such that for all $p \in \mathbb{N}$,

$$h_p^2 \leq C' \mathbb{E}^S[Q^p].$$

Proof. For the first case when $v_{min} > 0$ and Assumption 2 hold, we refer to [AFP17] in which the proof is found.

We place ourselves in the case when $v_{min} = 0$ and we adapt the proof of [AFP17]. Denote by f_T the density of $S_T^{m,n,J}$ under $\mathbb{P}^S$. We recall that, under $\mathbb{P}^S$, the distribution of the swap rate can be decomposed as

$$S_T^{m,n,J} \overset{a.s.}{=} \tilde{S}_T + \int_0^T \sqrt{V_u - \rho(u)^2 Q(V_u)} \boldsymbol{\lambda}(u) \cdot \mathrm{d}\boldsymbol{W}_u^*,$$

where $\tilde{S}_T := S_0 + \int_0^T \mu(u, V_u)\mathrm{d}u + \frac{1}{\epsilon}\int_0^T \rho(u)\|\boldsymbol{\lambda}(u)\|\mathrm{d}V_u - \frac{\kappa}{\epsilon}\int_0^T \rho(u)\|\boldsymbol{\lambda}(u)\|\big(\theta - \xi(u)V_u\big)\mathrm{d}u$ and we recall the expression of the cumulated volatility process $\Xi_t = \int_0^t \|\boldsymbol{\lambda}^{m,n}(s)\|^2\big(V_s - \rho(s)^2 Q(V_s)\big)\mathrm{d}s,\ t \leq T$ (cf. (4.10)). We refer to the proof of Lemma 4.3 for the justification of the fact that $\tilde{S}_T$ can be bounded by a finite constant $\Lambda \in \mathbb{R}_+$: $\tilde{S}_T \leq \Lambda < \infty$ almost surely.

First observe that the definition of the Hermite moments implies that

$$\begin{aligned}
h_p &= \int_{\mathbb{R}} \check{H}_p(x) f_T(x)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi n!}} \int_{\mathbb{R}} H_p\left(\frac{x-\mu_r}{\sigma_r}\right) \mathbb{E}^S\left[\exp\left(-\frac{(x-\tilde{S}_T)^2}{2\Xi_T^2}\right) \Big/ \sqrt{\Xi_T^2}\right] \mathrm{d}x \\
&= \mathbb{E}^S\left[\frac{1}{\sqrt{2\pi n! \Xi_T^2}} \int_{\mathbb{R}} H_p\left(\frac{x-\mu_r}{\sigma_r}\right) \exp\left(-\frac{(x-\tilde{S}_T)^2}{2\Xi_T^2}\right) \mathrm{d}x\right] =: \mathbb{E}^S[\tilde{h}_p].
\end{aligned}$$

The quantity inside the last expectation can rewrite as

$$\begin{aligned}
\tilde{h}_p &\overset{a.s.}{=} \frac{1}{\sqrt{2\pi p!}} \int_{\mathbb{R}} H_p\left(\frac{x-\mu_r}{\sigma_r}\right) \frac{\exp\left(-\frac{(x-\tilde{S}_T)^2}{2\Xi_T^2}\right)}{\sqrt{\Xi_T^2}} \mathrm{d}s \\
&= \frac{1}{\sqrt{p!}} \int_{\mathbb{R}} H_p\left(\Sigma x + M\right) \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} \mathrm{d}x =: \frac{1}{\sqrt{p!}} \hat{h}_p,
\end{aligned}$$

in which the random variables $\Sigma = \frac{1}{\sigma_r}\sqrt{\Xi_T^2}$ and $M = \frac{1}{\sigma_r}(\tilde{S}_T - \mu_r)$ have been set. Properties of the Hermite polynomials can be exploited now: for $p \geq 1$, $H_p'(x) = pH_{p-1}(x)$ along with $H_p(x) = xH_{p-1}(x) - H_{p-1}'(x)$ imply that for $p \geq 2$,

$$\hat{h}_p \overset{a.s.}{=} M\hat{h}_{p-1} - (p-1)(1-\Sigma^2)\hat{h}_{p-2}.$$

This relationship can be solved as

$$\hat{h}_p \overset{a.s.}{=} p! \sum_{k=0}^{\lfloor p/2 \rfloor} \frac{(\Sigma^2 - 1)^k M^{p-2k}}{2^k k!(p-2k)!} \tag{4.56}$$

with in particular $\hat{h}_0 = 1$ and $\hat{h}_1 = M$. Furthermore $\lambda_{min}^2 \int_0^T (V_s - Q(V_s))\,\mathrm{d}s \leq \Xi_T^2 \leq$

$\lambda_{max}^2 \int_0^T V_s \mathrm{d}s \leq \lambda_{max}^2 T v_{max}$ since $|\rho| \leq 1$ which yields almost surely

$$q := 1 - \frac{\lambda_{max}^2}{\sigma_r^2} T v_{max} \leq 1 - \Sigma^2 \leq 1 - \frac{\lambda_{min}^2}{\sigma_r^2} \int_0^T \frac{V_s^2}{v_{max}} \mathrm{d}s =: Q,$$

where q is checked to belong to $(0,1)$ and the random variable Q is similarly checked to belong almost surely tp $(0,1)$ under the Feller condition. Reltionship (4.56) can be expressed in terms of Hermite polynomials through

$$\hat{h}_p \overset{a.s.}{=} p!(1-\Sigma^2)^{p/2} \sum_{k=0}^{\lfloor p/2 \rfloor} \frac{(-1)^k}{k!(p-2k)!} \frac{\big((1-\Sigma^2)^{-1/2} M\big)^{p-2k}}{2^k} = (1-\Sigma^2)^{p/2} H_p\big((1-\Sigma^2)^{-1/2} M\big)$$

so that

$$h_p = \frac{1}{\sqrt{p!}} \mathbb{E}^S \left[(1-\Sigma^2)^{p/2} \hat{H}_p((1-\Sigma^2)^{-1/2} M) \right].$$

We bring the so-called Cramér's inequality (see [EMOT55] or [AS64] for details): there exists some positive constant K' such that for any $x \in \mathbb{R}$, $\frac{1}{\sqrt{p!}} e^{-x^2/4} |H_p(x)| \leq K'$. This inequality along with the Cauchy-Schwarz's one lead to

$$\begin{aligned} h_p^2 &\leq \mathbb{E}^S \left[(1-\Sigma^2)^p \right] \mathbb{E}^S \left[\frac{1}{p!} \hat{H}_p \left((1-\Sigma^2)^{-1/2} M \right)^2 \right] \\ &\leq K'^2 \mathbb{E}^S \left[(1-\Sigma^2)^p \right] \mathbb{E}^S \left[\exp\left(\frac{M^2}{2(1-\Sigma^2)} \right) \right] \\ &\leq K'^2 \mathbb{E}^S \Big[\exp\Big(m^2/(2q_1) \Big) \Big] \times \mathbb{E}^S \left[Q^p \right] \end{aligned}$$

where we have denoted by $m = \frac{\Lambda + \mu_r}{\sigma_r} \in \mathbb{R}_+$. □

4.4.3 Expression of Fourier coefficients

In order to derive some truncation error, the asymptotic behavior of Fourier coefficients $(\varphi_n)_{n \in \mathbb{N}}$ is necessary. We derive their exact expressions for two payoff functions in this paragraph.

Call option We first examine the particular case of call option for which $\varphi(x) = (x-K)_+$ with K the strike.

Proposition 4.16. *For any $n \geq 2$, the Fourier coefficients express as*

$$\varphi_n = \frac{\sigma_r}{\sqrt{n!}} H_{n-2}\left(\frac{K - \mu_r}{\sigma_r} \right) \frac{e^{-(K-\mu_r)^2/(2\sigma_r^2)}}{\sqrt{2\pi}}.$$

In addition, we have $\varphi_0 = \sigma_r \Big(\frac{1}{\sqrt{2\pi}} \exp\big(- \frac{(K-\mu_r)^2}{2\sigma_r^2} \big) - \frac{K-\mu_r}{\sigma_r} \Phi\big(\frac{\mu_r - K}{\sigma_r} \big) \Big)$ and $\varphi_1 = \sigma_r \Phi\big(\frac{\mu_r - K}{\sigma_r} \big)$.

Proof. The computations exploit the key property of the Hermite polynomials:

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2} = \hat{H}_n(x) e^{-x^2/2}.$$

Let $n \geq 2$.

$$\begin{aligned}
\left\langle \varphi, \check{H}_n \right\rangle_{\mathcal{L}_r^2} &= \frac{1}{\sqrt{n!}} \int_{\mathbb{R}} \varphi(x) H_n(x) \phi_r(x) \mathrm{d}x \\
&= \frac{1}{\sqrt{n!}} \int_K^{+\infty} (x-K) H_n\left(\frac{x-\mu_r}{\sigma_r}\right) \frac{e^{-(x-\mu_r)^2/2\sigma_r^2}}{\sqrt{2\pi\sigma_r^2}} \mathrm{d}x \\
&= \frac{1}{\sigma_r\sqrt{n!}} \int_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} (\sigma_r u + \mu_r - K) H_n(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \sigma_r \mathrm{d}u \\
&= \frac{1}{\sigma_r\sqrt{n!}} \Big\{ \sigma_r^2 \int_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} u H_n(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u + \sigma_r(\mu_r - K) \int_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} H_n(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u \Big\} \\
&= \frac{1}{\sigma_r\sqrt{n!}} \Big\{ \sigma_r^2 \Big[-u H_{n-1}(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \Big]_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} + \sigma_r \int_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} H_{n-1}(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \sigma_r \mathrm{d}u \\
&+ (\mu_r - K)\sigma_r \int_{\frac{K-\mu_r}{\sigma_r}}^{+\infty} H_n(u) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u \Big\} \\
&= \frac{\sigma_r}{\sqrt{n!}} H_{n-2}(K^*) \frac{e^{-(K^*)^2/2}}{\sqrt{2\pi}}.
\end{aligned}$$

where we set $K^* = \frac{K-\mu_r}{\sigma_r}$. Similar computations gives the result for $n = 0$ and $n = 1$. □

As noted in [DABB17] and [AFP17], it is important to note that the first term, φ_0 corresponds to an adjusted (through the parameters (μ_r, σ_r)) Bachelier price. Further terms in the expansions are thus adjustments for the stochastic volatility components.

Asset-or-Nothing options We also provide the example of a digital option of payoff $\varphi(x) = x\mathbb{1}_{[k,\infty)}(x)$, for some prescribed threshold $k \in \mathbb{R}$.

Proposition 4.17. *For $n \geq 2$, the Fourier coefficients express as*

$$\varphi_n = \frac{\sigma_r e^{-k^2/2}}{\sqrt{2\pi n!}} \Big((\sigma_r k + \mu_r) H_{n-1}(k) + \sigma_r H_{n-2}(k) \Big).$$

In addition, we have $\varphi_0 = \sigma_r \big(\sigma_r \frac{e^{-k^2/2}}{\sqrt{2\pi}} + \mu_r \Phi(-k) \big)$ and $\varphi_1 = \sigma_r(\sigma_r k + \mu_r) \frac{e^{-k^2/2}}{\sqrt{2\pi}} + \sigma_r^2 \Phi(-k)$.

Proof. Similar computations as in proof of Proposition 4.16. □

4.4.4 Truncation error

We now move onto the analyze of the truncation error. We recall it expresses as

$$\begin{aligned}
\varepsilon_2 &= \left| \mathbb{E}^S[\varphi(S_T^{m,n,J})] - P_T^N(\varphi) \right| \\
&= \left| P_T(\varphi) - P_T^N(\varphi) \right| \\
&= \Big| \sum_{p=N+1}^{+\infty} h_p \varphi_p \Big|.
\end{aligned}$$

Our purpose here is to derive the behavior of ε_2 fpr N large enough. To do so, we will exploit the previous results, notably Theorem 4.15. Beforehand, let us recall the well known

Stirling's formula which will be repeatedly employed hereafter:

$$p! \underset{p\to\infty}{\sim} \sqrt{2\pi}\sqrt{p}e^{-p}p^p.$$

Call option Let us first examine a particular case of centered Gram-Charlier expansion i.e. set $\mu_r = \mathbb{E}^S[S_T^{m,n}] = S_0^{m,n}$. Let φ still be the payoff function of a swaption of strike K: $\varphi(x) = (x-K)_+$ whose T maturity date is T. The following theorem is based on the result derived in Theorem 4.15 therefore we will place ourselves under similar assumptions.

Theorem 4.18. *Suppose that $\sigma_r^2 > \lambda_{max}^2 v_{max} T$.*

i. Let us first consider case when $v_{min} > 0$ and Assumption 2 holds; we denote by $q \in (0,1)$ the real number appearing in Theorem 4.15 and set $\tilde{q} = \sqrt{q}$. In case when $K = S_0^{m,n}$ (ATM options), the truncation error can be bounded as

$$\varepsilon_2 \leq C\frac{\tilde{q}^{N+1}}{N^{1/4}}$$

for some positive constant C.
In case when $K \neq S_0^{m,n}$ (AFM options), the truncation error can be bounded as

$$\varepsilon_2 \leq \frac{C'}{1-\tilde{q}}\tilde{q}^{N+1}.$$

for some positive constant C'.

ii. Consider now case when $v_{min} = 0$ and Feller condition (4.8) holds; we denote by $Q \in (0,1)$ the random variable appearing in Theorem 4.15. In case when $K = S_0^{m,n}$ (ATM options), the truncation error can be bounded as

$$\varepsilon_2 \leq C\frac{\sqrt{\mathbb{E}^S[Q^{N+1}]}}{N^{1/4}}$$

for some positive constant C.
In case when $K \neq S_0^{m,n}$ (AFM options), the truncation error can be bounded as

$$\varepsilon_2 \leq C'\sqrt{\mathbb{E}^S\left[\frac{Q^{N+1}}{1-Q}\right]}.$$

for some positive constant C'.

Proof. Throughout this proof, we denote by C a generic positive real number that may change from line to line. We begin with the first case when $v_{min} > 0$ and Assumption 2 holds. Let us first examine the particular case of At-The-Money swaptions for which $K = S_0^{m,n} = \mathbb{E}^S[S_T^{m,n}] = \mu_r$. In that particular case, $\varphi_p = \frac{\sigma_r}{\sqrt{2\pi p!}}H_{p-2}(0)$ for $p \geq 2$. The explicit expression of Hermite

polynomials yields $\check{H}_p(0) = (-1)^{p/2}\frac{1}{\sqrt{p!}}\frac{p!}{2^{p/2}(p/2)!}$ for even p, $H_p(0) = 0$ for odd p. For $p \in \mathbb{N}$,

$$\begin{aligned}
\left|\varphi_p h_p\right| &= \left|\frac{\sigma_r}{\sqrt{2\pi}}\frac{(-1)^{(p-2)/2}}{\sqrt{p!}}\frac{(p-2)!}{2^{(p-2)/2}((p-2)/2)!}h_p \mathbb{1}_{\{p \text{ even}\}}\right| \\
&\leq \frac{C}{\sqrt{p!}}\frac{(p-2)!\sqrt{q}^p}{2^{(p-2)/2}((p-2)/2)!} \\
&\underset{p\to+\infty}{\sim} \frac{C\sqrt{2\pi}\sqrt{p-2}e^{-(p-2)}(p-2)^{p-2}\sqrt{q}^p}{2^{(p-2)/2}\sqrt{\sqrt{2\pi}}p^{1/4}e^{-p/2}p^{p/2}\sqrt{2\pi}\sqrt{\frac{p-2}{2}}e^{-(p-2)/2}(\frac{p-2}{2})^{(p-2)/2}} \\
&= \frac{C(p-2)^{(p-2)/2}\sqrt{q}^p}{p^{p/2+1/4}} \\
&\underset{p\to+\infty}{\sim} C\frac{\sqrt{q}^p}{p^{5/4}} \xrightarrow[p\to+\infty]{} 0.
\end{aligned}$$

Furthermore, observe that the sign of the sequence $(H_{2p}(0))_{p\in\mathbb{N}}$ alternates. Consequently

$$\begin{aligned}
\varepsilon_2 &\leq \frac{\sigma_r}{\sqrt{2\pi}}\Big(\left|H_{N-1}(0)\,\mathbb{E}^S\left[H_{N+1}(S_T^{m,n})\right]\right|\mathbb{1}_{\{N \text{ odd}\}} \\
&\qquad + \left|H_N(0)\,\mathbb{E}^S\left[H_{N+2}(S_T^{m,n})\right]\right|\mathbb{1}_{\{N \text{ even}\}}\Big) \\
&\leq C\left(\frac{\sqrt{(N-1)!}}{2^{(N-1)/2}\left(\frac{N-1}{2}\right)!}\tilde{q}^{N+1}\mathbb{1}_{\{N \text{ odd}\}} + \frac{\sqrt{N!}}{2^{N/2}\left(\frac{N}{2}\right)!}\tilde{q}^{N+2}\mathbb{1}_{\{N \text{ even}\}}\right).
\end{aligned} \tag{4.57}$$

With the Stirling's formula we get that $\frac{\sqrt{n!}}{2^{n/2}(n/2)!} \underset{n\to\infty}{\sim} \frac{C}{n^{1/4}}$ which prove the first claim.

Let us now turn to the case of Away-from-The-Money options. We can associated the Cramèr's inequality (see the end of the proof of Theorem 4.15) with the expression of the Fourier coefficients associated with call option obtained in Proposition (4.17) to get that for any $p \in \mathbb{N}$,

$$\varphi_p \leq \frac{\sigma_r K'}{\sqrt{2\pi p(p-1)}}e^{-(K^*/\sigma_r)^2/4} \leq \frac{C}{p}.$$

It implies

$$\varepsilon_2 \leq \sum_{p=N+1}^{+\infty}\left|\varphi_p h_p\right| \leq C\sum_{p=N+1}^{+\infty}\frac{\tilde{q}^p}{p} \leq C\sum_{p=N+1}^{+\infty}\tilde{q}^p = \frac{C}{1-\tilde{q}}\tilde{q}^{N+1}. \tag{4.58}$$

Regarding the case when $v_{min} = 0$, preceding arguments can be adapted. First in case of ATM options and still using results of Theorem 4.15, we get the existence a random variable $Q \in (0,1)$ almost surely such that for all $p \in \mathbb{N}$,

$$\left|\varphi_p h_p\right| \leq C\frac{\sqrt{\mathbb{E}^S[Q^p]}}{p^{5/4}} \xrightarrow[p\to+\infty]{} 0,$$

for some positive constant C. In that case, the truncation error can be bounded as in (4.57) to get

$$\begin{aligned}
\varepsilon_2 &\leq C\left(\frac{\sqrt{(N-1)!}}{2^{(N-1)/2}\left(\frac{N-1}{2}\right)!}\sqrt{\mathbb{E}^S[Q^{N+1}]}\mathbb{1}_{\{N \text{ odd}\}} + \frac{\sqrt{N!}}{2^{N/2}\left(\frac{N}{2}\right)!}\sqrt{\mathbb{E}^S[Q^{N+2}]}\mathbb{1}_{\{N \text{ even}\}}\right) \\
&\leq C\frac{\sqrt{\mathbb{E}^S[Q^{N+1}]}}{N^{1/4}}
\end{aligned}$$

For AFM options, the linearity of the expectation and Cauchy-Schwarz inequality yield similarly as in (4.58):

$$\varepsilon_2 \leq C' \mathbb{E}^S \left[\frac{Q^{N+1}}{1-Q} \right]^{1/2}.$$

This concludes the proof. □

4.5 Numerical analysis

The following numerical experiments were performed using a single 2.80 GHz 64 bits CPU and the R programming language.

4.5.1 Matrix exponential computation

To perform Gram-Charlier type expansion at an arbitrary order $k \in \mathbb{N}$, we need to represent the action of the infinitesimal generator through a matrix before taking its exponential. Consider the standard basis of $\mathcal{P}_k(\mathbb{R}^2)$ composed of monomial functions: $\big((v,s) \mapsto v^p s^q\big)_{p+q \leq k}$. Given an enumeration $\Lambda : E_k \to \mathbb{N}$ of the set of exponents $E_k = \big\{(p,q) \in \mathbb{N}^2 : p+q \leq k\big\}$, the non-zero elements of the matrix $A^{(k)}$ representing the action of $\mathcal{A}_t$ on $\mathcal{P}_k(\mathbb{R}^2)$ are given by:

$$\begin{aligned}
A^{(k)}_{\Lambda(p,q),\Lambda(p,q)} &= -\left(\frac{p(p-1)\epsilon^2}{2(\sqrt{v_{max}} - \sqrt{v_{min}})^2} + p\kappa\xi^0(t) \right), \\
A^{(k)}_{\Lambda(p-1,q),\Lambda(p,q)} &= \kappa p \theta + \frac{p(p-1)\epsilon^2(v_{min} + v_{max})}{2(\sqrt{v_{max}} - \sqrt{v_{min}})^2}, \\
A^{(k)}_{\Lambda(p-2,q),\Lambda(p,q)} &= -\frac{p(p-1)\epsilon^2 v_{min} v_{max}}{2(\sqrt{v_{max}} - \sqrt{v_{min}})^2}, \\
A^{(k)}_{\Lambda(p+1,q-2),\Lambda(p,q)} &= \frac{q(q-1)\|\boldsymbol{\lambda}^{m,n}(t)\|^2}{2}, \\
A^{(k)}_{\Lambda(p+1,q-1),\Lambda(p,q)} &= -\frac{pq\epsilon\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|}{(\sqrt{v_{max}} - \sqrt{v_{min}})^2}, \\
A^{(k)}_{\Lambda(p-1,q-1),\Lambda(p,q)} &= -\frac{pq\epsilon v_{min} v_{max}\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|}{(\sqrt{v_{max}} - \sqrt{v_{min}})^2}, \\
A^{(k)}_{\Lambda(p,q-1),\Lambda(p,q)} &= \frac{pq\epsilon\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|(v_{min} + v_{max})}{(\sqrt{v_{max}} - \sqrt{v_{min}})^2}.
\end{aligned}$$

4.5.2 Parametrization of the DDSVLMM and its Jacobi version

In our settings, the vectors and multi-dimensional Brownian motions are of length 2. We consider a piecewise constant parametrization of the time dependency. This set-up corresponds to a *frozen* approximation of the Libor Market Model as motivated in [BGaM97], [BM07] or [WZ06]. The coefficients appearing in dynamics (4.1) and (4.7) are defined as:

$$\xi^0(t) = 1 + \frac{\epsilon}{\kappa} \sum_{j=m}^{n-1} \alpha_j(0)\xi^0_j(t),\ \xi^0_j(t) = \sum_{k=1}^{j} \frac{\Delta T_k \big(F_k(0) + \delta\big)}{1 + \Delta T_k F_k(0)} \rho_k(t)\|\boldsymbol{\gamma}_k(t)\|, \tag{4.59}$$

and

$$\boldsymbol{\lambda}^{m,n}(t) = \sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}(t),\ \rho(t) = \frac{1}{\|\boldsymbol{\lambda}^{m,n}(t)\|} \sum_{j=m}^{n-1} \omega_j(0)\,\|\boldsymbol{\gamma_j}(t)\|\,\rho_j(t). \tag{4.60}$$

The quantities $(F_j(0))_{j=m,\dots,n}$ are forward rates quoted on markets and the coefficients ω_j defined for $m \leq j \leq n-1$, by

$$\omega_j(0) := \frac{\Delta T_j P(0,T_{j+1})}{B^S(0)} \left(1 + \frac{\Delta T_j}{1+\Delta T_j F_j(0)} \sum_{l=m}^{j-1} \left(F_l(0) - S_0^{m,n}\right)\right)(F_j(0)+\delta),$$

$\delta \in \mathbb{R}$ is a parameter often named *shift*. The volatility vectors are specified as $\boldsymbol{\gamma_j}(T_i) = g(T_j - T_i)\boldsymbol{\beta_{j-i+1}}$ over the interval $[T_i, T_{i+1}[$. Parameters are set annually piecewise constant: $T_{i+1} - T_i = 1$. The $\boldsymbol{\beta_k}$ are 2−dimensional vectors with unitary Euclidian norm, while $g(u) = (bu+a)e^{-cu} + d$, where a, b, c and d are non-negative constants. Finally, the coefficients ρ_j are parametrized thanks to a coefficient $\rho \in [-1, 1]$ as

$$\rho_j(t) = \frac{\rho}{\sqrt{2}} \frac{\boldsymbol{\gamma_j}^{(1)}(t) + \boldsymbol{\gamma_j}^{(2)}(t)}{\|\boldsymbol{\gamma_j}(t)\|}.$$

Theoretical assumptions for Gram-Charlier convergence

We rewrite the theoretical assumptions made in Section 4.3. First about Assumption 2: we assume that $\rho^2 < 1$. In view of the form of the correlation coefficients ρ_k, we have that

$$\rho(t) = \frac{\rho}{\sqrt{2}\|\boldsymbol{\lambda}^{m,n}(t)\|} \sum_{j=m}^{n-1} \omega_j(0) \left(\boldsymbol{\gamma_j}^{(1)}(t) + \boldsymbol{\gamma_j}^{(2)}(t)\right).$$

Observe that for any time t,

$$\begin{aligned}\frac{1}{\rho(t)^2} &= \frac{2\|\boldsymbol{\lambda}^{m,n}(t)\|^2}{\rho^2\left(\sum_{j=m}^{n-1} \omega_j(0)\left(\boldsymbol{\gamma_j}^{(1)}(t) + \boldsymbol{\gamma_j}^{(2)}(t)\right)\right)^2} \\ &= \frac{2}{\rho^2} \frac{\left(\sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(1)}(t)\right)^2 + \left(\sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(2)}(t)\right)^2}{\left(\sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(1)}(t) + \sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(2)}(t)\right)^2} \geq \frac{1}{\rho^2},\end{aligned}$$

and thus Assumption 2 is satisfied. Assumption 1, that is $4\kappa\theta > \epsilon^2$, is straightforward to check. Besides Assumption 1 and Assumption 2, the following has to hold in order to ensure the convergence of the Gram-Charlier expansion (see Theorem 4.2 and assumption in Equation (4.11)):

$$\sigma_r^2 > \frac{T v_{max}}{2} \lambda_{max}^2 \Leftrightarrow \frac{2\sigma_r^2}{T v_{max}} > \lambda_{max}^2$$

where we recall that T is the maturity of the priced derivative and $\lambda_{max}^2 = \max\Big\{t \geq 0 : \|\boldsymbol{\lambda}^{m,n}(t)\|^2 = \left(\sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(1)}(t)\right)^2 + \left(\sum_{j=m}^{n-1} \omega_j(0)\boldsymbol{\gamma_j}^{(2)}(t)\right)^2\Big\}$. This constraint can be numerically checked.

4.5.3 Why bounding the volatility factor?

We discuss in present paragraph the underlying assumption to the dynamics (4.7): the volatility factor is bounded. Even if we are interested in modelling and pricing issues, we first take a look at historical market data to check if this assumption is reasonable on past beliefs of economic

agents. Secondly, we will calibrate the standard dynamics (4.1) on those past data and we study the distribution induced by the calibrated parameters of the volatility factor. To do so we recall that distribution of the Cox-Ingersoll-Ross process as a function of its parameters is well known (see Proposition 1.2.11 in [Alf15]).

4.5.3.1 From an historical point of view

To check the boundedness assumption, the difficulty comes from the fact that the stochastic volatility factor is not a directly observable quantity. As pointed out in [AFP17] (Theorem 3.6), the implied volatility when using the Jacobi dynamics for the volatility factor is bounded. In the following Figure 4.1, the empirical density of implied volatilities (Bachelier convention) associated with swaptions of various settings (various maturities/tenors, strikes) observed on Euro market between August 14th, 2012 and January 25th, 2021. On this historical density, the following estimates are obtained: 0.54% of the observed points are greater than 100bps, 0.16% are greater than 120bps, 0.04% greater than 150bps, 0.01% greater than 180bps and 0.004% greater than 200bps. Empirical density of the implied volatilities thus indicates that assuming the stochastic volatility factor to be bounded is acceptable.

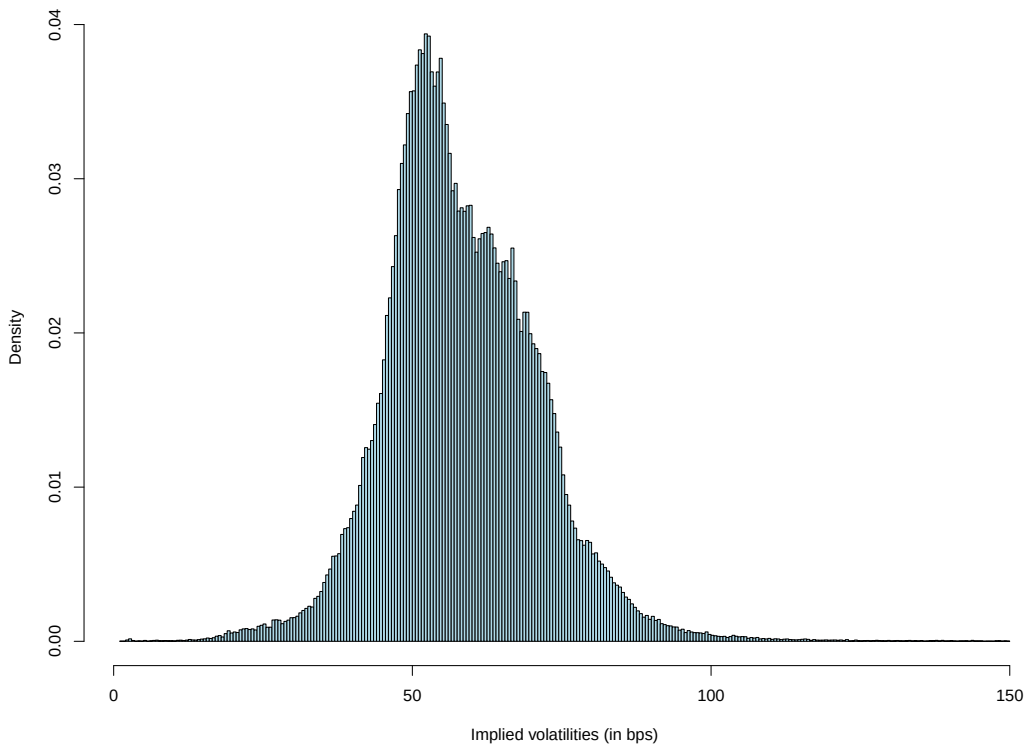


Figure 4.1: Empirical densities of implied volatilities on swaptions in the Euro market (in bps).

4.5.3.2 From a modelling point of view

We now take a modelling point of view. We have calibrated the standard DDSVLMM (4.1) on the end of year 2017, 2018, 2019 and 2020. Since the density of the CIR process is analytically known as a function of the CIR parameters, we have been able to compute the following

probabilities through analytical formulas. In Table 4.1, T is referred to as "Maturity" and M is the "threshold". Note that the initial value of the volatility process V_0 has been fixed to 1 in those experiments. It is observed that for quite small values of the ratio M/V_0 (from 1.5), probabilities for the CIR for going beyond the threshold M are quite low (at most around 5%). Calibrated parameters induces thus very thin tails on the densities of the CIR process illustrating that from a modelling point of view, replacing it by a bounded process has little impact on densities.

$\mathbb{P}\big(V_T \geq M\big)$ for various pairs (T, M).				
Threshold \ Maturity	1	5	10	20
2017				
0.8	0.7957	0.4821	0.3459	0.2152
1	0.4032	0.2921	0.2199	0.1390
1.5	0.005775	0.05292	0.05777	0.04313
2	4.327e-06	0.005884	0.01225	0.01231
2018				
0.8	0.7601	0.4538	0.3224	0.1980
1	0.3988	0.2834	0.2097	0.1296
1.5	0.01053	0.06118	0.06122	0.04259
2	2.728e-05	0.009035	0.01517	0.01319
2019				
0.8	0.8124	0.4035	0.2312	0.09753
1	0.3523	0.1986	0.1163	0.04775
1.5	0.001057	0.01661	0.01500	0.006994
2	4.508e-08	0.0006627	0.0013801	0.0008922
2020				
0.8	0.8543	0.4438	0.2614	0.1149
1	0.3585	0.2089	0.1265	0.05459
1.5	0.0003614	0.01275	0.01335	0.007027
2	1.559e-09	0.0002974	0.0009026	0.0007471

Table 4.1: Right tails distributions of Cox-Ingersoll-Ross processes.

4.5.4 Convergence illustrations

4.5.4.1 Convergence of the Jacobi process towards the CIR process

We start by providing some illustrations regarding convergence results on the volatility processes in case when $v_{min} = 0$ and $v_{max} \to +\infty$.

Strong error The $\mathbb{L}^1$ convergence of the Jacobi process towards the CIR has been theoretically discussed in Section 4.4.1. It is illustrated here. Beforehand, we describe the discretization schemes used to simulate volatility process. We chose to adapt the scheme presented in [LKD10] as follows:

$$V_{t_{i+1}} = V_{t_i} + \kappa\Big(\theta - \xi^0(t_i)\big((V_{t_i})_+ \wedge v_{max}\big)\Big)(t_{i+1} - t_i) + \epsilon\sqrt{Q\big((V_{t_i})_+ \wedge v_{max}\big)}(W_{t_{i+1}} - W_{t_i}), \tag{4.61}$$

with $x_+ = \max(x,0)$ and $x \wedge y = \min(x,y)$. CIR process is simulated by setting $v_{max} = \infty$ (and thus $Q(v) = v$ and $x \wedge v_{max} = x$) in this scheme.

In the following, all the expectations we are interested in are estimated using 10^5 Monte-Carlo samples, unless otherwise stated. When no confidence intervals are plotted it means that those could not be distinguished from the estimated curves in the plots. All our numerical results are obtained by setting $v_{min} = 0$ as theoretical properties are of particular interest for us in this case. The basic parametrization for the simulations below is the following: time horizon is fixed to $T = 5$ years, the volatility process parameters are $\kappa = 1$, $\theta = 0.3$, $\epsilon = 0.6$, $V_0^C = V_0^{J,v_{max}} = v_0 = 0.2$ and we set $\xi_0(t) \equiv 1$ for simplicity; the discretization time step is constant: $\forall i, t_{i+1} - t_i = \Delta t = 0.05$. Those parameters values may change for sensitivity analysis. Regarding the dependency towards to the upper bound parameter of the Jacobi process v_{max}, it will take values in

$$\{0.5, 0.7, 1.0, 1.2, 1.5, 1.8, 2, 3, 4, 5, 10, 10^2, 10^3, 10^4, 10^5\} \tag{4.62}$$

in our study.

We recall that Theorem 4.5 provides the following bounds: $\sup_{0\le t\le T} \mathbb{E}^S\left[|V_t^{J(p)} - V_t^C|\right] \le C/\log\left(\frac{v_{max}^{(p)}}{v_0}\right)$ and $\mathbb{E}^S\left[\sup_{0\le t\le T}|V_t^{J(p)} - V_t^C|\right] \le C'/\sqrt{\log\left(\frac{v_{max}^{(p)}}{v_0}\right)}$. In Figure 4.2, we plot the product $\sqrt{\log\left(\frac{v_{max}^{(p)}}{v_0}\right)} \times \mathbb{E}^S\left[\sup_{0\le t\le T}|V_t^{J(p)} - V_t^C|\right]$ as a function of $\log\left(v_{max}^{(p)}/v_0\right)$ and for different parametrizations. In particular, we study the impact of the vol-of-vol parameter $\epsilon \in \{0.6, 0.15, 2.4\}$ and the impact of the discretization time step $\Delta t \in \{0.05, 0.005, 1\}$. It turns out that the products remain bounded and converge towards zero as v_{max} increases in all cases. The case $\epsilon = 2.4$ is particular as the product first increases before going to zero. Observe that in this particular case the Feller condition does not hold. Regarding the impact of the discretization time step, it appears to be marginal. We also have investigated the impact of the discretization scheme. Alternatively, the Jacobi/CIR processes has been simulated using scheme from [DD+98]

$$V_{t_{i+1}} = V_{t_i} + \kappa\Big(\theta - \xi^0(t_i)V_{t_i}\Big)(t_{i+1} - t_i) + \epsilon\sqrt{Q\big((V_{t_i})_+ \wedge v_{max}\big)}(W_{t_{i+1}} - W_{t_i}), \tag{4.63}$$

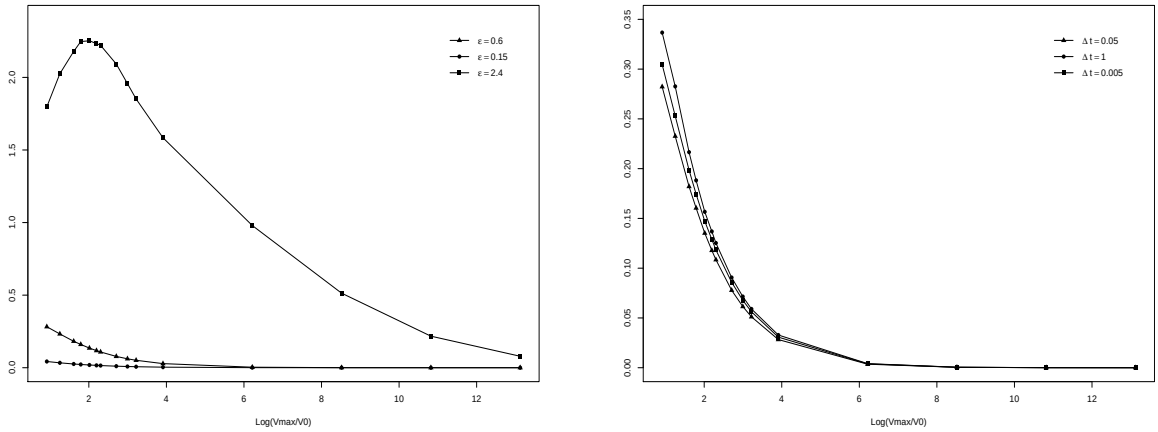
and using a scheme adapted from [BBD08]

$$V_{t_{i+1}} = \min\Big(\big|V_{t_i} + \kappa\big(\theta - \xi^0(t_i)V_{t_i}\big)\big)(t_{i+1} - t_i) + \epsilon\sqrt{Q\big(V_{t_i}\big)}(W_{t_{i+1}} - W_{t_i})\big|, v_{max}\Big). \tag{4.64}$$

These schemes produce results that are really close to those obtained with (4.61): we observed a maximum relative difference of 1.17% for $v_{max} = 10^4$ between schemes (4.63) and (4.61) and of 1.153% for $v_{max} = 10^5$ between schemes (4.64) and (4.61). Consequently, we do not provide plots using these schemes as the curves almost coincide. More elaborated numerical schemes may be adapted from those found in [Alf15] for the CIR process.

In Figure 4.3, we depict $\log\left(v_{max}^{(p)}/v_0\right) \times \mathbb{E}^S\left[|V_T^{J(p)} - V_T^C|\right]$ for maturities $T = 1, 10$ as a function of $\log\left(v_{max}^{(p)}/v_0\right)$. Similarly, we test different parametrizations. Here again, the products remain well bounded as v_{max} increases. As already observed in [AFP17], Jacobi process behaves very closely to CIR even for relatively small values of v_{max} (and of the ratio v_{max}/v_0).

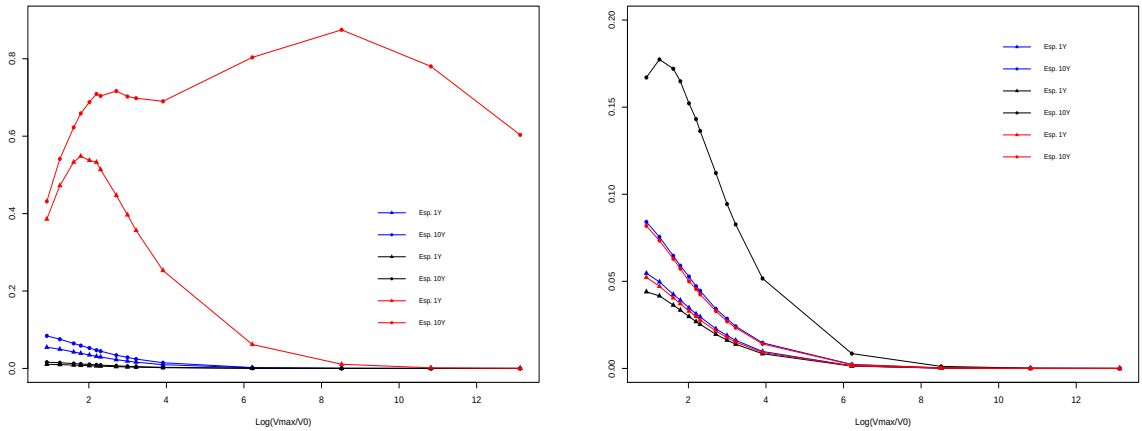
Investigations on a better convergence rate We now discuss the optimality of the convergence rate coming from Theorem 4.5. In left panel of Figure 4.4 the logarithms of previous esti-



(a) Sensitivity with respect to ϵ.

(b) Sensitivity with respect to Δt.

Figure 4.2: $\sqrt{\log\left(v_{max}^{(p)}/v_0\right)} \times \mathbb{E}^S[\sup_{0\leq s\leq 5}|V_s^{J(p)} - V_s^C|]$ with respect to $\log\left(v_{max}^{(p)}/v_0\right)$ for different values of the vol-of-vol parameter and of the time step of the discretization scheme.



(a) Sensitivity towards ϵ.

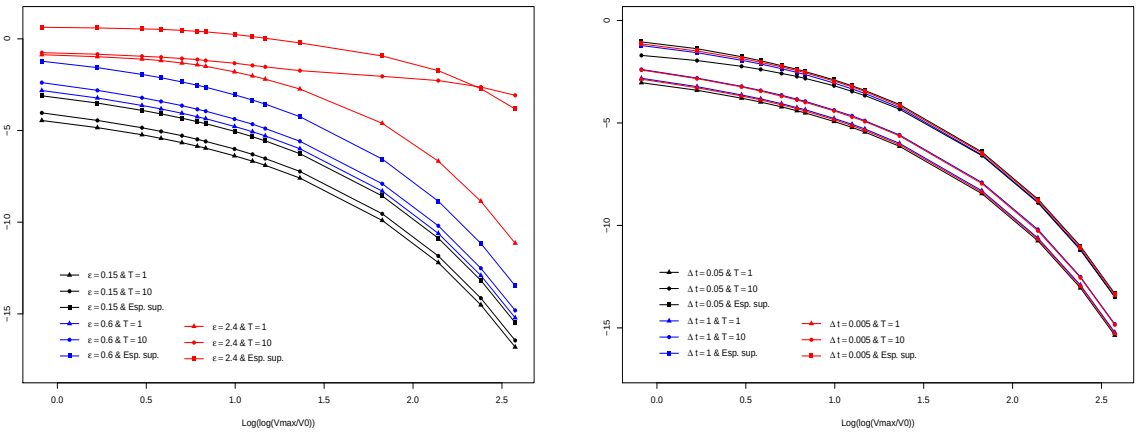
(b) Sensitivity towards Δt

Figure 4.3: $\log\left(v_{max}^{(p)}/v_0\right)\mathbb{E}^S[|V_T^{J(p)} - V_T^C|], T = 1, 10$ with respect to $\log\left(v_{max}^{(p)}/v_0\right)$ for different values of the vol-of-vol parameter and of the step time of the discretization scheme as $v_{max} \to +\infty$.

mates are plotted, that is $\log\left(\mathbb{E}^S[\sup_{0\leq s\leq 5}|V_s^{J(p)} - V_s^C|]\right)$ and $\log\left(\mathbb{E}^S[|V_T^{J(p)} - V_T^C|]\right)$, $T = 1, 10$ with respect to $\log\left(\log(v_{max}^{(p)}/v_0)\right)$ for the same parametrizations as above. The value of the vol-of-vol parameter has a significant impact of the global level of the errors but not on their behaviour as v_{max} increases, except for $\epsilon = 2.4$ for which Feller condition is not satisfied. The impact of the value of the step time appears as negligible for both the overall behavior and the level. The results indicate that the convergence rates estimated in Theorem 4.5 are not optimal, as the plots are clearly non linear.

We examine thus if the optimal convergence rate can be assimilated to a (negative) power of v_{max}. By analyzing the dependency of the logarithms of the empirical errors with respect to $\log\left(v_{max}/v_0\right)$ in Figure 4.5, we observe now an almost perfect linear behaviours, except again

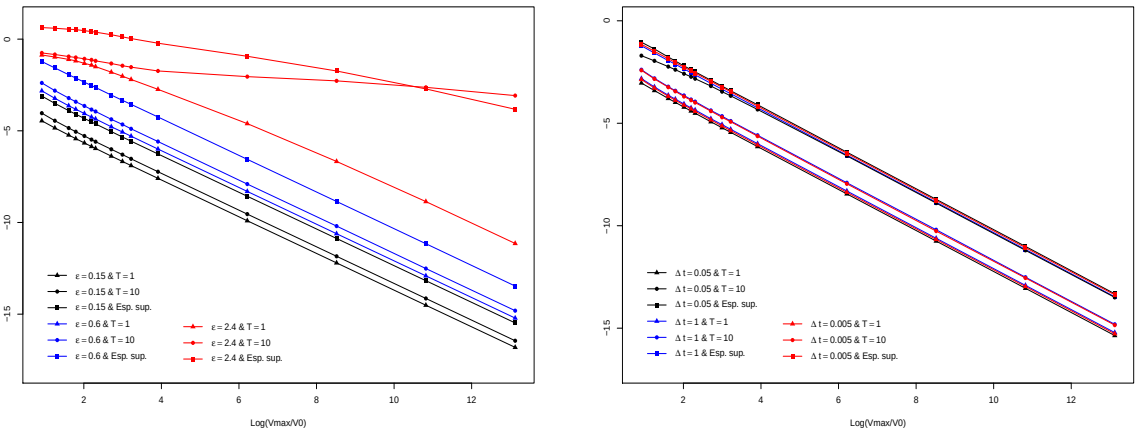
for $\epsilon = 2.4$. Performing linear regressions indicate slopes of around -1.0 for all considered errors with very high accuracy for each ($R^2 \geq 0.999$) which allows to conjecture that, whenever the Feller condition holds, the optimal rate would be such that Error $\propto \frac{1}{v_{max}/v_0}$.



(a) Sensitivity with respect to ϵ.

(b) Sensitivity with respect to Δt.

Figure 4.4: $\log \mathbb{E}^S\left[\sup_{0\leq s\leq 5} |V_s^{J(p)} - V_s^C|\right]$ and $\log \mathbb{E}^S\left[|V_T^{J(p)} - V_T^C|\right]$, $T = 1, 10$, as a function of $\log\left(\log\left(v_{max}^{(p)}/v_0\right)\right)$ for different values of ϵ and Δt.

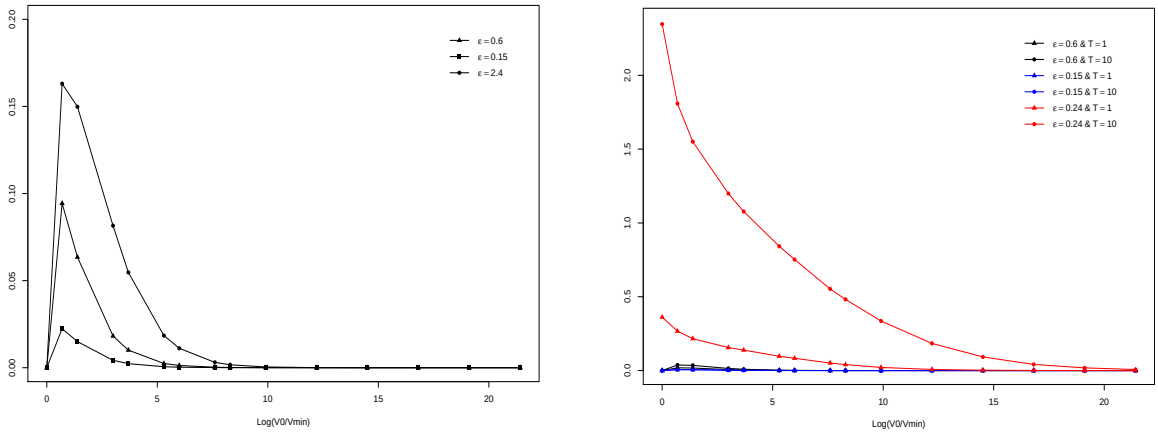


(a) Sensitivity with respect to ϵ.

(b) Sensitivity with respect to Δt.

Figure 4.5: $\log \mathbb{E}^S\left[\sup_{0\leq s\leq 5} |V_s^{J(p)} - V_s^C|\right]$ and $\log \mathbb{E}^S[|V_T^{J(p)} - V_T^C|]$, $T = 1, 10$, as a function of $\log\left(v_{max}^{(p)}/v_0\right)$ for different values of ϵ and Δt.

We now provide similar results in case when $v_{max} = \infty$ and $v_{min} \to 0$. As the step time of discretization scheme has not great impact, we focus in this on sensitivities with respect to other parameters of the model. In Figure 4.6 we illustrate Theorem 4.5. Again, numerical results reinforce theoretical ones. Graphs in Figure 4.7 also suggest that a better convergence can be found when Feller condition holds.



(a) $\log\left(v_0/v_{min}^{(p)}\right)^{1/2}\mathbb{E}^S\left[\sup_{0\leq s\leq 5}|V_s^{J(p)} - V_s^C|\right]$ with respect to $\log\left(v_0/v_{min}^{(p)}\right)$ for different values of the vol-of-vol parameter ϵ.

(b) $\log\left(v_0/v_{min}^{(p)}\right)\mathbb{E}^S\left[|V_T^{J(p)} - V_T^C|\right], T = 1, 10$ with respect to $\log\left(v_0/v_{min}^{(p)}\right)$ for different values of the vol-of-vol parameter ϵ.

Figure 4.6: Strong convergence as $v_{min} \to 0$.

Weak convergence Results relative to weak approximation of (4.1) by (4.7) are illustrated in Figure 4.8. First, some illustrations of the weak convergence of the Jacobi process towards the Cox-Ingersoll-Ross one (see Theorem 4.7). We test the convergence of $\mathbb{E}^S[f(V_T^J)]$ towards $\mathbb{E}^S[f(V_T)]$ for different functions f. The theoretical convergence rate proportional to $1/v_{max}$ obtained in of Theorem 4.7 seems to be acceptable for a wider class of functions than the one set in the statement of the theorem: in Fig. 4.8b, the payoff function is not of polynomial growth, in Fig. 4.8c the payoff function is not continuous and in Fig. 4.8d the payoff function explodes near zero. The case $f(x) = 1/x$ displayed in Figure 4.8a is remarkable as the confidence interval is too wide to conclude on the empirical convergence.

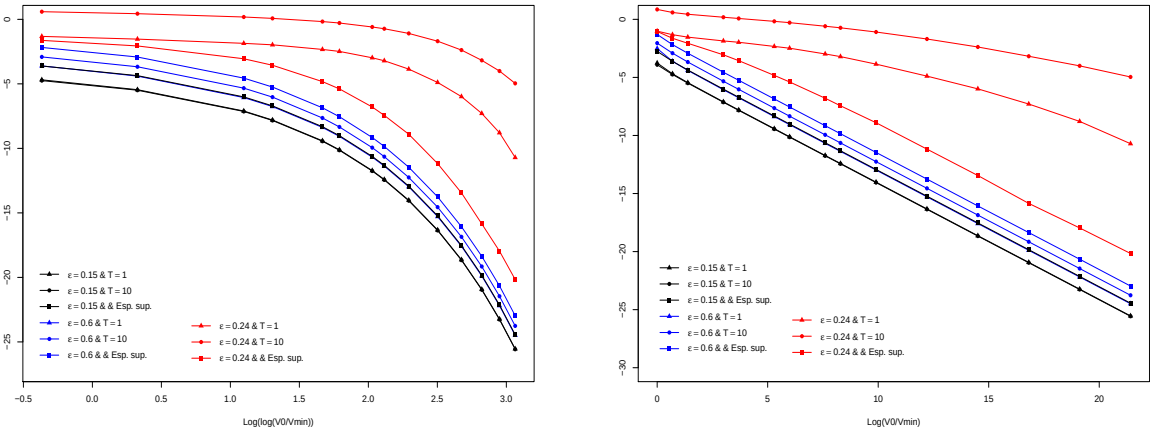
4.5.4.2 Convergence of the swap rate processes

Let us now move onto the case of the convergence on swap rates. The volatility scheme has been introduced above in Equation (4.61); the numerical scheme we used for the swap rate writes

$$S_{t_{i+1}}^{m,n} = S_{t_i}^{m,n} + \sqrt{Q\Big(\min\big((V_{t_i})_+, v_{max}\big)\Big)}\rho(t_i)\|\boldsymbol{\lambda}^{m,n}(t_i)\|(W_{t_{i+1}} - W_{t_i})$$
$$+ \sqrt{\min\big((V_{t_i})_+, v_{max}\big) - \rho(t_i)^2 Q\Big(\min\big((V_{t_i})_+, v_{max}\big)\Big)}\boldsymbol{\lambda}^{m,n}(t_i)\cdot(\boldsymbol{W}_{t_{i+1}}^{S,*} - \boldsymbol{W}_{t_i}^{S,*}).$$

Strong convergence In Theorem 4.8, we derive a strong convergence rate on the swap rate process itself to deduce a pricing error. We illustrate strong convergence in Figure 4.9. In Figure 4.9, the logarithms of $\mathbb{E}^S\left[|S_T^{m,n,J} - S_T^{m,n}|\right]$ and $\mathbb{E}^S\left[\sup_{u\leq T'}|S_u^{m,n,J} - S_u^{m,n}|\right]$ are displayed for different parametrizations (given below) and values of v_{max} taken in (4.62). In particular, we assess the impact of the volatility on the convergence. For this example we took constant parameters as:

$$(\|\lambda^{m,n}\|, \rho, \kappa, \theta, \epsilon, v_0, s_0) \equiv (0.15, 0.5, 1, 0.3, 0.6, 0.2, 0.01) \text{ (in black below)},$$
$$(\|\lambda^{m,n}\|, \rho, \kappa, \theta, \epsilon, v_0, s_0) \equiv (0.60, 0.5, 1, 0.3, 2.4, 0.2, 0.01) \text{ (in red below)},$$
$$(\|\lambda^{m,n}\|, \rho, \kappa, \theta, \epsilon, v_0, s_0) \equiv (0.05, 0.5, 1, 0.3, 0.15, 0.2, 0.01) \text{ (in blue below)}.$$



(a) $\log \mathbb{E}^S\left[|V_T^{J(p)} - V_T^C|\right]$, $T = 1, 10$ and $\log \mathbb{E}^S\left[\sup_{0\leq s\leq 5} |V_s^{J(p)} - V_s^C|\right]$ as functions of $\log\left(\log\left(v_0/v_{min}^{(p)}\right)\right)$ for different values of ϵ.

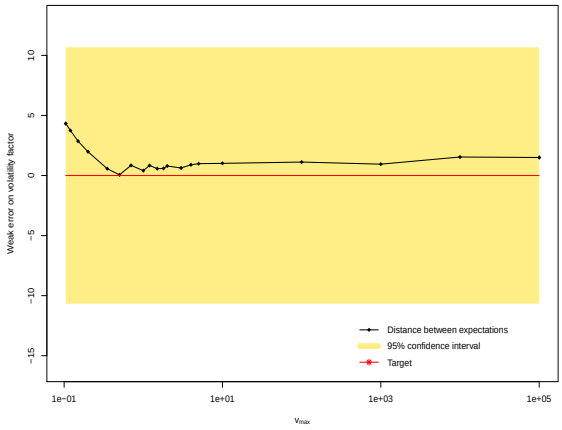
(b) $\log \mathbb{E}^S\left[|V_T^{J(p)} - V_T^C|\right]$, $T = 1, 10$ and $\log \mathbb{E}^S\left[\sup_{0\leq s\leq 5} |V_s^{J(p)} - V_s^C|\right]$ as functions of $\log\left(v_0/v_{min}^{(p)}\right)$ for different values of ϵ.

Figure 4.7: Convergence rates as $v_{min} \to 0$.

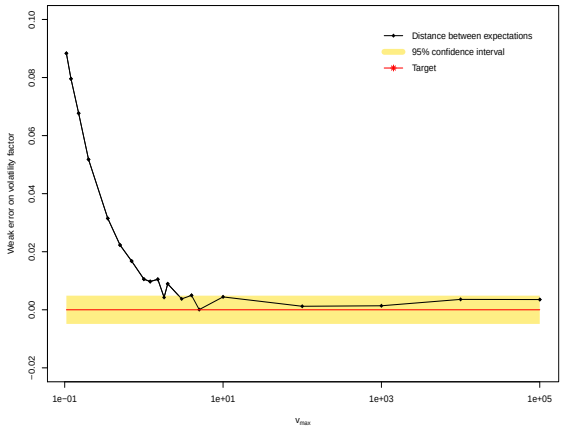
Conclusions are similar to those of the strong convergence of the volatility processes: numerical experiments suggst that a better convergence rate than the one derived in Theorem 4.8 such that Error $\propto 1/v_{max}$ as $v_{max} \to +\infty$ in case when Feller condition holds.

Weak convergence We first depict the convergence of moments of $S^{m,n,J}$ towards those of $S^{m,n}$ as stated in Proposition 4.9. In Figure 4.10, we display the logarithm of $\left|\mathbb{E}^S\left[\left(S_1^{m,n,J}\right)^p\right] - \mathbb{E}^S\left[\left(S_1^{m,n}\right)^p\right]\right|$ as a function of v_{max} (logarithm scale on x-axis) and for different values of p. Those results suggest that the theoretical rate is optimal. Observe that in Figure 4.10b, parameters are the same as in Figure 4.10a but the vol-of-vol parameter ϵ: it is such that for small values of v_{max}, the Feller condition (3.8) is not satisfied. For small v_{max}, it is then observed that the distance between moments of each process does not decrease. However, as v_{max} goes to infinity the error eventually behaves as expected. Interestingly, performing a linear regression of the whole curve still provide consistent results.

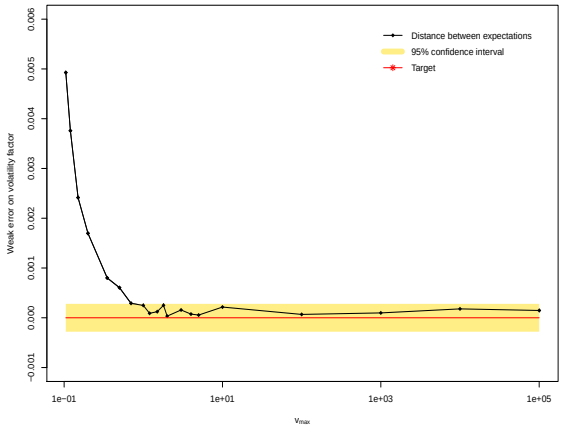
Let us now discuss main result on pricing error stated in Theorem 4.11 illustrated in Figure 4.11. The payoff functions we chose do not satisfy all the restrictions of the statements. Indeed, the parametrization we chose is so that major part of empirical distribution of the swap rate used to obtain visuals of Figure 4.11 is on negative line close to zero. In particular, denominator of the payoff function associated with Figure 4.11c is «often» close to zero. Even though the payoff functions in Figures 4.11a, 4.11b and 4.11c we chose are smooth, they do not necessarily satisfy the restricted polynomial growth condition required in Theorem 4.11. Conversely, Figure 4.11d is associated with a function whose asymptotic growth do satisfy the hypothesis of theorem but that is not smooth enough.
We now provide some convergence illustrations on prices of swap rates derivatives in Figure 4.11. For all considered payoff functions, the Jacobi induced prices are out of the confidence interval built around the reference price obtained in (4.1) for small values of v_{max} but reach it very quickly. The regularity of the payoff function seems to have few impact on this behaviour. Though, even for small upper bounds parameters, v_{max} is quite close from the confidence interval. It is in line with comments made in this thesis but also in also in [AFP17]: distribution



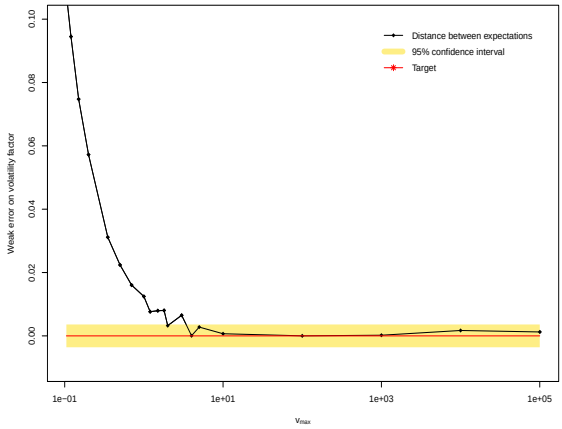
(a) $f(x) = \frac{1}{x}$



(b) $f(x) = e^{10x}$



(c) $\varphi(x) = x\mathbb{1}_{x\geq 0.05}$



(d) $\varphi(x) = \log(|x|)$

Figure 4.8: $\left|\mathbb{E}^S\left[f(V_1)\right] - \mathbb{E}^S\left[f(V_1^J)\right]\right|$ for different payoff functions f as a function of v_{max} (logarithmic scale on x-axis). Estimations made on 10^5 Monte-Carlo paths.

of $S^{m,n,J}$ is surprisingly very closely to that $S^{m,n}$ even for small values of v_{max}. Note that in present experiment, we took $v_0 = 0.1$ and v_{max} took values in

$$\left\{0.105, 0.12, 0.15, 0.2, 0.35, 0.5, 0.7, 1, 1.2, 1.5, 1.8, 2, 3, 4, 5, 10, 10^2, 10^3, 10^4, 10^5\right\},$$

so that the first values taken by the ratio v_0/v_{max} is close to one; reducing this ratio leads to an even quicker convergence of the distributions. Finally, we underline that those experiments were obtained by setting $v_{min} = 0$: though we have not been able to set $v_{min} = 0$ in the statement of the Theorem 4.11, those illustrations pledge in favour of an extension of the result to that particular case.

4.5.5 Pricing with Gram-Charlier series

We now turn to the illustrations of the Gram-Charlier pricing of swaptions.

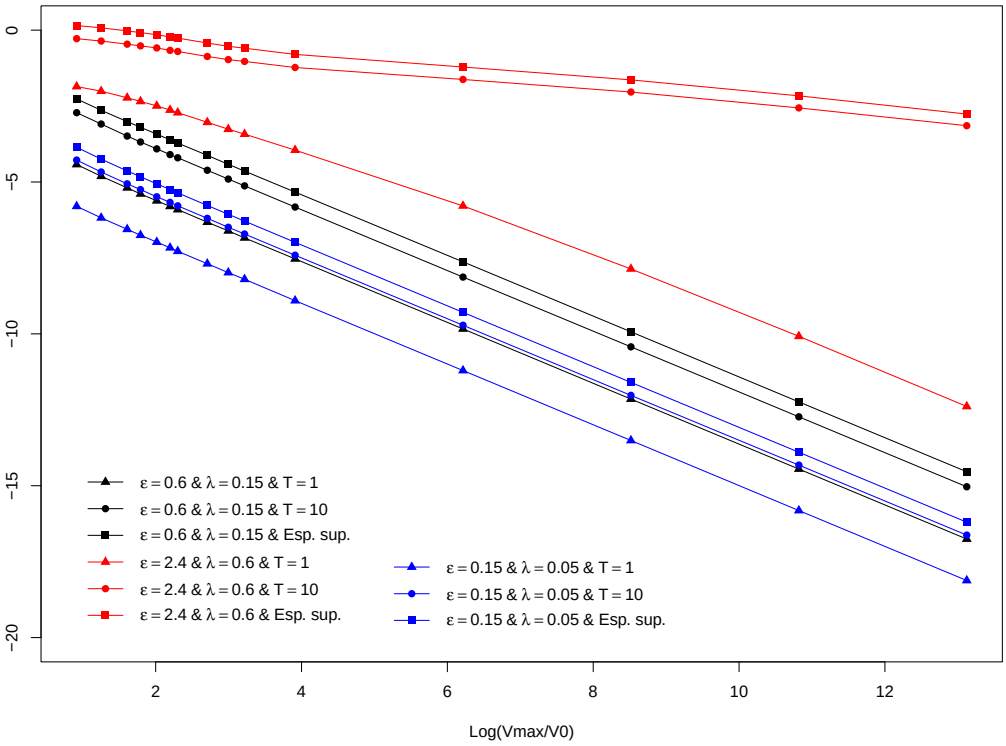
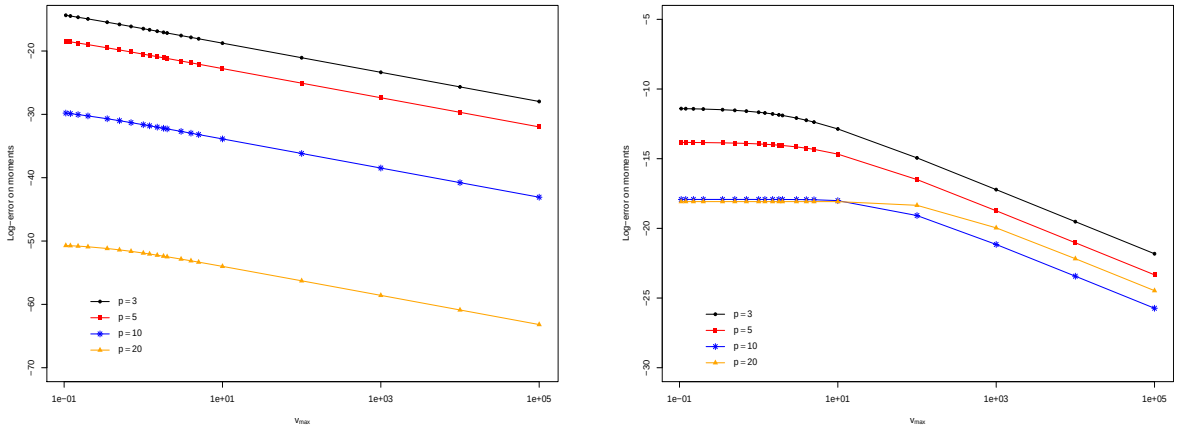


Figure 4.9: $\log \mathbb{E}^S\big[\sup_{t\leq 10} |S_t^{m,n,J(p)} - S_t^{m,n}|\big]$ and $\log \mathbb{E}^S\big[|S_T^{m,n,J(p)} - S_T^{m,n}|\big]$, $T = 1, 10$, as functions of $\log(v_{max}^{(p)}/v_0)$ for different parametrizations. 10^5 paths for Monte-Carlo estimations.

4.5.5.1 Convergence of the expansion series

We illustrate the pricing of swaptions with the Gram-Charlier expansion under model (4.7). We recall that a swaption is a call option on the swap rate: its discounted payoff writes $\varphi(x) = B^S(0) \times (x - K)_+$ where K denotes the strike of the swaption (recall that B^S is annuity of the swap appearing in the denominator of Equation (A.21)). Prices obtained with Gram-Charlier expansions are compared to reference prices that are computed thanks to Monte-Carlo simulations. Except for Figure 4.13b in which a narrow confidence interval was required, Monte-Carlo prices were computed using 10^5 paths with a time-step being now set to 0.001. Monte-Carlo prices will be considered as reference (or target) quantities. In the following, parameters of the model (4.7) change from plot to plot.

The choice of auxiliary density g_r used to build the space $\mathcal{L}_r^2$ is central when pricing using Gram-Charlier expansion. As discussed in [FMS13], [AFP17] or [AF18], the parameters of the auxiliary density should be chosen so that a maximum number of moments of this auxiliary density match those of the unknown one. Namely, setting $\mu_r = \mathbb{E}^S[S_T^{m,n}]$ and $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$ allows to match the first two moments of the unknown density and by doing so, the convergence rate (with respect to the truncation order of the series) of the Gram-Charlier series could be significantly improved. However, such a choice is realistic due to the number of constraints imposed to the model parameters, and notably on the variance parameter since it is generally conflicting with condition (4.11) that guarantees the convergence of the Gram-Charlier approximation. This point is also discussed in [AFP17]. Following the work of [AF18], g_r can be alternatively chosen as a Gaussian mixture (i.e. a linear combination of Gaussian densities).



(a) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta) = (0.06, 0.03, 0.08, 0.02, 0.70, 0.03, 0.16, 0.70, 0.10)$. Linear regressions gave -1.008 (resp. -1.012, -1.025 and -1.068) as slope for $p = 3$ (resp. $p = 5$, $p = 10$ and $p = 20$).

(b) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta) = (0.06, 0.03, 0.08, 0.02, 0.70, 0.03, 2.40, 0.70, 0.10)$. Linear regressions gave -1.233 (resp. -1.34, -1.621 and -1.995) as slope for $p = 3$ (resp. $p = 5$, $p = 10$ and $p = 20$).

Figure 4.10: $\log \left| \mathbb{E}^S \left[\left(S_1^{m,n,J} \right)^p \right] - \mathbb{E}^S \left[\left(S_1^{m,n} \right)^p \right] \right|$ as a function of v_{max} for different p. Moments are computed thanks to matrix representation.

The value of σ_r^2 is indicated for each experiment and whether it matches $\mathrm{Var}^S(S_T^{m,n})$ or not. The parameter μ_r is always chosen so that $\mu_r = \mathbb{E}^S[S_T^{m,n}] = S_0^{m,n}$.

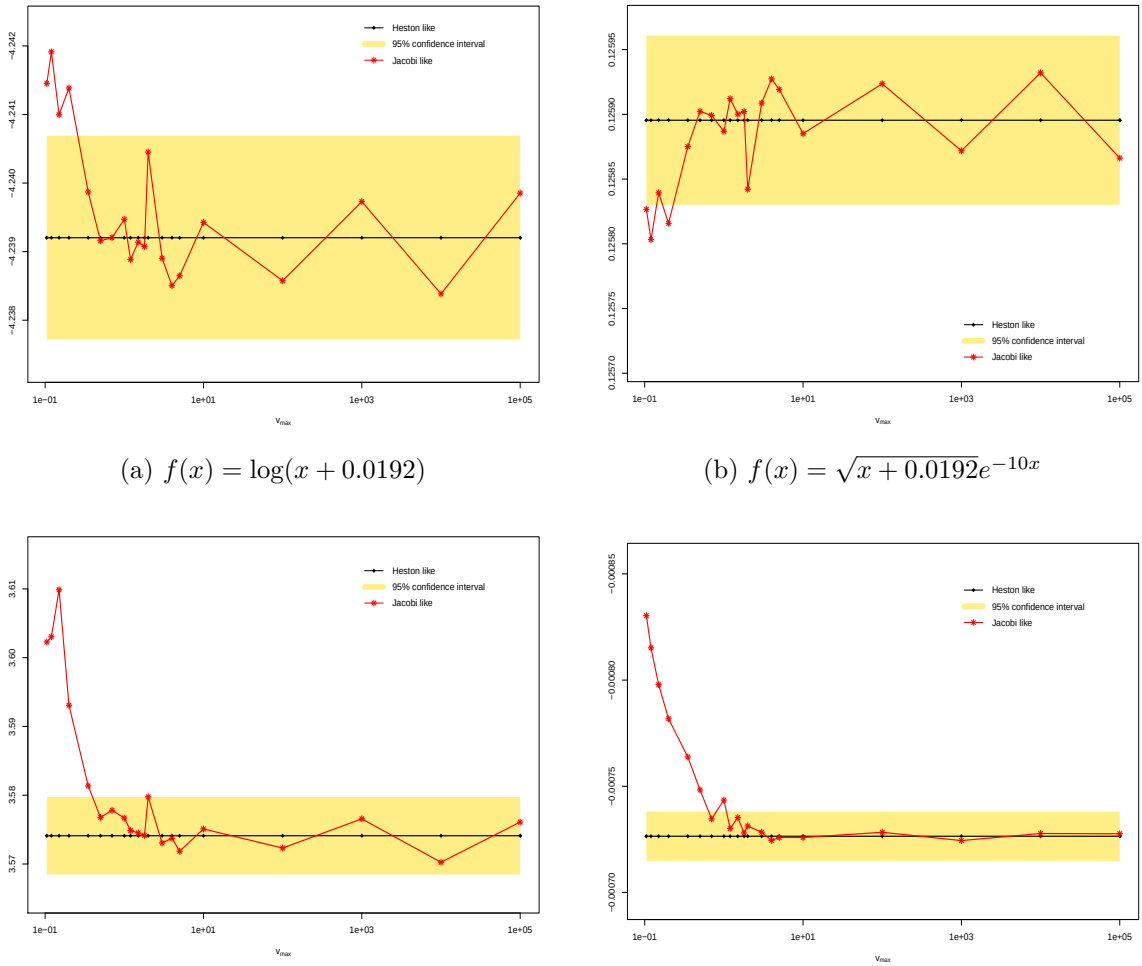
Example 1

We first illustrate that the condition (4.11) is sharp and how the variance of the auxiliary density monitors the behavior of the Gram-Charlier series. In Figure 4.12 we gather several experiments to illustrate it. First, in Figure 4.12a, $v_{max} < \infty$ and we take $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$ but parameters are so that (4.11) is not satisfied. The divergence is clear in the left panel; in the right on (that is a zoom of the left one) we observe some typical oscillations relatively close to the target price (Monte-Carlo) before diverging. In Figure 4.12b we still take $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$ but with $v_{max} = \infty$: we observe again oscillations around the target value before a clear divergence; however, the divergence appears earlier in the expansion in that case. In a third experiment we arbitrarily fix $\sigma_r = 0.0085$; the divergence becomes material now around the 8-th order expansion.

Example 2

We now select finite v_{max} and consider parametrizations so that (4.11) holds. In Figure 4.13a the parameter σ_r is chosen close to the empirical variance but in such a way that the sufficient condition (4.11) is satisfied: we take $\sigma_r = \sqrt{\mathrm{Var}^S(S_T^{m,n})} + 10^{-4}$. The Gram-Charlier expansion is now stable and converge towards the Monte-Carlo price, in line with results of [AFP17]. When moving to a parametrization allowing to match the first two moments of the density of $S_T^{m,n}$, we observe a remarkable accuracy and stability of the Gram-Charlier prices. Indeed in Figure 4.13b it is observed that from the very first truncating orders, Gram-Charlier prices are very close to the reference one as they are in the confidence interval - obtained with 10^6 simulations - of Monte-Carlo price and remain in it as the truncation order increases. The same observation can be made for Figure 4.13d in which we price a relatively deep In-The-Money option of rather long maturity.

On the contrary, the more the parameter σ_r is chosen "far" from $\mathrm{Var}^S(S_T^{m,n})$, the more the



(a) $f(x) = \log(x + 0.0192)$

(b) $f(x) = \sqrt{x + 0.0192}e^{-10x}$

(c) $f(x) = \frac{0.05+x^2}{0.0192+x}$

(d) $f(x) = x\mathbb{1}_{x \geq -0.0045}$

Figure 4.11: $\mathbb{E}^S\left[f(S_1^{m,n,J})\right]$ (in red) and $\mathbb{E}^S\left[f(S_1^{m,n})\right]$ (black lines) for different payoff functions f as a function of v_{max} (logarithm scale on x-axis). Estimations made on 10^5 Monte-Carlo paths; $v_0 = 0.1$.

Gram-Charlier prices converge slowly (and sometimes may diverge due to numerical instabilities). In Figure 4.13c, we set $\sigma_r = \sqrt{\mathrm{Var}^S(S_T^{m,n})} + 0.002$ - corresponding to $\frac{0.002}{\sqrt{\mathrm{Var}^S(S_T^{m,n})}} \approx 0.63$ in that case. The Gram-Charlier prices do not reach the confidence interval by the expansion order $N = 20$.

4.5.5.2 Truncation error analysis

In this paragraph, we analyze the error made when truncating the Gram-Charlier series as a function of the truncating order N. Illustrations of this paragraph are related to results of Section 4.4.2. Below we test several parametrizations to plot the error between successive Gram-Charlier prices and limiting one defined as being Gram-Charlier price of order $N = 50$. The truncation error is depicted for At-The-Money, In-The-Money and Out-of-the-Money swaptions. All the numerical results are in lines with Theorem 4.18. In Figure 4.14 it is remarkable that the parameter whose impact is bigger on the convergence is the shift parameter δ: when δ is large,

all derivatives behave as the ATM ones. More surprisingly, it turns out that the truncation error is not a monotonic function of the truncation order in case of AFM options.

Theorem 4.18 relies on Theorem 4.15 describing the behaviour of Hermite coefficients. For the sake of completeness, we provide in Figure 4.15 the logarithm of the coefficients $(h_n^2)_{n\in\mathbb{N}}$ as a function of n computed under the same parametrization as in Figure 4.14 but enriched with other couples of maturity and tenor. We set $v_{min} = 0$ so that we are in the second case of the statement. In that particular case, the decreasing of the Hermite coefficients is slower than in the first case. It is actually observed in Figure 4.15 that the decreasing of Hermite coefficients is quite slow. Observe also that interestingly, as the order increases, the impact of the maturity and tenor (i.e. the couple (T_m, T_n)) is less pronounced.

4.5.5.3 Benchmark

To assess the impact of the approximation of (4.1) by the Jacobi-based dynamics (4.7), we price swaptions under a fixed set of parameters. We set:

$$(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta) = (0.06, 0.03, 0.08, 0.02, 0.70, 0.03, 0.16, 0.7, 0.1);$$

$v_0 = 0.1$ in both dynamics (4.1) and (4.7); $v_{max} = 9 \times v_0 = 0.9$ for (4.7). Various pricing methods are compared:

- the Heston's standard approach based on integration of the characteristic function of (4.1) as in [WZ06] (integrals are numerically computed using a Laguerre quadrature with 180 points) - referred to as *Heston* (for $v_{max} = \infty$);
- a Gram-Charlier expansion of the density of (4.1) as in [DABB17] but generalized[1] to higher truncation orders - referred to as *GC-CIR* (for $v_{max} = \infty$);
- a Gram-Charlier expansion of the density of (4.7) as introduced in Section 4.3.2.2 for which we consider two alternatives in the way we chose the auxiliary density g_r:
  - as a Gaussian density of mean $\mu_r = S_0^{m,n}$ and variance $\sigma_r = \sqrt{\frac{T v_{max} \lambda_{max}^2}{2}} + 0.0001$ - referred to as *GC-Jacobi* (for $v_{max} < \infty$);
  - as a mixture of Gaussian densities (following [AF18]) whose first two moments are the same of $S_T^{m,n}$ - referred to as *GC-Mixture* (for $v_{max} < \infty$).

We stress that *Heston*, *GC-CIR* and *Monte-Carlo* methods are employed to price under dynamics (4.1) while *GC-Jacobi* and *GC-Mixture* allow to get *approximating prices* associated to approximating dynamics (4.7). We still consider that the *true* swaption prices come from *Monte-Carlo* simulations based on 10^5 paths of (4.1) using the aforementioned discretization schemes with time step 10^{-3}.

Before going into the details of the results, we describe quickly the method *GC-Mixture*. The reference density is chosen as a linear combination of Gaussian densities:

$$g_r(x) = \sum_{m=1}^{M} c_m g_m(x) = \sum_{m=1}^{M} c_m \frac{e^{-(x-\mu_m)^2/2\sigma_m^2}}{\sqrt{2\pi\sigma_m^2}}$$

for a chosen $M \in \mathbb{N}^*$, with $0 \le c_m \le 1$ for all $m \in \{1, \dots, M\}$, $\sum_{m=1}^{M} c_m = 1$ and $(\mu_m, \sigma_m^2) \in (\mathbb{R}, \mathbb{R}_+^*)$ for all $m \in \{1, \dots, M\}$. [AF18] showed that convergence of Gram-Charlier expansion

[1] The generalization is done using the formulas introduced in Section 4.3.2.2 in which we set $v_{max} = \infty$.

(4.18) is guaranteed as long as (4.11) is imposed for at least one σ_m. In our experiments, we have set $M = 2$ and $c_1 = 0.99$ and constraint (4.11) is imposed on σ_2; σ_1 is determined so that $c_1 \times \sigma_1^2 + (1 - c_1) \times \sigma_2^2 = \mathbb{E}^S[(S_T^{m,n})^2]$; we also set $\mu_1 = \mu_2 = S_0^{m,n}$. As argued in [AF18], the reason behind these choices is that we want to approximate the unknown density mainly through g_1 whose variance is expected to be closer to that of the target density which is the case in our experiments.

Table 4.5 we find prices obtained with (4.1): as expected, all Gram-Charlier expansions eventually diverge after observing some oscillations around the reference prices. Those fluctuations are characteristic of the divergence of the expansion series as illustrated in Figure 4.12. The chosen reference density has the same variance as the swap rate, and thus the Gram-Charlier prices for small truncation orders are relatively close to the target one as already shown in [DABB17]. Recall that this behaviour is very dependent on the choice of the auxiliary density as discussed for Figure 4.12.

In Table 4.6 we find the approximating swaption prices obtained under (4.7) and recall the target prices (i.e. *Monte-Carlo* prices) under (4.1). First, we observe that bounding the volatility process does not substantially impact the pricing of swaptions. All approximating prices of the method *GC-Jacobi* do converge to the targets, expressing the robustness of the method. However, the convergence can be quite slow as most of the expansions do not reach the confidence intervals. This is due to the choice of the variance parameter of the reference density: σ_r^2 is chosen so that (4.11) is satisfied which degrades the speed of convergence (as illustrated in Figure 4.13). The Gaussian mixture method *GC-Mixture* allows to get rid of this difficulty by choosing a reference density that shares a number of moments with the unknown density to speed the convergence up. The improvement is significant since from the very first truncation orders, the Gram-Charlier prices are close to the Monte-Carlo ones and that they remain so through the expansion. Interestingly for some swaptions the *GC-Mixture* better replicates the reference price than *Heston* pricing do, e.g. 5×30 OTM swaption of relative strike 100bps. We also observe a remarkable stability of the approximating prices that makes this pricing method very attractive: it is satisfactory to rely on a 4-th order expansion using a Gaussian mixture to obtain an accurate price approximation. However this approach may also suffer from numerical instabilities as already noted in [AF18]. For instance, the ATM swaption 1×30 is badly approximated at high orders by the *GC-Mixture* approach.

4.5.5.4 Computational time discussion

For practical applications in which intensive pricing must be performed (e.g. for calibration), the computational time of the different proposed approach is of interest. Relying on the matrix exponentiation can be cumbersome since it is a costly operation. We discuss it below.

Cost of matrix exponentiation

Whereas [AFP17] worked in a time independent framework, our study allows for time dependency of coefficients driving the evolution of our processes of interest. We have detailed our piecewise constant parametrization in Section 4.5.2. The computation of moments of swap rate process is based on the matrix exponentiation common to all polynomial processes - see Section 3.1.9, notably Theorem 3.33. In present piecewise constant time dependency, this computation is easier as discussed in Remark 21, Equation 3.35. Yet, computation of matrix exponential is numerically costly, as discussed above.

In the expression of Gram-Charlier prices (4.18), the Hermite moments $h_p = \mathbb{E}^S[H_p(S_T^{m,n})]$ are linear combinations of marginal moments of swap rate process at terminal date T. Namely, for a given $p \in \mathbb{N}$, h_p requires the computation of moments of $S_T^{m,n}$ up to p-th order. In

standard parametrization, we consider that time dependent functions are annually piecewise constant (see Section 4.5.2). We now split the successive intervals $[T_i, T_{i+1}]$ (of length one in our parametrization) into $L \in \mathbb{N}$ sub-intervals of length $1/L$. Coefficients are set piecewise constant according to this new grid. In Figure 4.16, we provide the time computation of the p-tuple $U_p := \left(\mathbb{E}^S[(S_T^{m,n})], \ldots, \mathbb{E}^S[(S_T^{m,n})^p]\right)$ as a function of (p, L) for $T = 2$ (meaning that we split both $[0, 1]$ and $[1, 2]$ into L sub-intervals). We took $(p, L) \in \{3, 5, 10, 15, 20, 30, 40\} \times \{1, 3, 5, 10, 20\}$; computations were performed using the `expm` function of R language. We observe a quasi exponential growth of the computational time (recall that Figure 4.16 is displayed with log-scale) meaning that repeated computations of Hermite moments using matrix exponentiation of $S^{m,n}$ is not doable in a reasonable time even for relatively small orders. Indeed, for $L = 10$ computing a single 30-th order Gram-Charlier price takes around 1sec which is quite slow; a similar computational cost would be required for a 30-th order price with $L = 1$ and maturity $T = 20$.

Cost of various pricing approaches

We now compare the cost of various pricing methods introduced in the previous section:

- *Heston*, based on a Laguerre's quadrature;
- a 4-th order expansion in the Jacobi-based dynamics with moments computed thanks to matrix exponentiation: it is the *GC-4* method;
- a 20-th order expansion in the *GC-20* method;
- a 4-th order expansion in the Jacobi-based dynamics with moments represented by integrals: it is *GC-Integral* method.

Indeed, to overcome the computational cost associated to the matrix exponentiation, we propose to rely on an alternative moments representation for the third method. We propose to compute them based on alternative closed-form formulas involving some integrals (see Chapter 5, Section 5.4 for more details) which significantly reduces the computational cost. In practice, this can be done for computation of relatively small order moments but can hardly be generalized to high orders. The following Table 4.2 gives the time for a call to the pricing function associated to each method: those are average durations obtained on 1000 calls (100 for *GC-Matrix*). As the computational time depends mainly on the maturity of the swaptions, we do not detail its dependency regarding tenor and strike.

As expected from the durations displayed in Figure 4.16, the *GC-Matrix* pricing method for a relatively high truncation order is quite slow, and much slower than the *GC-Integral* approach. Note also that only the latter allows to reduce the call compared to that associated with the standard *Heston*-like pricing approach for all maturities. Efficient calibration methods should thus be based on integral representation of moments.

Average call time (in sec., $\times 10^{-2}$)				
Maturity	*GC-20*	*GC-4*	*Heston*	*GC-Integral*
1	0.13	0.056	0.050	0.019
5	0.64	0.22	0.16	0.043
10	1.37	0.43	0.22	0.056
20	2.24	0.91	0.42	0.11

Table 4.2: Average time of a call to the pricing function.

4.5.5.5 Calibration experiments

We propose some calibrations experiments to compare the results obtain when using the Jacobi-based dynamics (4.1) or the reference one (4.1). For the latter, we will still use Fourier-based pricing methods. It will be compared to two different settings to calibrate the proposed dynamics (4.1) using Gram-Charlier expansion:

- g_r is a Gaussian density whose variance parameter is chosen as $\sigma_r = 1.5\sqrt{\mathrm{Var}(S_T^{m,n})}$ –we will refer to this setting as the "Adjusted GC" approach;
- g_r is a Gaussian density whose variance parameter is set to $\sigma_r = \sqrt{\frac{1}{2}Tv_{max}\lambda_{max}^2} + 10^{-5}$ –we will refer to this setting as the "Converging GC" approach;

In any case, we take $\mu_r = S_0^{m,n} = \mathbb{E}^S[S_T^{m,n}]$ and we set $v_{max} = 0.9 = 9 \times v_0$ and still $v_{min} = 0$. Note that the coefficients 1.5 in the first method is chosen based on empirical observations of the ratio $\mathrm{Var}(S_T^{m,n})/\sqrt{\frac{v_{max}T\lambda_{max}^2}{2}}$. The reference approach we will compare those methods to is the well-known Fourier based approach (see formulas in Proposition 1.9) associated with the dynamics (4.1)–this is what we will call the «Heston» approach. In the latter, integrals were computed using quadratures with 180 points. The results below are obtained thanks to Nelder-Mead optimisation algorithm (3 sequences of 500 maximum iterations are performed in a row –see Chapter 5 for more details). The optimisation consists in the minimization of the squared relative distance between market and model volatilities[2]. Those are extracted from Euro market on the December 30th, 2020. They are associated with swaptions of maturities and tenors in $\{1, 2, \ldots, 5, 7, 10, 15, 20, 25, 30\}$ years for At-The-Money data; regarding Away-From-The-Money options residual strikes ranged in $\pm\{0, 25, 50, 100, 150, 200\}$ bps, maturities as previously in $\{1, 2, \ldots, 5, 7, 10, 15, 20, 25, 30\}$, except for the relative strike of -200 bps for which maturities $1, 2, 3, 20, 25, 30$ are missing; for those away-from-the-money options, tenor is fixed to 10 years. In total, 269 market volatilities are to be replicated. The reported times account for the 3 sequences of Nelder-Mead runs.

Not surprinsingly, the Heston approach accounts for the better replication of market data. However, Gram-Charlier methods provide very satisfactory data replication in less than the half of the Heston time.

Calibrating the upper bound parameter To conclude this chapter, we propose to use (4.7) as a model on its own and calibrate v_{max}. We use a 4-th order Gram-Charlier expansion using a reference density of standard deviation $\sigma_r = \sqrt{\frac{\lambda_{max}^2 T_m v_{max}}{2}} + 10^{-5}$ (which thus depends on the parameter of the model) for a derivative of maturity T_m. Calibrations are performed on December 31st of last 4 years. The replication error in the last line is computed as the normalized sum of squared relative differences between market and model volatilities. Results are given in Table 4.4. Note that $v_0 = 0.1$ in all calibration experiments. It is observed that this additional parameter does not perturbed the calibration process and provide satisfactory results in term of data replication. They are better than those presented in Table 4.3, which was expected since the present setting is more parametrized. It is worth noting the v_{max} takes quite small values around 0.45.

[2] If $(\sigma_i^{\mathrm{Mkt}})_{i=1,\ldots,N}$ are market swaptions prices and $(\sigma_i^{\mathrm{Model}})_{i=1,\ldots,N}$ are models ones depending on the set of parameters $\boldsymbol{\Theta}$, the target function is $f(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{i=1}^N \left(\frac{\sigma_i^{\mathrm{Mod}} - \sigma_i^{\mathrm{Mkt}}}{\sigma_i^{\mathrm{Mkt}}}\right)^2$.

Calibrating the DDSVLMM			
Method	Heston	Adjusted GC	Converging GC
a	$2.86 \cdot 10^{-5}$	0.03621	0.04767
b	0.02392	0.003590	0.01013
c	0.1314	0.07334	0.06824
d	0.01538	0.04724	0.02405
κ	0.4331	0.4694	0.2385
θ	0.2030	0.3104	0.2021
ϵ	0.4129	0.4672	0.2480
ρ	0.3405	0.4615	0.4442
δ	0.1724	0.1417	0.1212
Target fun.	0.01350	0.03458	0.08080
Time (in sec.)	1065.0	426.6	570.6

Table 4.3: Proposed methodologies for calibration compared to standard one.

Calibrating the upper bound on volatility				
Date	2020	2019	2018	2017
a	0.03866	0.04299	0.09908	0.05083
b	0.01860	0.01400	0.03320	0.01783
c	0.08890	0.09010	0.11420	0.10540
d	0.004850	0.02876	0.00968	0.02942
κ	0.3622	0.27649	0.32454	0.21358
θ	0.2692	0.16236	0.13561	0.10788
ϵ	0.2982	0.22900	0.19022	0.17104
ρ	0.5117	0.46931	0.53643	0.48931
δ	0.1264	0.14806	0.10583	0.16063
v_{max}	0.4494	0.41838	0.51168	0.51804
Target fun.	0.03744	0.03163	0.02670	0.02948

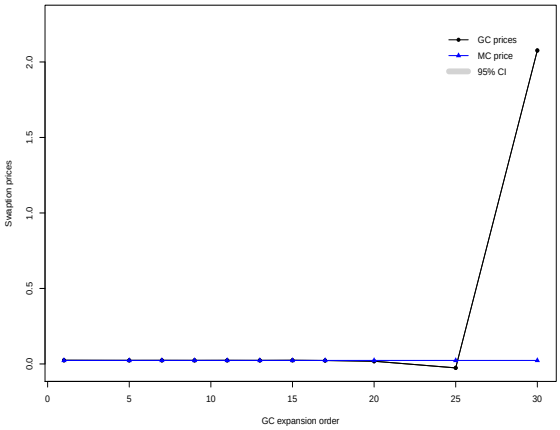
Table 4.4: Calibrating v_{max} among other parameters.

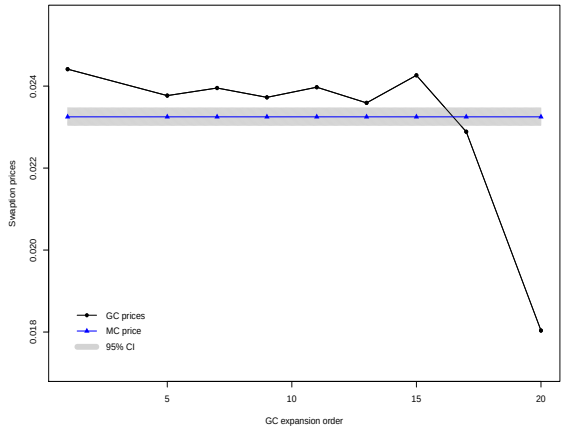
Numerical experiments associated with benchmark analysis of Paragraph 4.5.5.3 In the following tables: (i) maturities ("Mat.") and tenors ("Ten.") are expressed in years; (ii) for OTM derivatives, the column "Strike" actually refers to residual strikes: for a swaption of strike K whose underlying is a swap rate starting at $S_0^{m,n}$, the residual strike is defined as $k := K - S_0^{m,n}$; (iii) in the column containing Monte-Carlo prices, the numbers in brackets relate to the half-width of the 95% confidence interval (expressed in percent); (iv) for the columns associated to Gram-Charlier expansions, the sub-columns N indicate the truncation order in the Gram-Charlier price expansion (4.19); (v) for each swaption, initial values of swap rate $S_0^{m,n}$ and of annuity $B^S(0)$ –defined in (A.21)– are extracted from the Euro market on June 30th, 2020.

ATM swaptions prices (in %)									
Mat.	Ten.	Monte-Carlo	*Heston*	*GC - CIR*					
				$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 20$	$N = 30$
1	5	0.7002 (0.006940)	0.6906	0.6965	0.7034	0.6992	0.7044	0.6644	4.4864
	10	1.6405 (0.01600)	1.6310	1.6328	1.6421	1.6366	1.6422	1.6213	2.5517
	30	4.6268 (0.04452)	4.6367	4.6543	4.6707	4.6607	4.6694	4.6499	5.1194
5	5	1.4564 (0.01502)	1.5446	1.5503	1.5723	1.5477	1.5847	-0.9482	2970.52
	10	3.1950 (0.03217)	3.3075	3.3064	3.3531	3.3007	3.3794	-2.0222	6335.111
	30	8.0585 (0.07943)	7.8981	7.9933	8.0823	7.9816	8.1231	0.9332	8078.586
10	5	2.0861 (0.02072)	2.1998	2.1896	2.2083	2.1896	2.2142	1.2091	822.3816
	10	4.2375 (0.04185)	4.3843	4.3934	4.4121	4.3980	4.4123	4.2713	34.5032
	30	9.7160 (0.09498)	9.5697	9.5292	9.6056	9.5277	9.6293	5.2439	3975.008
20	5	2.9465 (0.02856)	2.9815	2.9734	2.9864	2.9767	2.9866	2.8880	23.9651
	10	5.7149 (0.05481)	5.6405	5.6257	5.6496	5.6317	5.6499	5.4694	44.1814
	30	11.1824 (0.1076)	10.7078	10.6129	10.6979	10.6113	10.7244	5.8403	4427.085

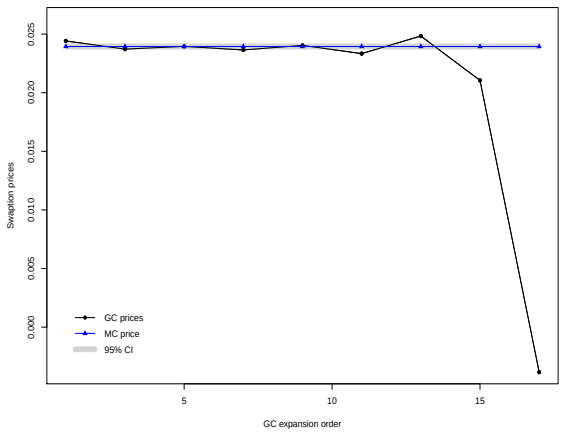
OTM swaptions prices (in %)									
(Mat., Ten.)	Strike (in bps)	Monte-Carlo	*Heston*	*GC - CIR*					
				$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 20$	$N = 30$
$(5, 30)$	-100	28.65237 (0.12036)	28.82812	28.6733	28.6471	28.7101	28.6098	25.2996	4478.3490
	-50	16.72938 (0.10512)	16.67944	16.7130	16.7877	16.7131	16.7571	22.7759	-7.6910
	50	3.20459 (0.05151)	3.26834	3.1732	3.1179	3.2228	3.0760	11.1263	-4723.8150
	100	1.06831 (0.02966)	1.18024	1.1110	1.0272	1.0868	1.0630	-5.0592	713.33970
$(10, 5)$	-100	5.50335 (0.02930)	5.58216	5.5795	5.5871	5.5836	5.5759	6.4377	-185.7491
	-50	3.55659 (0.02560)	3.66747	3.6595	3.6809	3.6572	3.6810	3.5445	-351.0481
	50	1.11335 (0.01559)	1.20400	1.1998	1.1985	1.2080	1.1954	2.0900	-844.7266
	100	0.54746 (0.01105)	0.60535	0.6112	0.5930	0.6172	0.5895	0.8453	435.6163
$(10, 10)$	-100	11.00521 (0.05915)	11.12600	11.1472	11.1552	11.1506	11.1469	11.2711	2.6359
	-50	7.16137 (0.05161)	7.30679	7.3328	7.3546	7.3344	7.3496	7.3161	-4.4963
	50	2.28681 (0.03158)	2.40144	2.3921	2.3905	2.4000	2.3914	2.5263	-29.8883
	100	1.13646 (0.02241)	1.20324	1.1917	1.1730	1.1931	1.1763	1.2102	18.7957
$(10, 30)$	-100	27.54480 (0.13794)	27.52106	27.5144	27.5244	27.5402	27.4778	29.5721	1734.475
	-50	17.35986 (0.11982)	17.26971	17.2397	17.3222	17.2344	17.3129	18.2796	-2999.649
	50	4.81972 (0.06864)	4.66882	4.6597	4.6377	4.6953	4.6194	9.2027	-4100.454
	100	2.14114 (0.04586)	2.02785	2.0630	1.9797	2.0712	1.9799	0.41024	3399.006
$(20, 10)$	-100	11.27873 (0.07188)	11.21118	11.2028	11.2253	11.2055	11.2159	11.3135	-26.4230
	-50	8.21365 (0.06387)	8.13932	8.1257	8.1554	8.1283	8.1515	8.0025	22.8451
	50	3.79619 (0.04542)	3.72652	3.7168	3.7235	3.7265	3.7246	3.8034	-21.5172
	100	2.40971 (0.03641)	2.34795	2.3481	2.3352	2.3566	2.3369	2.5359	-27.1140

Table 4.5: Swaption prices using dynamics (4.1) ($v_{max} = \infty$).

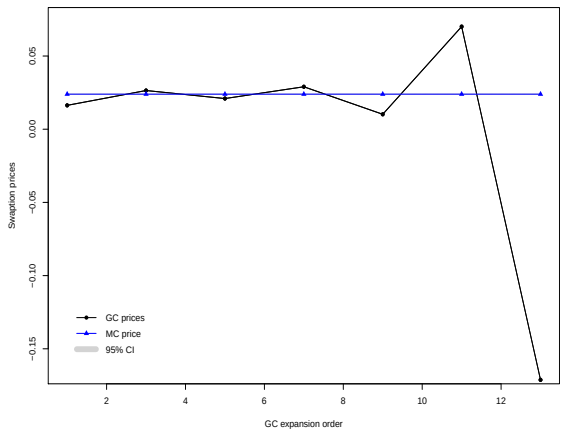




(a) $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$, $s_0^{m,n} = K = 0$, $v_0 = 0.025$, $v_{max} = 0.5$, $a = b = c = 0$, $d = \kappa = 1$, $\theta = 0.25$, $\epsilon = 0.6$, $\rho = \delta = 0.0$ $\sigma_r = 1.27305 \cdot 10^{-2}$. Maturity and tenor of the swaption: $T = T_m = 5$ and $T_n - T_m = 5$. Number of simulations: 10^5.

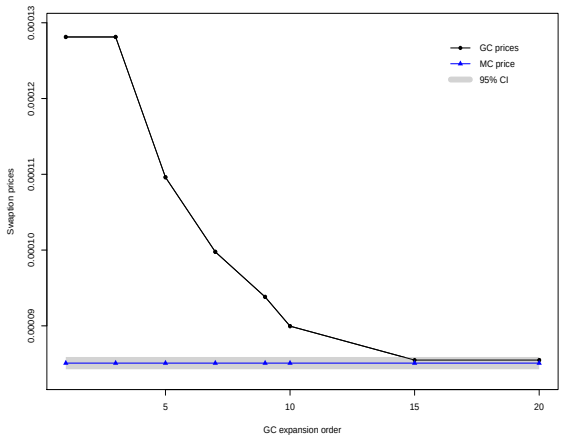


(b) $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$, $s_0^{m,n} = K = 0$, $v_0 = 0.025$, $v_{max} = \infty$, $a = b = c = 0$, $d = \kappa = 1$, $\theta = 0.25$, $\epsilon = 0.6$, $\rho = \delta = 0.0$ $\sigma_r = 1.27305 \cdot 10^{-2}$. Maturity and tenor of the swaption: $T = T_m = 5$ and $T_n - T_m = 5$. Number of simulations: 10^5.

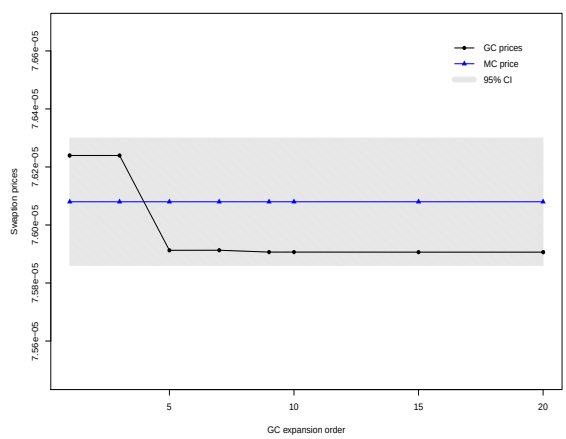


(c) $\sigma_r^2 \neq \mathrm{Var}^S(S_T^{m,n})$, $s_0^{m,n} = K = 0$, $v_0 = 0.025$, $v_{max} = \infty$, $a = b = c = 0$, $d = \kappa = 1$, $\theta = 0.25$, $\epsilon = 0.6$, $\rho = \delta = 0.0$ $\sigma_r = 8.5 \cdot 10^{-3}$. Maturity and tenor of the swaption: $T = T_m = 5$ and $T_n - T_m = 5$. Number of simulations: 10^5.

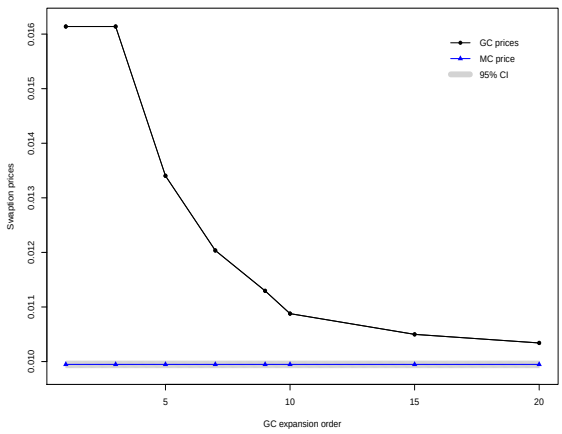
Figure 4.12: Diverging Gram-Charlier prices series.



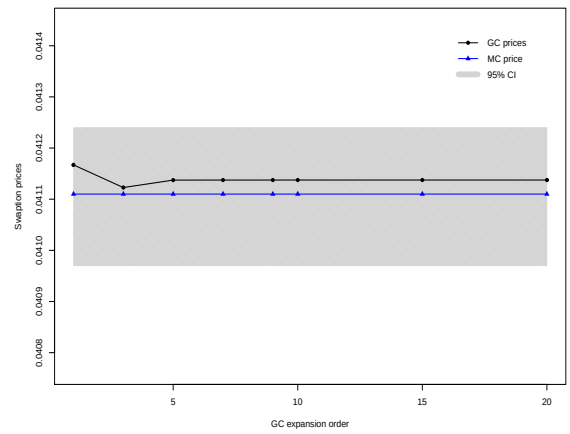
(a) $\sigma_r^2 \neq \mathrm{Var}^S(S_T^{m,n})$, $s_0^{m,n} = 0$, $v_0 = 0.025$, $v_{max} = 0.25$, $a = b = c = 0$, $d = \kappa = 1$, $\theta = 0.12$, $\epsilon = 0.4$, $\rho = \delta = 0.0$ $\sigma_r = 3.203183 \cdot 10^{-4}$, $K = 0$. Maturity and tenor of the swaption: $T = T_m = 1$ and $T_n - T_m = 1$. Number of simulations: 10^5.



(b) $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$, $s_0^{m,n} = 0$, $v_0 = 0.025$, $v_{max} = 0.089$, $a = b = c = 0$, $d = 1$, $\kappa = 2$, $\theta = 0.06$, $\epsilon = 0.1$, $\rho = \delta = 0.0$, $\sigma_r = 1.906037 \cdot 10^{-4}$, $K = 0$. Maturity and tenor of the swaption: $T = T_m = 1$ and $T_n - T_m = 1$. Number of simulations: 10^6.

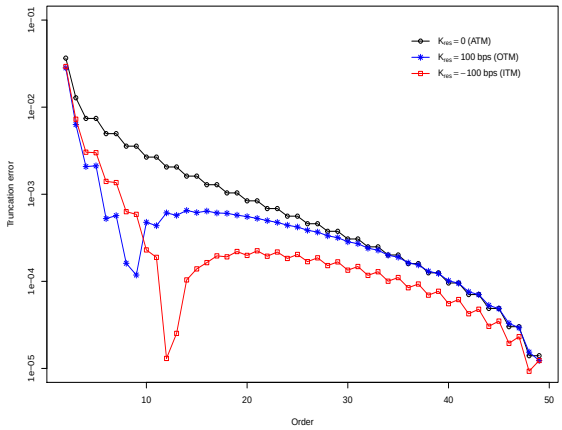


(c) $\sigma_r^2 \neq \mathrm{Var}^S(S_T^{m,n})$. $s_0^{m,n} = 0.008982445$, $v_0 = 0.025$, $v_{max} = 0.089$, $a = 10^{-4}$, $b = 10^{-1}$, $c = 2.5$, $d = 10^{-1}$, $\kappa = 1.5$, $\theta = 0.06$, $\epsilon = 0.13$, $\rho = 0.4$, and $\delta = 0.1$. $\sigma_r = 0.005183955$. $K = s_0^{m,n}$. Maturity and tenor of the swaption: $T = T_m = 2$ and $T_n - T_m = 8$. Number of simulations: 10^5.

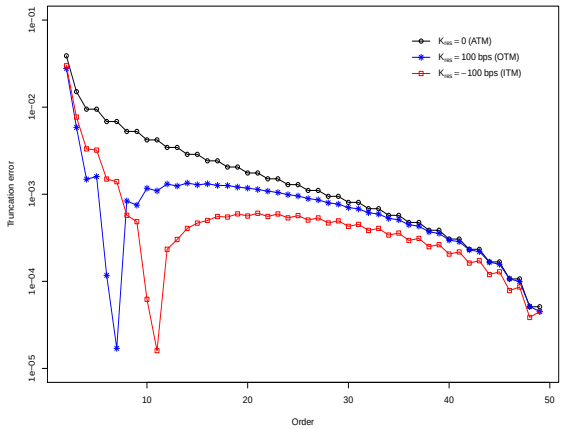


(d) $\sigma_r^2 = \mathrm{Var}^S(S_T^{m,n})$. $s_0^{m,n} = 0.002307454$, $v_0 = 0.015$, $v_{max} = 0.05$, $a = 10^{-4}$, $b = 10^{-1}$, $c = 2.5$, $d = 10^{-1}$, $\kappa = 1.5$, $\theta = 0.03$, $\epsilon = 0.13$, $\rho = 0.4$, and $\delta = 0.1$. $\sigma_r = 0.005461044$. $K = s_0^{m,n} - 0.01$. Maturity and tenor of the swaption: $T = T_m = 10$ and $T_n - T_m = 4$. Number of simulations: 10^5.

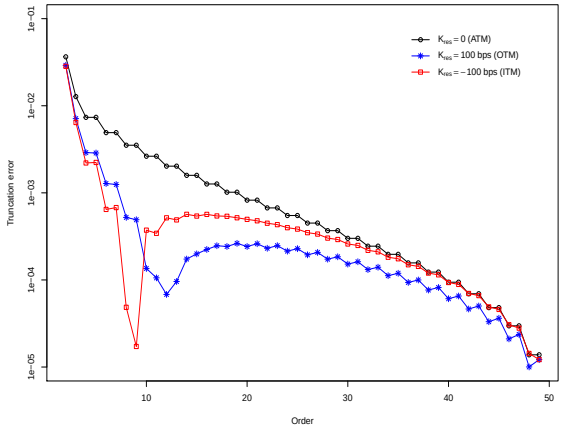
Figure 4.13: Converging Gram-Charlier pricing series.



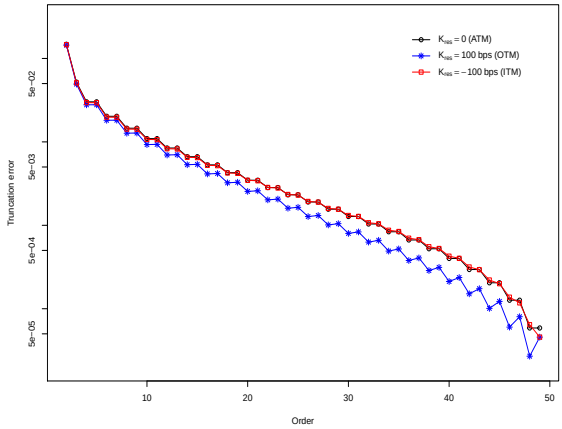
(a) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, 0.7, 0.1, 0.9)$



(b) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.9, 0.7, 0.1, 0.9)$

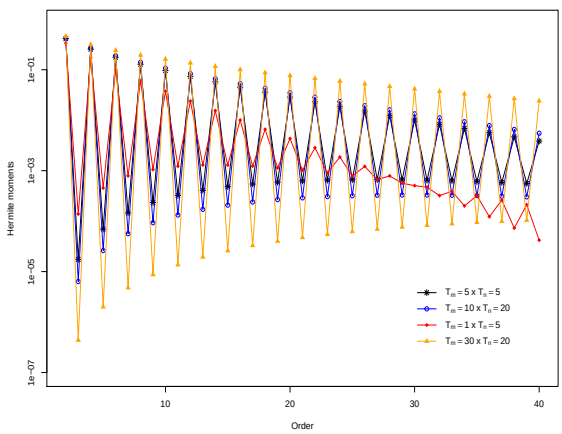


(c) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, -0.5, 0.1, 0.9)$

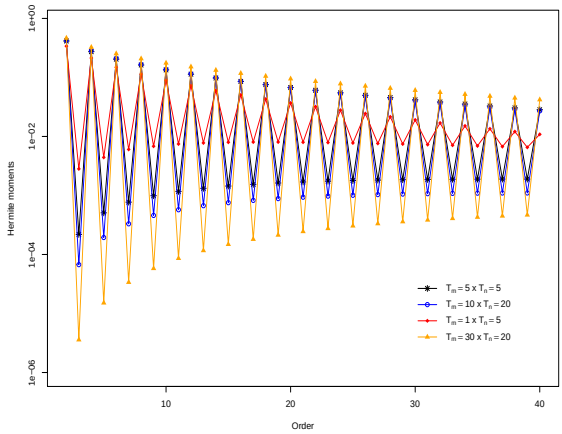


(d) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, 0.7, 0.4, 0.9)$

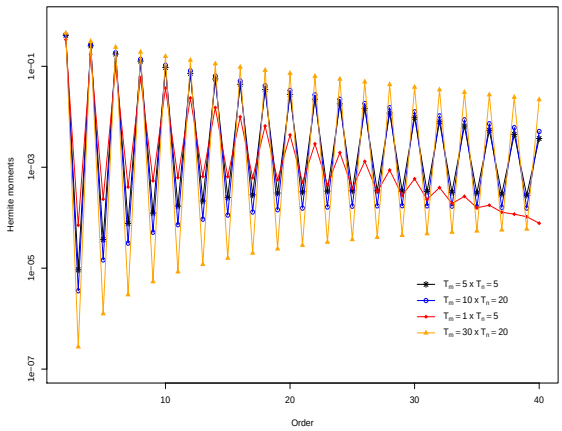
Figure 4.14: $|P_N(\varphi) - P_{50}(\varphi)|$ as a function of N for different parametrizations. $(T_m \times T_n) = (1 \times 5)$; $s_0^{m,n} = -0.004372$. Log-scale on y-axis.



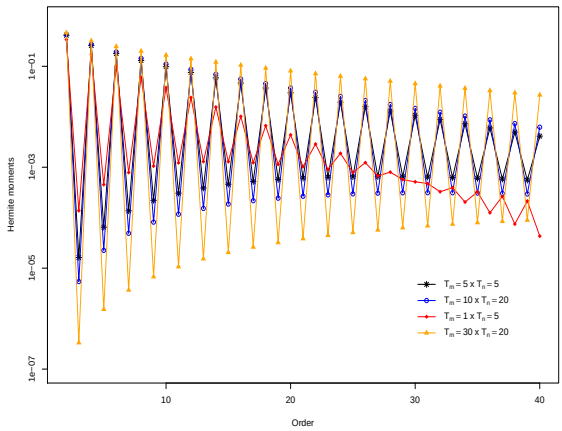
(a) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, 0.7, 0.1, 0.9)$



(b) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.9, 0.7, 0.1, 0.9)$



(c) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, -0.5, 0.1, 0.9)$



(d) $(a, b, c, d, \kappa, \theta, \epsilon, \rho, \delta, v_{max}) =$
$(0.06, 0.03, 0.08, 0.02, 0.7, 0.03, 0.16, 0.7, 0.4, 0.9)$

Figure 4.15: Hermite coefficients h_n^2 as function of n. Log-scale on y-axis.

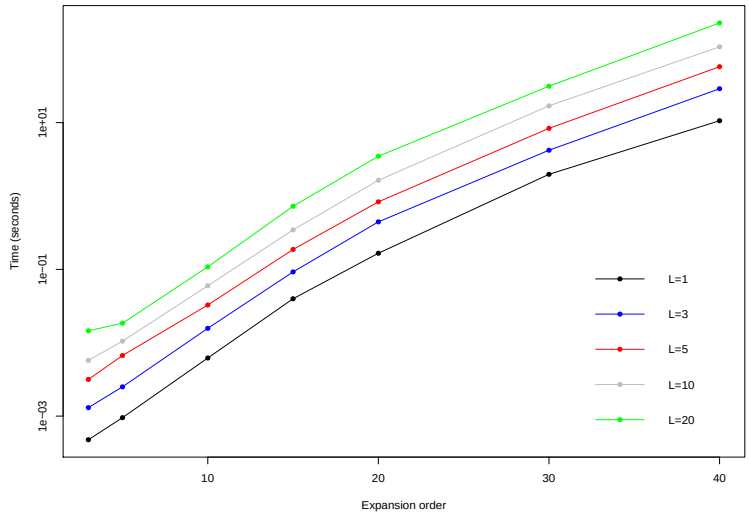


Figure 4.16: Computational time (in seconds) of U_p as a function of (p, L).

ATM swaptions prices (in %)

Mat.	Ten.	Monte-Carlo	*GC-Jacobi*						*GC-Mixture*					
			$N=4$	$N=6$	$N=8$	$N=10$	$N=20$	$N=30$	$N=2$	$N=4$	$N=6$	$N=8$	$N=10$	$N=30$
1	5	0.7002 (0.006940)	0.9189	0.8387	0.7950	0.7682	0.7189	0.7076	0.7011	0.6994	0.7053	0.7012	0.7051	0.70172
	10	1.6405 (0.01600)	1.9777	1.8397	1.7681	1.7264	1.6577	1.6453	1.6381	1.6382	1.6460	1.6402	1.6445	1.63469
	30	4.6268 (0.04452)	5.6326	5.2391	5.0348	4.9158	4.7192	4.6831	4.6612	4.6679	4.6813	4.6697	4.6761	2.08387
5	5	1.45643 (0.01502)	2.3170	2.0713	1.9313	1.8418	1.6587	1.6054	1.5456	1.5572	1.5825	1.5519	1.5925	1.6288
	10	3.19498 (0.03217)	4.83331	4.3358	4.0540	3.8750	3.5141	3.4117	3.3047	3.3249	3.3693	3.3160	3.3823	3.6979
	30	8.05853 (0.07943)	11.3839	10.2535	9.6179	9.2169	8.4216	8.2029	7.9995	8.0243	8.1235	8.0035	8.1512	8.8896
10	5	2.08612 (0.02072)	3.4591	3.0660	2.8388	2.6917	2.3815	2.2855	2.1524	2.1985	2.2240	2.1931	2.2293	2.5756
	10	4.23755 (0.04185)	6.5617	5.7564	5.2837	4.9729	4.2939	4.0687	4.1835	4.3736	4.4388	4.3600	4.4600	5.1794
	30	9.71601 (0.09498)	14.9916	13.2946	12.3148	11.6806	10.3464	9.9348	9.3687	9.5661	9.6718	9.5419	9.6937	10.4277
20	5	2.94648 (0.02858)	5.0916	4.4667	4.0999	3.8587	3.3319	3.1572	2.8629	2.9874	3.0098	2.9831	3.0059	2.9876
	10	5.71492 (0.05481)	9.7187	8.5178	7.8119	7.3470	6.3283	5.9880	5.4073	5.6523	5.6949	5.6437	5.6875	6.5303
	30	11.18240 (0.10759)	19.2066	16.7650	15.3213	14.3649	12.2382	11.5069	10.7889	10.7614	10.7974	10.7390	10.7734	12.4307

OTM swaptions prices (in %)

(Mat., Ten.)	Strike (in bps)	Monte-Carlo	*GC-Jacobi*						*GC-Mixture*					
			$N=4$	$N=6$	$N=8$	$N=10$	$N=20$	$N=30$	$N=2$	$N=4$	$N=6$	$N=8$	$N=10$	$N=30$
$(5,30)$	-100	28.65237 (0.12036)	29.5814	28.9283	28.6504	28.5263	28.4836	28.5619	28.8411	28.6988	28.6563	28.7271	28.6240	29.7973
	-50	16.72938 (0.10512)	19.4613	18.4547	17.9122	17.5834	16.9851	16.8441	16.9134	16.7409	16.8244	16.7458	16.7795	17.8866
	50	3.20459 (0.05151)	5.4611	4.4787	3.9622	3.6597	3.1763	3.1142	2.8944	3.2096	3.1175	3.2348	3.0642	3.4971
	100	1.06831 (0.02966)	1.5783	0.9677	0.7334	0.6509	0.7612	0.9125	0.8032	1.1145	1.0095	1.0583	1.0574	1.0807
$(10,5)$	-100	5.50335 (0.02930)	6.4633	6.1289	5.9478	5.8380	5.6413	5.5993	5.6147	5.5873	5.5954	5.5942	5.5765	5.4864
	-50	3.55659 (0.02560)	4.8483	4.4696	4.2536	4.1154	3.8312	3.7464	3.6685	3.6654	6973	3.6622	3.6952	4.0310
	50	1.11335 (0.01559)	2.2996	1.9237	1.7108	1.5758	1.3081	1.2371	1.1178	1.2146	1.2037	1.2169	1.1921	1.1202
	100	0.54746 (0.01105)	1.3658	1.0366	0.8615	0.7578	0.5900	0.5704	0.5132	0.6261	0.5896	0.6228	0.5816	0.8945
$(10,10)$	-100	11.00521 (0.05915)	12.5747	11.8972	11.5274	11.3017	10.8966	10.8149	11.1162	11.1286	11.1471	11.1516	11.0984	10.7787
	-50	7.16137 (0.05161)	9.3364	8.5632	8.1163	7.8268	7.2158	7.0255	7.2140	7.2949	7.3749	7.2898	7.3759	7.9791
	50	2.28681 (0.03158)	4.2593	3.4893	3.0460	2.7604	2.1702	1.9991	2.1346	2.4209	2.3938	2.4291	2.3607	2.3798
	100	1.13646 (0.02241)	2.4204	1.7489	1.3860	1.1676	0.7989	0.7479	0.9576	1.2530	1.1637	1.2483	1.1408	1.8347
$(10,30)$	-100	27.54480 (0.13794)	30.6997	29.3516	28.6457	28.2332	27.5701	27.4769	27.6674	27.5508	27.5467	27.5841	27.4702	28.3477
	-50	17.35986 (0.11982)	22.1742	20.5646	19.6529	19.0736	17.9024	17.5649	17.2942	17.2652	17.3859	17.2584	17.3617	18.2517
	50	4.81972 (0.06864)	9.1856	7.5892	6.6925	6.1290	5.0358	4.7631	4.2957	4.7246	4.6526	4.7353	4.5952	5.1662
	100	2.14114 (0.04586)	4.7217	3.3987	2.7205	2.3365	1.8044	1.8067	1.6704	2.1175	1.9624	2.0812	1.9591	2.1803
$(20,10)$	-100	11.27873 (0.07188)	14.7645	13.6335	12.9839	12.5659	11.6982	11.4378	11.1446	11.2191	11.2642	11.2241	11.2388	0.12142175
	-50	8.21365 (0.06387)	12.1150	10.9311	10.2387	9.7849	8.8002	8.4762	7.9853	8.1423	8.2049	8.1382	8.1938	9.0687
	50	3.79619 (0.04542)	7.5772	6.3960	5.7067	5.2563	4.2899	3.9833	3.4456	3.7587	3.7510	3.7496	3.7350	4.5494
	100	2.40971 (0.03641)	5.6889	4.5631	3.9197	3.5083	2.6756	2.4472	2.0652	2.4005	2.3428	2.3851	2.3249	3.0961

Table 4.6: Swaption prices using dynamics (4.7) ($v_{max} < \infty$).

Chapter 5

Fast calibration of the LIBOR Market Model with Stochastic Volatility based on analytical gradient

Present chapter presents augmented version of [AAB+20].

5.1 Introduction

To speed up the calibration procedure of the DDSVLMM, two main strategies can be considered. First, reduce the computational time required for numerical calculation of one (or multiple) swaption price. In [WZ06], pricing under the SVLMM is performed based on both the classical [Hes93] numerical integration method and the famous Fast Fourier Transform (FFT) approach of [CM99], which has become a standard for option valuation for models with known characteristic function, as it is particularly the case for affine diffusion processes. As an alternative to moment generating function calculation, [LDB20] have shown the efficiency of using Edgeworth and Gram-Charlier expansions in the calibration of the DDSVLMM. Further work on this type of strategies that reduce the computational cost of the objective function is provided in a companion paper, see [ABLM20].
Note that a typical number of parameters to be specified during a DDSVLMM calibration is 8 or 9, depending on whether the displacement parameter is included in the calibration process or not, whereas a standard Heston model calibration involves 5 parameters. In such calibration problem, the regularity of the objective function is complex to determine. This latter issue can be tackled using optimization algorithms unconcerned about the regularity of the target function; it has been done in [LDB20] in which authors used the Nelder-Mead method.

In this chapter, we develop a second strategy which consists in decreasing the number of objective function calls required by the optimization algorithm. In this context, the use of gradient-based algorithms is of interest. This can be done by relying on either numerical or analytical gradient computation, the latter being particularly efficient in terms of accuracy and computational cost. A central reference for the present chapter is [CdBRG17], that developed the analytical gradient for fast calibration of the Heston model, based on an alternative formulation of the Heston moment generating function proposed by [dBRFCU10]. Alternatives focusing on «learning» the gradient, without analytical calculation, have been proposed by [LBGO19] based on Artificial Neural Networks. Apart from the Heston model, analytical gradient-based

methods have also been developed by [GV19] for the calibration of the so-called 3/2 model.

By adapting the approach of [CdBRG17], the present chapter provides a numerically stable representation of analytical gradient for the DDSVLMM and uses it in gradient-based optimization routines. Note that the work of [CdBRG17] focused on the Heston model, did not make use of real market data and focused on the Levenberg-Marquardt optimization algorithm. We specify the methodology in the context of DDSVLMM and we demonstrate the efficiency of our approach by inputting the analytical gradient into both the Levenberg-Marquardt algorithm (denoted by LM in the following, see [Mar63]) and the Broyden–Fletcher–Goldfarb–Shanno algorithm (denoted by BFGS, see [BLNZ95]) using real market swaption data.

In some experiments, we will require Feller's condition to be satisfied by using an alternative version of the Levenberg-Marquardt algorithm including constraints; note that although the Feller condition is not targeted in [CdBRG17], in our context this condition is of interest for further uses as it preserves the stability of some numerical discretization schemes. As an example, such Feller condition ensures strong convergence of order 1/2 of discretization schemes as discussed in [Alf15] (Chapter 3) and references therein.

It is worth mentioning that in our experiment, the computation of the objective function based on the moment generating function representation proposed by [AMST07] is in average slightly faster compared to that of [dBRFCU10]. This leads us to consider an 'optimal' optimization routine made of 1) the analytical gradient as an input of the gradient-based method (LM or BFGS), coming from the differentiation of the moment generating function as provided in [dBRFCU10] and 2) objective function calls during the optimization procedure that rely on the numerical evaluation of the moment generating function representation of [AMST07].

Finally, the efficiency of our approach is compared to the following methods: the classical Heston-type pricing method [Hes93] or the Edgeworth expansion of swaptions prices as developed in [LDB20], both combined with the Nelder-Mead algorithm as well as the Levenberg-Marquart algorithm with numerical gradient. As a main result, this chapter shows that the analytical gradient method is highly competitive both from a computational standpoint and calibration accuracy, as it achieves to significantly limit the number of steps in the optimization algorithm while still offering accurate data replication. Our calibration experiments on real data also show that expansion methodologies that reduce the computational cost of an objective function call (see for instance [LDB20] and [ABLM20]) is a complementary efficient alternative; the combination of the two strategies (expansion approach and computation of the related analytical gradient) appears as a promising direction which is left for further research.

Section 5.2 presents the alternative representation of the moment generating function we propose along with the analytical gradient calculation of the objective function it implies; the optimization algorithm used is also detailed. Section 5.3 illustrates the efficiency of the proposed calibration method and compares it to the classical alternative methods listed above. Section 5.4 is dedicated to the study of the calibration of Jacobi based dynamics (4.7) introduced in previous Chapter.

5.2 Analytical gradient calculation and optimization routines

5.2.1 Swaptions pricing

We retake the notations of Chapter 1: $S^{m,n}$ represents the swap rate associated with contract of maturity T_m and tenor T_n. Spot swaption price expresses as a call options on the swap rate

$$PS(0,K) = B^S(T_m)\mathbb{E}^S\left[(S^{m,n}_{T_m} - K)_+\right]$$

K is the strike of the option. When modelling the swap rate using a *frozen* approximation of the standard DDSVLMM defined in (1.48) with piecewise constants parameters, this swaption price can be expressed in terms of integrals in the complex field of the characteristic function of $S_{T_m}^{m,n}$ since it is a Heston-like model. This characteristic function is itself analytically expressed in terms of the parameters $\mathbf{\Theta}$ defining the dynamics (1.48) via Riccati equations as given in Proposition 1.8. To underline the dependency of the swaption prices with respect to these parameters, we will rather denote swaptions prices below $PS(\mathbf{\Theta}; 0, K)$.

Before going any further, we recall the expression of the moment generating function involved in (log-normal) swaption prices of Proposition 1.8 defined over $\mathcal{D} \subset \mathbb{C}$ as

$$\Psi(x; t, S_t^{m,n}, V_t) = \mathbb{E}^S \left[\exp\left(x \ln\left(\frac{S_{T_m}^{m,n} + \delta}{S_0^{m,n} + \delta} \right) \right) \middle| \mathcal{F}_t \right], \ x \in \mathcal{D}.$$

Beforehand, we recall some quantities defined in Chapter 4:

$$\xi(t) := \tilde{\xi}_0^S(t), \quad \lambda(t) := \left\| \sum_{j=m}^{n-1} \omega_j(0) \boldsymbol{\gamma_j}(t) \right\|, \quad \tilde{\rho}(t) := \frac{1}{\lambda(t)} \sum_{j=m}^{n-1} \omega_j(0) \|\boldsymbol{\gamma_j}(t)\| \rho_j(t).$$

Then, for x in the domain of definition of the moment generating function, the moment generating function can be expressed as

$$\Psi(x; t, s, v; \mathbf{\Theta}) = e^{A(\tau,x) + B(\tau,x)v + sx} \tag{5.1}$$

where $\tau = T_m - t$ and the deterministic functions A and B (whose dependency with respect to $\mathbf{\Theta}$ is omitted for the sake of simplicity) are recursively computed on the grid $(\tau_j, \tau_{j+1}]$ where $\tau_j = T_m - T_{m-j}$, $j = 0, \cdots, m-1$: for each $j \in \{0, 1, \ldots, m-1\}$, for each τ in $(\tau_j, \tau_{j+1}]$,

$$\begin{cases} A(\tau, z) &= A(\tau_j, z) + \dfrac{\kappa\theta}{\epsilon^2} \left[(\mu + \nu)(\tau - \tau_j) - 2 \ln \dfrac{1 - g_j e^{\nu(\tau - \tau_j)}}{1 - g_j} \right], \\ B(\tau, z) &= B(\tau_j, z) + \dfrac{(\mu + \nu - \epsilon^2 B(\tau_j, z))(1 - e^{\nu(\tau - \tau_j)})}{\epsilon^2 (1 - g_j e^{\nu(\tau - \tau_j)})}, \end{cases}$$

with initial condition $A(0, z) = B(0, z) = 0$ and

$$\mu = \kappa\xi(\tau_j) - \tilde{\rho}(\tau_j)\epsilon\lambda(\tau_j)z, \quad \nu = \sqrt{\mu^2 - \lambda^2(\tau_j)\epsilon^2(z^2 - z)}, \quad g_j = \frac{\mu + \nu - \epsilon^2 B(\tau_j, z)}{\mu - \nu - \epsilon^2 B(\tau_j, z)}.$$

For the sake of simplicity, we omitted the time dependency of μ, ν and g_j.

5.2.2 Analytic characteristic function gradient

As pointed out by [AMST07], the representation of the characteristic function by [Hes93] suffers from numerical discontinuities for long maturities. To overcome this difficulty, they proposed an equivalent representation that is continuous. Another alternative formulation of the Heston moment generating function has been developed by [dBRFCU10] that is also continuous. This alternative formulation is easier to differentiate and thus well suited for gradient calculation, as derived by [CdBRG17] in the context of the Heston model.

For our purpose, we rely on the following modifications of the functions A and B: for

$\tau_j \leq \tau < \tau_{j+1},\ j = 0, 1, \ldots, m-1,$

$$\begin{cases} A(\tau, z) &= A(\tau_j, z) - \dfrac{\kappa\theta\tilde{\rho}\lambda z(\tau - \tau_j)}{\epsilon} + \dfrac{2\kappa\theta}{\epsilon^2} D_j(\tau), \\ B(\tau, z) &= B(\tau_j, z) - \dfrac{A_j(\tau)}{V_0}, \end{cases} \tag{5.2}$$

where we have set

$$\begin{aligned} E_j(\tau) &= \nu + \mu - \epsilon^2 B(\tau_j, z) + (\nu - \mu + \epsilon^2 B(\tau_j, z)) e^{-\nu(\tau - \tau_j)}, \\ D_j(\tau) &= \ln \frac{\nu}{V_0} + \frac{\kappa\xi - \nu}{2}(\tau - \tau_j) - \ln \frac{E_j(\tau)}{2V_0}, \\ A_j^1(\tau) &= \left[B(\tau_j, z)(2\mu - \epsilon^2 B(\tau_j, z)) + \lambda^2(z - z^2)\right] \tanh \frac{\nu(\tau - \tau_j)}{2}, \\ A_j^2(\tau) &= \frac{\nu}{V_0} + \frac{\mu - \epsilon^2 B(\tau_j, z)}{V_0} \tanh \frac{\nu(\tau - \tau_j)}{2}, \\ A_j(\tau) &= \frac{A_j^1(\tau)}{A_j^2(\tau)}. \end{aligned}$$

These formulas have been adapted from [CdBRG17] by using hyperbolic tangent functions in order to avoid numerical instabilities.

The following result is key for present chapter and is based on an equivalent pricing formula of Proposition 1.8.

Proposition 5.1. *The gradient of the swaption price under the approximate Heston dynamics (1.49) is given by:*

$$\nabla PS(\boldsymbol{\Theta}; 0, K) = B^S(0)\left((S_0^{m,n} + \delta)\nabla P_1(\boldsymbol{\Theta}; 0, K) - (K + \delta)\nabla P_2(\boldsymbol{\Theta}; 0, K)\right)$$

with

$$\nabla P_1(\boldsymbol{\Theta}; 0, K) = \frac{1}{\pi}\int_0^{+\infty} Re\left(\frac{e^{-iu\ln\frac{K+\delta}{S_0^{m,n}+\delta}}\nabla\Psi(u - i; 0, s_0^{m,n}, v_0; \boldsymbol{\Theta})}{iu}\right) du,$$

$$\nabla P_2(\boldsymbol{\Theta}; 0, K) = \frac{1}{\pi}\int_0^{+\infty} Re\left(\frac{e^{-iu\ln\frac{K+\delta}{S_0^{m,n}+\delta}}\nabla\Psi(u; 0, s_0^{m,n}, v_0; \boldsymbol{\Theta})}{iu}\right) du,$$

and

$$\nabla\Psi(\boldsymbol{\Theta}; u) = \Psi(u; 0, s_0^{m,n}, v_0; \boldsymbol{\Theta})\boldsymbol{\chi}(\boldsymbol{\Theta}; u),$$

where Ψ is given by (5.1) and $\boldsymbol{\chi}(\boldsymbol{\Theta}; u)$ is the gradient vector whose components are the partial derivatives of the characteristic function with respect to each parameter. It is given in details in Appendix C.

Remark 35. *The particular form of the analytical gradient allows one to compute $\Psi(\boldsymbol{\Theta}; u)$ only once.*

5.2.3 Formulation of the calibration

The calibration problem corresponds to the finding of model parameters that allow to best replicate market data. In this chapter, we choose to calibrate on market swaptions prices rather than on implied volatilities since we derived the analytical gradient of the swaptions prices.

Let us consider the following standard parametrizations for the maps $\boldsymbol{\gamma_j}$ and ρ_j introduced in (1.48): for $t \in [T_k, T_{k+1})$,

$$\boldsymbol{\gamma_j}(t) \equiv \boldsymbol{\gamma_j}(T_k) = g(T_j - T_k)\boldsymbol{\beta_{j-k+1}},$$
$$\rho_j(t) \equiv \rho_j(T_k) = \frac{\rho}{\sqrt{N_f}} \frac{1}{\|\boldsymbol{\gamma_j}(T_k)\|} \sum_{p=1}^{N_f} \boldsymbol{\gamma_j}^{(p)}(T_k),$$

where $g(u) = (a + bu)e^{-cu} + d$ with a, b, c and d non-negative parameters to calibrate, $\boldsymbol{\beta_{j-k+1}}$ is a N_f-dimensional unit vector representing the inter-forward correlation structure and ρ is a correlation parameter to calibrate in $(-1, 1)$. Given this parametrization, one can compute the derivatives of g (which is the norm of $\boldsymbol{\gamma_j}$ by construction) with respect to a, b, c and d as follows, for $u \geq 0$:

$$\begin{aligned} \frac{\partial g(u)}{\partial a} &= e^{-cu}, & \frac{\partial g(u)}{\partial b} &= u\frac{\partial g(u)}{\partial a}, \\ \frac{\partial g(u)}{\partial c} &= -(a + bu)\frac{\partial g(u)}{\partial b}, & \frac{\partial g(u)}{\partial d} &= 1. \end{aligned} \tag{5.3}$$

Finally, the set of parameters to be calibrated writes $= (a, b, c, d, \kappa, \theta, \epsilon, \rho)$.

Assume that we have a set $\left(PS^{mkt}_{m,n}(0, K_j)\right)_{j\in J, m\in M, n\in N}$ of market swaptions prices of different strikes, maturities and tenors. For the same set of strikes, maturities and tenors, we denote the model prices computed as in Proposition 1.8 by $(PS_{m,n}(0, K_j))_{j\in J, m\in M, n\in N}$. We formulate the calibration problem as an inverse least squares minimization problem with bound constraints and with the Feller condition as an inequality constraint in the following way:

$$\begin{array}{lrcl} \text{argmin}_{\boldsymbol{\Theta}} & \dfrac{1}{2W} \displaystyle\sum_{(j,m,n)\in J\times M\times N} w_{j,m,n} \left(\dfrac{PS_{m,n}(\boldsymbol{\Theta}; 0, K_j) - PS^{mkt}_{m,n}(0, K_j)}{PS^{mkt}_{m,n}(0, K_j)} \right)^2 & & \\ \text{such that} & 2\kappa\theta & \geq & \epsilon^2 \\ & (a, b, c, d, \kappa, \theta, \epsilon, \rho) & \in & (\mathbb{R}_+)^4 \times (\mathbb{R}^*_+)^3 \times (-1; 1) \end{array} \tag{5.4}$$

where $w_{j,m,n}$ are fixed positive weights associated to each swaption and $W = \displaystyle\sum_{(j,m,n)\in J\times M\times N} w_{j,m,n}$. In the following, F stands for the objective function in the previous optimization problem and $\boldsymbol{f}$ stands for the vector of residuals defined by

$$\boldsymbol{f}(\boldsymbol{\Theta}) = \left[\sqrt{\frac{w_{j,m,n}}{W}} \frac{PS_{m,n}(\boldsymbol{\Theta}; 0, K_j) - PS^{mkt}_{m,n}(0, K_j)}{PS^{mkt}_{m,n}(0, K_j)} \right]_{j,m,n}$$

and so that $F(\boldsymbol{\Theta}) = \frac{1}{2}\|\boldsymbol{f}(\boldsymbol{\Theta})\|^2$. As mentioned in [GV19], a regularization term of the form $\frac{1}{2}\|\Gamma\boldsymbol{\Theta}\|^2$ can be added to F to promote some solution. As an example, a classical choice is to take $\Gamma = Id$ to prevent the norm of $\boldsymbol{\Theta}$ of becoming too large.

Such an optimization problem can be solved numerically using general optimization methods like the Nelder-Mead algorithm. It is a direct search method, i.e. it does not require any information about the gradient of the objective function. Instead, it relies on the concept of simplex that is a special polytope of $n+1$ vertices in a n-dimensional space. The algorithm starts by evaluating the objective function on each vertex of the simplex. The vertex with the highest value is replaced by a new point where the objective function is lower. The simplex is updated in this manner until the sample standard deviation of the function values on the current simplex falls below some preassigned tolerance. More details on the Nelder-Mead method can be found in [Nas79]. Note that with this algorithm, one can enforce bound and inequality constraints in

(5.4) by modifying the objective function in the following way:

$$\tilde{F}(\boldsymbol{\Theta}) = \begin{cases} F(\boldsymbol{\Theta}), & \text{if } \boldsymbol{\Theta} \in (\mathbb{R}_+)^4 \times (\mathbb{R}_+^*)^3 \times]-1;1[\text{ and } 2\kappa\theta \geq \epsilon^2, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.5}$$

Although the Nelder-Mead method has proven to be efficient in some contexts, we observed that it turns out to be very time-consuming in our framework. As a matter of fact, this method requires a lot of evaluations of the objective function and the computation of the swaptions prices is very expensive due to the calculation of integrals in the complex field and the recursive definition of the characteristic function. This has already been pointed out by [LDB20]. Consequently, an optimization algorithm that does not require numerous calls to the objective function is preferred in order to achieve a fast calibration. Since we have been able to derive an analytical formula for the gradient of model swaption price, we study gradient-based optimization algorithms. We present two of such methods in the next section.

5.2.4 Calibration using gradient-based algorithms

In this section, we quickly remind the main ideas behind gradient-based algorithms before presenting more specifically two of these algorithms, namely the BFGS and the LM algorithms.

Gradient-based algorithms are iterative methods that start from a given point and proceed by successive adjustments. Each improvement is obtained by moving from the current point along a conveniently chosen descent direction in such a manner that the value of the objective function decreases. We describe in Algorithm 1 the general algorithmic scheme of gradient-based algorithms. More details on gradient-based algorithms can be found in [NW06].

Algorithm 1: Gradient-based algorithms in pseudo code

Input: Initial guess $\boldsymbol{\Theta}_0$, objective function F, objective function gradient ∇F, tolerance ϵ_{tol} and maximum number of iterations k_{max}

```
begin
    k ← 0
    while k ≤ k_max do
        Compute ∇F(Θ_k)
        if ∇F(Θ_k) ≤ ε_tol then
            break
        end
        Compute a descent direction d_k, generally by using ∇F(Θ_k)
        Find α_k such that F(Θ_k + α_k d_k) is reasonably lower than F(Θ_k) (line search)
        Θ_{k+1} ← Θ_k + α_k d_k
    end
end
```

Output: Last computed $\boldsymbol{\Theta}_k$

We now take a closer look at the BFGS and LM algorithms. The BFGS algorithm is a quasi-Newton method, which means that it uses an approximation of the inverse of the Hessian matrix instead of the exact inverse used in Newton's method. It is designed for solving all kinds of unconstrained non-linear optimization problems. The descent direction $\boldsymbol{d}_k$ is given by

$$\boldsymbol{d}_k = -\mathbf{H}_k \nabla F(\boldsymbol{\Theta}_k),$$

where $\mathbf{H}_k$ is an approximation of the inverse of the Hessian matrix that is defined recursively

by,

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{\mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^T \mathbf{H}_k}{\mathbf{y}_k^T \mathbf{H}_k \mathbf{y}_k} + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}$$

with initial value $\mathbf{H}_0 = \mathbf{I}$ ($\mathbf{I}$ is the identity matrix), $\mathbf{s}_k = \mathbf{\Theta}_{k+1} - \mathbf{\Theta}_k$ and $\mathbf{y}_k = \nabla F(\mathbf{\Theta}_{k+1}) - \nabla F(\mathbf{\Theta}_k)$. The gradient of the objective function F writes as a function of the gradient of the swaption price $\nabla PS_{m,n}$:

$$\nabla F(\mathbf{\Theta}) = \frac{1}{W} \sum_{(i,m,n) \in I \times M \times N} w_{j,m,n} \nabla PS_{m,n}(\mathbf{\Theta}; 0, K_j) \frac{PS_{m,n}(\mathbf{\Theta}; 0, K_j) - PS_{m,n}^{mkt}(0, K_j)}{PS_{m,n}^{mkt}(0, K_j)^2}.$$

The calibration problem (5.4) is a constrained optimization problem and thus, the BFGS algorithm can not be used. We therefore rely on an extension of the classical BFGS algorithm, known as L-BFGS-B, that has been developed to handle bound constraints [BLNZ95]. However, inequality constraints can not be easily enforced. Consequently, we relax the Feller condition when using this method.

Furthermore, the Levenberg-Marquardt algorithm is specifically designed for solving non-linear least squares problems, which is exactly the type of problem we are coping with. This algorithm has the particularity of behaving like the steepest descent method when the current point is far (in some sense) of a (the) solution and of behaving like the Gauss-Newton method when the current point is near of a (the) solution. This is achieved by introducing a damping parameter μ_k in the expression of the descent direction, as follows

$$\boldsymbol{d}_k = -(\boldsymbol{J}_k^T \boldsymbol{J}_k + \mu_k \mathbf{I})^{-1} \boldsymbol{g}_k$$

where $\boldsymbol{J}_k$ is the Jacobian matrix of $\boldsymbol{f}$ in $\mathbf{\Theta}_k$, $\boldsymbol{g}_k = \boldsymbol{J}_k^T \boldsymbol{f}(\mathbf{\Theta}_k)$ is the gradient of F in $\mathbf{\Theta}_k$. For large values of μ_k (compared to coefficients of $\boldsymbol{J}_k^T \boldsymbol{J}_k$), we have $\boldsymbol{d}_k \simeq -\frac{1}{\mu_k} \boldsymbol{g}_k$ which corresponds to the descent direction in the steepest descent method. For small values of μ_k, we have $\boldsymbol{d}_k \simeq -(\boldsymbol{J}_k^T \boldsymbol{J}_k)^{-1} \boldsymbol{g}_k$ which corresponds to the descent direction in the Gauss-Newton method. The updating strategy of μ_k is described in Algorithm 2 in Appendix E. There are a plenty of different updating strategies leading to several versions of the Levenberg-Marquardt algorithm. More details on this algorithm can be found in [MNT99].

Adding bound constraints to the LM algorithm

As for the BFGS algorithm, the classic LM algorithm does not handle bound and inequality constraints. However, it can be extended to do so. Bound constraints can be ensured by using a projection of $\mathbf{\Theta}_k$ onto the feasible set. We detail the modifications in Algorithm 3 in Appendix E. Handling linear inequality constraints requires many modifications so we will not detail them here but the interested reader can find a discussion on this topic in [NW06] (Chapter 15). Observe that the Feller condition is not linear in the parameters of the DDSVLMM. However it can be easily linearized using the following change of variables:

$$\tilde{\kappa} = \ln \kappa, \quad \tilde{\theta} = \ln \theta, \quad \tilde{\epsilon} = \ln \epsilon.$$

The Feller condition writes in term of the new volatility parameters as:

$$\tilde{\kappa} + \tilde{\theta} + \ln 2 \geq 2\tilde{\epsilon}.$$

To account for this change of variables, the gradient of the swaption price has to be modified by replacing the partial derivatives with respect to κ, θ and ϵ by the corresponding derivatives

with respect to $\tilde{\kappa}$, $\tilde{\theta}$ and $\tilde{\epsilon}$:

$$\frac{\partial PS}{\partial \tilde{\kappa}}(\mathbf{\Theta}; 0, K_j) = \kappa \frac{\partial PS}{\partial \kappa}(\mathbf{\Theta}; 0, K_j),$$
$$\frac{\partial PS}{\partial \tilde{\theta}}(\mathbf{\Theta}; 0, K_j) = \theta \frac{\partial PS}{\partial \theta}(\mathbf{\Theta}; 0, K_j),$$
$$\frac{\partial PS}{\partial \tilde{\epsilon}}(\mathbf{\Theta}; 0, K_j) = \epsilon \frac{\partial PS}{\partial \epsilon}(\mathbf{\Theta}; 0, K_j).$$

5.3 Calibration results

In this section, we present our experimental results for the calibration of the DDSVLMM (1.49) using the BFGS and LM algorithms. We first detail the market data to be replicated and discuss some implementation aspects. Then, we compare the BFGS and LM routines with existing calibration methods with regards to the objective function value and to the computational time.

For the calibration, we used a set of 280 market EURO swaptions volatilities. During the numerical procedure, these volatilities are converted into prices using the Bachelier formula based on a rate curve as used under the Solvency II regulation[1] (see Chapter 1). The ATM swaptions maturities and tenors considered range into $\{1, \ldots, 10, 15, 20, 25, 30\}$. For away-from-the-money swaptions, we consider the same range for maturities and focus on a 10-year reference tenor; the strikes (in bps) range into $+/-\{25, 50, 100\}$. As mentioned previously, the shift δ is objectified otherwise: we take $\delta = 0.1$. The inter-forward correlation structure, captured by the $\boldsymbol{\beta_k}$ parameters, is assessed by an PCA technique we do not detail here. The number N_f of risk factors is set to 2.

We implemented the pricing and gradient functions in C++. We used the R base function `optim` for the Nelder-Mead and BFGS algorithms and we used the C++ `LEVMAR` package [L+05] for the Levenberg-Marquardt algorithm. This choice is particularly motivated by the fact that the `LEVMAR` package implements the extended version of the Levenberg-Marquardt algorithm allowing to handle bound constraints and the extended version allowing to handle both bound and linear inequality constraints. As for the computation of the numerical integral required in the pricing and the gradient functions, we resorted to the Gauss-Laguerre quadrature with 90 nodes.

5.3.1 Methods accuracy

We compare the BFGS and Levenberg-Marquardt algorithms with existing calibration methods based on the criteria of the objective function value. First, let us introduce the three reference calibration methods used for the purpose of comparison.

The first one is the classical Heston method in which the price is computed through the formula of Proposition 1.8 and the optimization is performed via the Nelder-Mead algorithm. We set the maximum number of iterations to 500 and we repeat the optimization 3 times in order to achieve a better convergence.

The second calibration method relies on Edgeworth approximations: it consists in using an approximate swaption price obtained using an Edgeworth expansion of the unknown density of the swap rate $\mathbb{R}$ (see [LDB20] for a thorough description of the method). The associated optimization method is the Nelder-Mead algorithm. We use the same parametrization for the Nelder-Mead method as for the classical Heston method.

[1] Available at www.eiopa.europa.eu

The last method is based on the LM algorithm but associated with a numerical gradient estimation instead of using the derived analytical gradient. The price is still computed with pricing formula of Proposition 1.8. We use the central difference scheme in order to approximate the gradient as:

$$\nabla PS(\boldsymbol{\Theta}; 0, K_j) \simeq \frac{PS(\boldsymbol{\Theta} + \boldsymbol{h}; 0, K_j, T_m, T_n) - PS(\boldsymbol{\Theta} - \boldsymbol{h}; 0, K_j)}{2\boldsymbol{h}}$$

where $\boldsymbol{h} := h\boldsymbol{e}$ with $\boldsymbol{e}$ a vector whose components are 1 and h a small quantity. We take $h = 10^{-8}$ and a maximal number of 15 iterations.

Let us also present the different parametrizations studied for the BFGS and Levenberg-Marquardt algorithms relying on the analytical gradient as derived in this chapter. For the BFGS algorithm, we test one configuration in which the maximum number of iterations is set to 30. For the Levenberg-Marquardt algorithm, we test two configurations having the same tolerance levels ϵ_1, ϵ_2 and ϵ_3 that are set to 10^{-10} (see Appendix E.2.1). The two configurations, respectively denoted by LM-BC-15 and LM-BC-30, use the version of the LM algorithm allowing to handle bound constraints only with a maximum number of iterations of 15 and 30 respectively.

We summarize the studied methods and their main characteristics in Table 5.1.

Method name	Optimization method	Maximum number of iterations	Ensures Feller condition	Features
Heston	Nelder-Mead	500	Yes	Repeated 3 times
Edgeworth	Nelder-Mead	500	Yes	Repeated 3 times. Uses a different swaption pricing formula.
LM-NUM	Levenberg-Marquardt	15	No	Uses the numerical gradient
BFGS	L-BFGS-B	30	No	Uses the analytical gradient
LM-BC-30	Levenberg-Marquardt	30	No	Uses the analytical gradient
LM-BC-15	Levenberg-Marquardt	15	No	Uses the analytical gradient

Table 5.1: Summary of studied methods.

Before going further, we precise the bounds, particularly the lower bounds, for the 8 parameters of the DDSVLMM to calibrate. Indeed, we experienced numerical instabilities when some parameters equal zero or are very close to zero. For instance if the speed reversion parameter κ or the volatility of volatility ϵ become almost zero, the behavior of the model is pathologic. Therefore, we give the following lower ($\boldsymbol{LB}$) and upper ($\boldsymbol{UB}$) bounds for $\boldsymbol{\Theta} = (a, b, c, d, \kappa, \theta, \epsilon, \rho)$:

$$\begin{aligned} \boldsymbol{LB} &:= (10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, -0.999), \\ \boldsymbol{UB} &:= (+\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty, 0.999). \end{aligned}$$

The procedure that has been led in order to compare the various calibration methods is the following: for each set of data, we draw randomly 100 initial parameter starting values between $\boldsymbol{LB}$ and $\boldsymbol{UB}$ that satisfy the Feller condition and we perform the calibration for each described method starting from each of these initial guess. From this procedure, we retrieve the boxplots of Figure 5.1 using EURO data, which provide statistics on the objective function value over the 100 calibrations.

Note first that the variance (induced by the randomness in the initial guesses) of replication errors when using Nelder-Mead and BFGS algorithms are comparable and significantly lower to that obtained when using Levenberg-Marquardt routines. To this extent, Nelder-Mead and BFGS seem less dependent to initial guess. However, median errors reached by the Levenberg-Marquardt are lower: this can be explained by the fact that this algorithm is particularly suited for least-squares problems.

Concerning the Levenberg-Marquardt approach more specifically, we first note that increas-

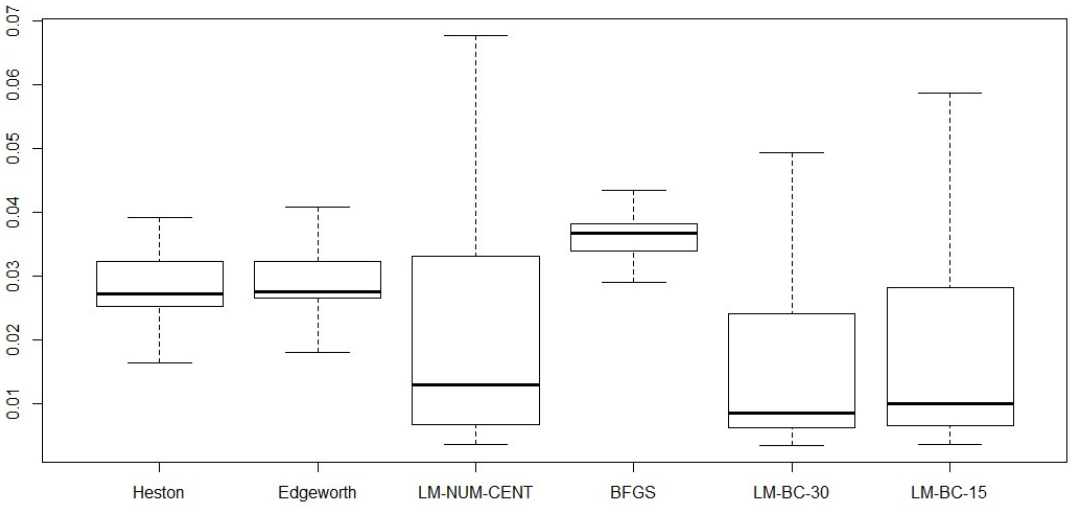


Figure 5.1: Boxplots of our selected benchmark of calibration methods.

ing the number of iterations yields a better calibration (in average and in variance), which is an expected behaviour. Using the Levenberg-Marquardt technique coupled with a numerical approximation of the gradient of the target function leads to a rather wide range of objective function values, which illustrates the benefit of using an analytical gradient in comparison. In addition, the median value for this approach is higher to the one obtained when using Levenberg-Marquardt optimization with analytical gradient. Therefore, the information conveyed by the analytical Hessian matrix turns out to be valuable in order to stabilize the calibration process and reduce the dependency on the starting point.

Moreover, it is worth mentioning that those results suggest that the target function is not convex; otherwise we could have reasonably expect less dependency of the algorithms with respect to the starting points.

So far we did not impose the Feller's condition to be satisfied by the outputs of the calibration procedures. The percentages of obtained parameters (over 100 calibrations) that do not satisfy it are given in Table 5.2. The number of unsatisfied Feller conditions is rather significant especially for the LM algorithm. Note that such condition is always ensured for Nelder-Mead based calibrations as the Feller condition has been imposed as depicted in (5.5).

Method	Percentage of unsatisfied Feller condition
Heston	0 %
Edgeworth	0 %
LM-NUM	34 %
BFGS	13 %
LM-BC-30	33 %
LM-BC-15	32 %

Table 5.2: Percentages of unsatisfied Feller condition for each method.

Numerical results with bound constraints

In view of the previous results, we propose to study two other configurations of the Levenberg-Marquardt algorithm. The first one, denoted by LM-BLEIC, uses the version of the Levenberg-

Marquardt algorithm allowing to handle both bound constraints and linear inequality constraints. The maximum number of iterations is set to 50. The second one, denoted by LM-BLEIC-NM, uses the same algorithm as the LM-BLEIC configuration but here 200 iterations of the Nelder-Mead method are peformed at the end of the Levenberg-Marquardt algorithm when the latter converged towards a point whose objective function value is greater than a given threshold set to 0.3. We present the boxplots for the LM-BLEIC and LM-BLEIC-NM methods in Figure 5.2 using EURO data. We observe that for some initial guesses, the LM-BLEIC

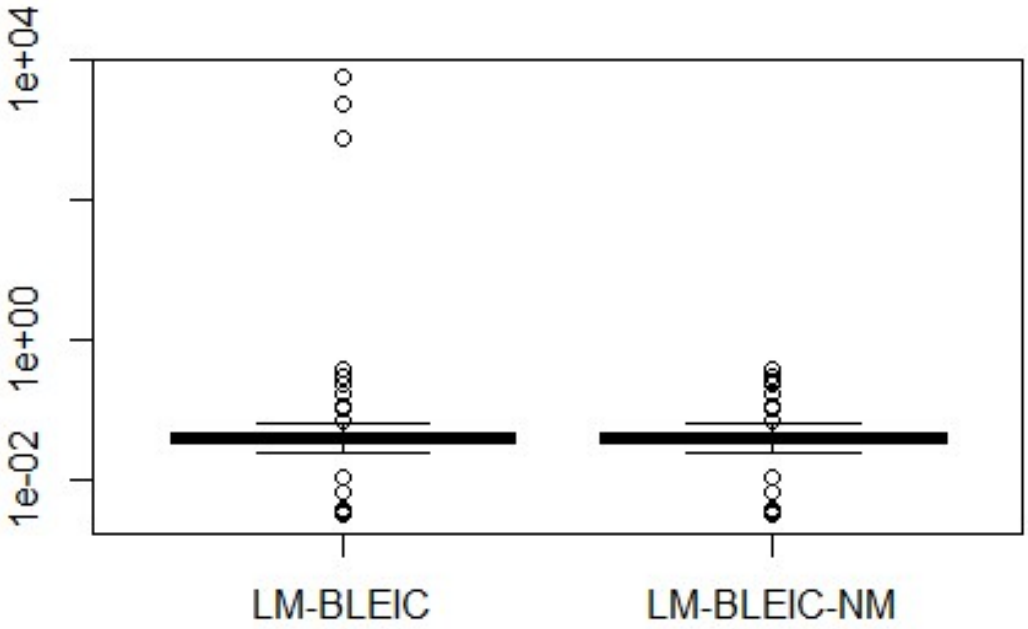


Figure 5.2: Boxplots for LM-BLEIC and LM-BLEIC-NM (log-scale).

method converges towards points at which the objective function takes extremely high values. This is actually due to the fact that the implementation of the Levenberg-Marquardt algorithm handling linear inequality constraints can not ensure that the points stay in the feasible set. As a consequence, some parameters may take negative values leading to numeric precision errors forcing the algorithm to stop prematurely. This is the reason why we introduced the LM-BLEIC-NM. When the objective function value at the exit of the LM-BLEIC method exceeds a given threshold, we perform some Nelder-Mead iterations in order to achieve a better optimum. This method gives us quite satisfying results compared to LM-BLEIC, as far as it allows to get rid of the extreme points. Using the Nelder-Mead allows thus to reduce the variance of the outputs of the calibration procedures but does not significantly reduce the median value of the replication errors which remain close.

5.3.2 Time efficiency

Following the same procedure as the one presented in the previous section, we compute the average CPU time, the average number of calls to the objective function and the average duration of a call. For gradient-based methods, we also compute the average number of calls to the gradient and the average duration of a call. The computations were performed on a computer with a 2.6 gigahertz Intel Core i7 processor and 8 gigabytes of RAM. The results are presented in Table 5.3. Note that when using the Levenberg-Marquardt algorithm, we compute the average call time to the residual function f and its gradient ∇f instead of the objective function F and

its gradient ∇F because the `LEVMAR` implementation takes as inputs the functions f and ∇f, see Section 5.2.3. This explains why there is a difference between BFGS and Levenberg-Marquardt methods in terms of average call time to the gradient. Note also that calling the objective function in the Heston method appears to take in average less time than in gradient-based methods since F is actually replaced by $\tilde{F}$ defined in (5.5) for Nelder-Mead methods, which simply returns a large value in all cases where the Feller condition is not satisfied.

Method	Average CPU time	Average number of calls to F/f	Average call time to F/f	Average number of calls to $\nabla F/\nabla f$	Average call time to $\nabla F/\nabla f$
Heston	159.45 s	1489.20	0.11 s	0	0 s
Edgeworth	8.47 s	1469.46	5.55 $\times 10^{-3}$ s	0	0 s
LM-NUM	40.06 s	277.36	0.14 s	0	0 s
BFGS	36.40 s	39.34	0.14 s	39.34	0.78 s
LM-BC-30	33.89 s	104.80	0.14 s	30.00	0.63 s
LM-BC-15	14.94 s	38.93	0.14 s	15.00	0.63 s
LM-BLEIC	44.29 s	95.76	0.14 s	48.34	0.63 s
LM-BLEIC-NM	45.51 s	102.29	0.14 s	48.34	0.63 s

Table 5.3: Computational times.

The gradient-based algorithms (including those using numerical gradient) appear to be much faster than the classical Heston calibration method using the Nelder-Mead algorithm, since they provide reductions in computational time ranging from 71 % (LM-BLEIC-NM) to 91 % (LM-BC-15). This gain in time results directly from the massive reduction of the number of calls to the objective function and thus, to the characteristic function. However, these methods are still not as fast as the Edgeworth method which uses a large number of function calls but for which each call is very fast, which was the purpose of the method at its origin, see [LDB20]. However one needs to keep in mind that the reduction in computational time achieved by the Edgeworth expansion comes at the cost of a lower fitting accuracy to market data, as pointed out in [LDB20].

As a main conclusion, the use of the analytical gradient rather than the numerical gradient allows to reduce the calibration duration by a factor of 2.7 when looking at LM-BC-15.

Finally, let us observe that the LM algorithm using analytical gradient and including bound constraints only (LM-BC) is faster than the BFGS algorithm. This can be explained by the fact that the Levenberg-Marquardt algorithm takes advantage from the particular shape of the calibration problem, i.e. a least squares optimization problem.

To conclude this section, we justify numerically why we use the characteristic function by [AMST07] instead of that of [dBRFCU10] in the computation of the swaption price. We compared the average call time of both characteristic function representations over 1000 randomly chosen parameters and observed an average call time to Albrecher's representation 5% lower than the average call time to Cui's representation. The time difference between these two representations of the characteristic function is essentially due to the fact that the coefficients A and B in Albrecher's representation can be easily written as function of a few quantities (see e.g. Appendix 5.3 of [LDB20]) which allows to perform less computations than for Cui's representation.

5.4 Extension of gradient-based approach to polynomial model

We integrate now an approximated gradient in the calibration of the Jacobi-based LIBOR Market Model defined in (4.7) (and thus in a Bachelier framework). We have seen in Chapter 4 that a 4-th order Gram-Charlier expansion provide satisfactory results in terms of data replication and computational time. In this section, we propose to integrate the gradient of swaptions

prices expessed under the proposed model (4.7). Those will be expressed thanks to Gram-Charlier expansions based on Gaussian density g_r as auxiliary one whose mean is still set equal to $\mathbb{E}^S[S_T^{m,n}]$. From (4.18), we approximate swaption prices by

$$PS(\boldsymbol{\Theta};0,K) = B^S(0)\Big(\varphi_0 + \varphi_1\mathbb{E}^S\Big[H_1\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big] + \varphi_2\mathbb{E}^S\Big[H_2\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big] + \varphi_3\mathbb{E}^S\Big[H_3\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big] + \varphi_4\mathbb{E}^S\Big[H_4\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big]\Big).$$

The partial derivative with respect to any parameter of the model can be deduced: for any $x \in \{a,b,c,d,\kappa,\theta,\epsilon,\rho\}$,

$$\partial_x PS(\boldsymbol{\Theta};0,K) = B^S(0)\Big\{\partial_x\varphi_0 + \partial_x\Big(\varphi_1\mathbb{E}^S\Big[H_1\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big]\Big) + \partial_x\Big(\varphi_2\mathbb{E}^S\Big[H_2\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big]\Big) + \partial_x\Big(\varphi_3\mathbb{E}^S\Big[H_3\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big]\Big) + \partial_x\Big(\varphi_4\mathbb{E}^S\Big[H_4\Big(\frac{S_T^{m,n}-K}{\sigma_r}\Big)\Big]\Big)\Big\}.$$

To go further in the calculations, we thus have to compute the gradient of the moments of swap rate process. This could be done using the matrix exponential representation of them as in [AF18]. We propose here an alternative method for computing moments possible for small order ones.

Integral representation of swap rates moments

As a polynomial process, moments of the swap rate defined by (4.7) can be expressed in terms of integrals of deterministic quantities, as mentioned in Section 4.5 of Chapter 4.

Proposition 5.2. *Let $T \le T_m$ and set $v_{min} = 0$. We have:*

- $\mathbb{E}^S[S_T^2] = S_0^2 + \int_0^T \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\|^2\mathbb{E}^S[V_s]\mathrm{d}s$;
- $\mathbb{E}^S[(S_T^{m,n})^3] = (S_0^{m,n})^3 + 3\int_0^T \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\|^2\Big(S_0^{m,n}V_0 + \epsilon\int_0^s \rho(u)\|\boldsymbol{\lambda}^{m,n}(u)\|\big(\mathbb{E}^S[V_u] - \frac{1}{v_{max}} \times \mathbb{E}^S[V_u^2]\big)\mathrm{d}u\Big)\mathrm{d}s$;
- $\mathbb{E}^S[(S_t^{m,n})^4] = (S_0^{m,n})^4 + 6(S_0^{m,n})^2V_0\int_0^T \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(u)\|^2\mathrm{d}u + 12\epsilon\int_0^T \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(u)\|\int_0^u \rho(t)\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(t)\| \Big(S_0^{m,n}V_0 + \epsilon\int_0^t \rho(s)\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\|\big(\mathbb{E}^S[V_s] - \frac{1}{v_{max}}\mathbb{E}^S[V_s^2]\big)\mathrm{d}s - \frac{1}{v_{max}}\Big\{S_0^{m,n}V_0^2 + 2\epsilon\int_0^t \rho(s)\|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\| \big(\mathbb{E}^S[V_s^2] - \frac{1}{v_{max}}\mathbb{E}^S[V_s^3]\big)\mathrm{d}s\Big\}\Big)\mathrm{d}t\mathrm{d}u$,

where the moments of the Jacobi process are expressed as:

- $\mathbb{E}^S[V_t] = \exp\Big(-\kappa\int_0^t \xi^0(s)\mathrm{d}s\Big)\Big(v_0 + \kappa\theta\int_0^t \exp\big(\kappa\int_0^s \xi^0(u)\mathrm{d}u\big)\mathrm{d}s\Big)$;
- $\mathbb{E}^S[V_t^2] = \exp\Big(-\int_0^t(2\kappa\xi^0(u) + \epsilon^2/v_{max})\mathrm{d}u\Big)\Big(V_0^2 + (2\kappa\theta + \epsilon^2)\int_0^t \mathbb{E}^S[V_u]\exp\big(\int_0^u(2\kappa\xi^0(s) + \epsilon^2/v_{max})\mathrm{d}s\big)\mathrm{d}u\Big)$;
- $\mathbb{E}^S[V_t^3] = \exp\Big(-3\int_0^t(\kappa\xi^0(u) + \epsilon^2/v_{max})\mathrm{d}u\Big)\Big(V_0^3 + 3(\kappa\theta + \epsilon^2)\int_0^t \mathbb{E}^S[V_u^2]\exp\big(3\int_0^u(\kappa\xi^0(s) + \epsilon^2/v_{max})\mathrm{d}s\big)\mathrm{d}u\Big)$,

for all $t \ge 0$.

Proof. From (4.7), we express the Jacobi process as the sum of a Lebesgue's and an Ito's integral and from Lemma 4.4 we get that $\mathbb{E}^S[V_t] = v_0 + \int_0^t \kappa\big(\theta - \xi^0(s)\mathbb{E}^S[V_s]\big)\mathrm{d}s$. It is an ordinary differential equation that is uniquely solved by: $\mathbb{E}^S[V_t] = \exp\Big(-\kappa\int_0^t \xi^0(s)\mathrm{d}s\Big)\Big(v_0 + \kappa\theta\int_0^t \exp\big(\kappa\int_0^s \xi^0(u)\mathrm{d}u\big)\mathrm{d}s\Big)$. Squared Jacobi process evolves following $\mathrm{d}V_t^2 = \big(2\kappa\theta - (2\kappa\xi^0(t) + \epsilon^2)V_t - \epsilon^2 V_t^2/v_{max}\big)\mathrm{d}t + 2\epsilon\sqrt{Q(V_t)}\mathrm{d}W_t$ according to Ito's formula. Hence the ordinary differential equation satisfied by the second order moment of the Jacobi process writes $\mathbb{E}^S\left[V_t^2\right] = V_0^2 + \int_0^t \big(2\kappa\theta - (2\kappa\xi^0(s) + \epsilon^2)\mathbb{E}^S[V_s] - \epsilon^2\mathbb{E}^S[V_s^2]/v_{max}\big)\mathrm{d}s$. Solving it yields $\mathbb{E}^S\left[V_t^2\right] = \exp\Big(-\int_0^t(2\kappa\xi^0(u) + \epsilon^2/v_{max})\mathrm{d}u\Big)\Big(V_0^2 + (2\kappa\theta+\epsilon^2)\int_0^t \mathbb{E}^S[V_u]\exp\big(\int_0^u(2\kappa\xi^0(s)+\epsilon^2/v_{max})\mathrm{d}s\big)\mathrm{d}u\Big)$ which is fully known since we have previously determined the expression of the mean of the Jacobi process. Third order moment of the Jacobi process can be determined analogously.

We are now ready to compute small orders moments of the swap rate process. We directly get from (4.7) that $\mathbb{E}^S[S_T^{m,n}] = S_0^{m,n}$. Using Ito's isometry, we have $\mathbb{E}^S[S_T^2] = S_0^2 + \int_0^T \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\|^2\mathbb{E}^S[V_s]\mathrm{d}s$ in which the expression of mean of the Jacobi process can be injected. Concerning the third order moment, we resort again on the Ito's formula to get that $\mathrm{d}\tilde{S}_t^{m,n} = 3(S_t^{m,n})^2\Big(\rho(t)\|\boldsymbol{\lambda}^{m,n}(t)\|\sqrt{Q(V_t)}\mathrm{d}W_t + \sqrt{V_t - \rho(t)^2Q(V_t)}\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(t)\cdot\mathrm{d}\boldsymbol{W}_t^{S,*}\Big) + 3\|\boldsymbol{\lambda}^{m,n}(t)\|^2 S_tV_t\mathrm{d}t$ where $\tilde{S}_t^{m,n} := (S_t^{m,n})^3$. Hence $\mathbb{E}^S[(S_t^{m,n})^3] = (S_0^{m,n})^3 + 3\int_0^t \|\boldsymbol{\lambda}^{\boldsymbol{m},\boldsymbol{n}}(s)\|^2\mathbb{E}^S[S_s^{m,n}V_s]\mathrm{d}s$. By definition of the quadratic covariation, the process $\Big(S_t^{m,n}V_t - \big\langle S_\cdot^{m,n}, V_\cdot\big\rangle_t\Big)_{t\geq 0}$ is a martingale and thus $\mathbb{E}^S[S_t^{m,n}V_t] = S_0^{m,n}V_0 + \mathbb{E}^S[\langle S_\cdot^{m,n}, V_\cdot\rangle_t]$. Combined with the fact that $\langle S_\cdot^{m,n}, V_\cdot\rangle_t = \epsilon\int_0^t \rho(u)\|\boldsymbol{\lambda}^{m,n}(u)\|Q(V_u)\mathrm{d}u$, we get $\mathbb{E}^S[S_t^{m,n}V_t] = S_0^{m,n}V_0 + \epsilon\int_0^t \rho(u)\|\boldsymbol{\lambda}^{m,n}(u)\|\mathbb{E}^S[Q(V_u)]\mathrm{d}u$ which is fully known since $\mathbb{E}^S[Q(V_t)] = \mathbb{E}^S[V_t] - \mathbb{E}^S[(V_t)^2]/v_{max}$. The computation of the fourth order moment of the swap rate process is done similarly using the quadratic covariation between the processes $\Big((S_t^{m,n})^2\Big)_{t\geq 0}$ and $\big(V_t\big)_{t\geq 0}$. □

5.4.1 Numerical results

In all presented experiments, we have set $v_{min} = 0$ and $v_{max} = 9v_0 = 0.9$. We now set the following bounds similarly to the previous one, but with an additional restriction on the parameter θ ensuring the proper definition of the Jacobi process:

$$\boldsymbol{LB} = (10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, -0.999),$$
$$\boldsymbol{UB} = (+\infty, +\infty, +\infty, +\infty, +\infty, v_{max}, +\infty, 0.999).$$

We worked with the `nls.lm` function of R (package `minpack.lm`). On 100 different starting points still randomly generated but satisfying all constraints required under Jacobi-based model, namely

$$2\kappa\min(\xi_{min}^0 v_{max} - \theta, \theta) \geq \epsilon^2 \text{ (Feller condition) and } 4\kappa\theta > \epsilon^2 \text{ (Assumption 1).} \tag{5.6}$$

In all the experiments we present below, we have worked with the numerical gradient of the objective function, approximated by forward difference scheme with step size of 10^{-8}. Indeed, several integrated functions that are involved in the analytical expressions of the gradient (based on the formulas obtained in Proposition 5.2) are of exponential behavior for large time. Consequently, to calculate the gradient with enough precision (using trapezoidal rule or other schemes[2]), a tiny step size has to be chosen in numerical schemes; numerous calls to the in-

[2]Note however that the functions we are dealing with are simply continuous and not differentiable.

tegrated functions would then be necessary. The computational time required to calibrate Jacobi-based model using analytical gradient is thus significant (several hours) making this method out of the operational scope of the present study.

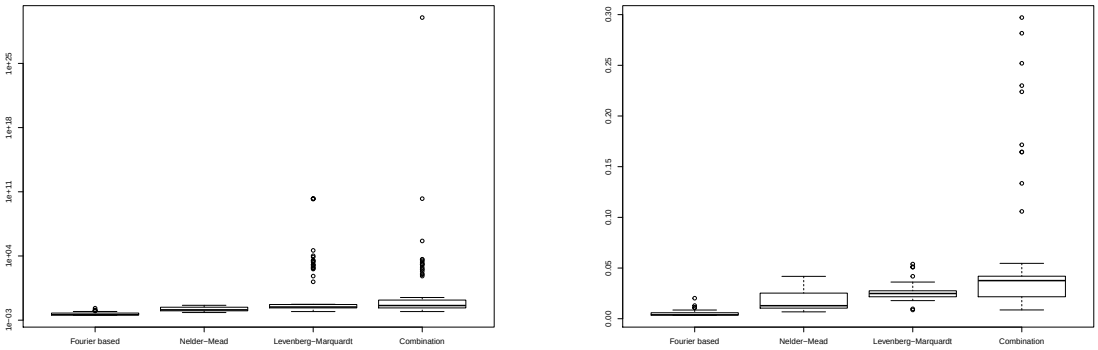
We discuss different methodologies under the Jacobi-based dynamics. In all Gram-Charlier based approaches, we will set $\mu_r = \mathbb{E}^S[S_{T_m}^{m,n}]$ and $\sigma_r = \sqrt{\frac{v_{max}T}{2}\lambda_{max}^2} + 10^{-5}$. First one is the standard Nelder-Mead optimization performed 3 times in a row with a maximum number of iterations set to 500 for each sequence. It is associated with the (truncated) price representation (4.18); it is referred simply as "Nelder-Mead". Second one is Levenberg-Marquardt based optimization with a maximum number of iteration set to 30; it is the "Levenberg-Marquardt" (or "LM") approach (with a maximum number of iterations set to 30). A number of pathologic calibrations (not plotted below) are obtained using such method, and this why we propose to combine it with a run of 200 iterations maximum of Nelder-Nead algorithm when the objective function is above 0.1 at the end of Levenberg-Marquardt algorithm: this is the "Combination" method. To handle parameters constraints, we project the parameters obtained as outputs of the Levenberg-Marquardt approach on the acceptable set before running Nelder-Mead algorithm. Fourth method we propose also aims at handling constraints on parameters. We propose to integrate them directly in the objective function as follows: with the notation above, the objective function F is modified as

$$\tilde{F}(\boldsymbol{\Theta}) = \frac{1}{2}\|\boldsymbol{f}(\boldsymbol{\Theta})\|^2 + \frac{1}{2}\|\boldsymbol{c}(\boldsymbol{\Theta})\|^2$$

where we set $\boldsymbol{c}(\boldsymbol{\Theta}) = \big(\max(0, \epsilon^2 - 2\kappa\min(\xi_{min}^0 v_{max} - \theta, \theta)), \max(0, \epsilon^2 - 4\kappa\theta)\big)$. One advantage of this technique is that it modifies the objective function in a smooth way since $\boldsymbol{c}$ is differentiable once. This modified objective function will be minimized using the Levenberg-Marquardt algorithm (still with 30 iterations max.): this is the "LM with constraints" approach. These four methods are compared to the standard Fourier based approach of Proposition 1.9 associated with the reference dynamics (1.50). Note that for the latter, the relaxed Feller condition $2\kappa\theta \geq \epsilon^2$ is imposed. The optimization is performed using 3 runs of 500 iterations maximum of the Nelder-Mead algorithm. It is referred to as "Fourier based".

The distribution of the replication error under each parametrization are given in Boxplots (5.3). As previously, the Fourier based methodology provides the best replication errors, followed by the Nelder-Mead approach. The Levenberg-Marquardt approach provides satisfactory results close to the Nelder-Mead approach in average. However, the latter induced bigger variance than the previous methods, illustrating again its sensitivity towards starting points. The combination method allows to correct some pathologic calibrations of the LM one but not all of them. It also induces a larger variance explained by the additional constraints taken into account during Nelder-Mead runs. In Boxplots 5.4, we compare the Levenberg-Marquardt associated with the objective function F and the Levenberg-Marquardt associated with $\tilde{F}$. Not surprisingly, the second one is a little less accurate than the unconstrained one but still provides very satisfactory results. However, it turns out that only methods using Nelder-Mead algorithm or the "Combination" method properly handle the constraints (5.6) on parameters, as shown in Table 5.5. Adding constraints in the standard Levenberg-Marquardt –as done for defining $\tilde{F}$ above– is not satisfactory from this point of view. This can be explained by the (*a priori*) non qualification of the constraints in that particular case (see [Gil03] notably Section 4.4 for a deeper discussion on the topic).

Regarding computational times, we gather the average calibration and call times in Tables (5.4). Gradient-based approaches (in which we include the Combination method) allows for significant reduction of the computational time compared to the Fourier-based method (up to almost 93.8% reduction).



(a) All results in logarithmic-scale.

(b) Removal of pathological calibrations.

Figure 5.3: Boxplots of our selected benchmark of calibration methods for Jacobi-based model.

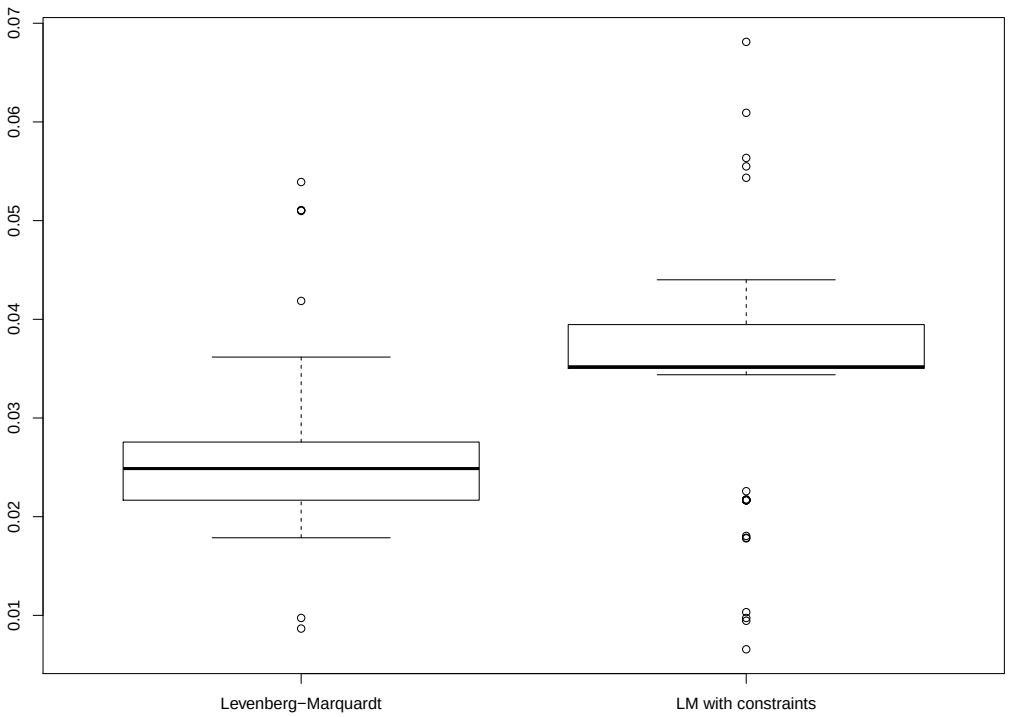


Figure 5.4: Boxplots of our selected benchmark of calibration methods for Jacobi-based model.

As main conclusion, the combination of gradient-based runs followed by a Nelder-Mead algorithm proved to be competitive both in computational time and replication accuracy in present Jacobi-based DDSVLMM.

Method	Average CPU time
Fourier based	829.50
Nelder-Mead	311.00
LM	51.48
Combination	82.60
LM with constraints	61.36

Table 5.4: Computational times (in sec.).

Method	Percentage of unsatisfied constraints
LM	53 %
Combination	1 %
LM with constraints	60 %

Table 5.5: Percentages of unsatisfied parameters constraints for each method.

Appendices

Appendix A

Short rates models in insurance and pricing of interest rates derivatives

A.1 Short rates models in insurance

A.1.1 Hull-White model

This model has been introduced in [HW90] as an extension of the Vasicek and CIR models. In the Vasicek model, the short rate dynamics is given by an Ornstein-Uhlenbeck dynamics. The CIR model uses the square-root process. The Hull-White model generalized the mentioned models by allowing the parameters to be time dependent. Its dynamics writes

$$\mathrm{d}r_t = \kappa(t)\big(\theta(t) - r_t\big)\mathrm{d}t + \sigma(t) r_t^{\alpha}\mathrm{d}W_t,$$

where κ, θ, σ are deterministic functions of time, $\alpha \geq 0$ and $(W_t)_{t\geq 0}$ is a standard Brownian motion. To avoid some over parametrization, it is generally set $\kappa(t) \equiv \kappa$, $\sigma(t) \equiv \sigma$ and $\alpha = 0$ so that the Hull-White model generally refers to the following dynamics:

$$\mathrm{d}r_t = \kappa\big(\theta(t) - r_t\big)\mathrm{d}t + \sigma\mathrm{d}W_t. \tag{A.1}$$

The Hull-White model is particularly convenient as it provides a number of analytical expression easy to implement. First, the short rate can be obtain using Itô's formula as

$$r_t = e^{-\kappa t} r_0 + \kappa \int_0^t e^{-\kappa(t-s)}\theta(s)\mathrm{d}s + \sigma e^{-\kappa t}\int_0^t e^{\kappa s}\mathrm{d}W_s.$$

This expression justifies that the short rate process $(r_t)_{t\geq 0}$ is a Gaussian process and thus so is the integrated process $\left(\int_0^t r_s \mathrm{d}s\right)_{t\geq 0}$.

Negative rates Finally, we take a look at the ability of the model to generate non-positive rates. As a Gaussian process, the short rate defined by (A.1) can take negative values. Mean and variance processes, respectively $(M(t))_{t\geq 0}$ and $(\Sigma(t))_{t\geq 0}$, allow to characterize the distribution of $(r_t)_{t\geq 0}$; they respectively write:

$$M(t) = e^{-\kappa t} r_0 + \kappa \int_0^t e^{-\kappa(t-s)}\theta(s)\mathrm{d}s,$$
$$\Sigma(t) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}).$$

We then obtain, for any $x \in \mathbb{R}$,

$$\mathbb{P}(r_t \leq x) = \mathbb{P}\left(M(t) + \sqrt{\Sigma(t)}G \leq x\right) = \Phi\left(\frac{x - M(t)}{\sqrt{\Sigma(t)}}\right) > 0, \tag{A.2}$$

where $G \sim \mathcal{N}(0,1)$ and Φ is the cumulative distribution function of the standard normal distribution.

Pricing of ZC bonds and derivatives Since dealing with Gaussian process, computing the ZC bond price reduces to the computation of the moment generating function of the integrated short rate process according to the pricing formula (A.17): it can thus be done through closed-form formula here. Since, for $\kappa > 0$,

$$\begin{aligned}
\mathbb{E}\left[\int_t^T r_u \mathrm{d}u \middle| \mathcal{F}_t\right] &= r_t \frac{1 - e^{-\kappa(T-t)}}{\kappa} + \int_t^T \theta(s)\big(1 - e^{-\kappa(T-s)}\big)\mathrm{d}s, \\
\mathrm{Var}\left(\int_t^T r_u \mathrm{d}u \middle| \mathcal{F}_t\right) &= \frac{\sigma^2}{\kappa^2}\int_t^T (1 - e^{-\kappa(T-s)})^2 \mathrm{d}s \\
&= \frac{\sigma^2}{\kappa^2}\left((T-t) - \frac{2}{\kappa}\big(1 - e^{-\kappa(T-t)}\big) + \frac{1}{2\kappa}(1 - e^{-2\kappa(T-t)})\right),
\end{aligned}$$

and using fact that if $G \sim \mathcal{N}(m, s^2)$, then $\mathbb{E}[e^{-G}] = e^{-m+s^2/2}$ we finally get:

$$P(t,T) = \exp\left(-r_t \frac{1 - e^{-\kappa(T-t)}}{\kappa} - \int_t^T \big(1 - e^{-\kappa(T-s)}\big)\theta(s)\mathrm{d}s + \frac{\sigma^2}{2\kappa^2}\int_t^T (1 - e^{-\kappa(T-s)})^2 \mathrm{d}s\right). \tag{A.3}$$

An important consequence: forward rates and swap rates can also be analytically express in this model. Forward rates in (1.25) express as

$$\begin{aligned}
F(t,T,S) = \frac{1}{S-T}\Bigg(\exp\Bigg(&-\frac{r_t}{\kappa}\left(e^{-\kappa(S-t)} - e^{-\kappa(T-t)}\right) + \int_t^S \big(1 - e^{-\kappa(S-u)}\big)\theta(u)\mathrm{d}u \\
&- \int_t^T \big(1 - e^{-\kappa(T-u)}\big)\theta(u)\mathrm{d}u + \frac{\sigma^2}{2\kappa^2}\Big\{\int_t^T \big(1 - e^{-\kappa(T-u)}\big)^2 \mathrm{d}u - \int_t^S \big(1 - e^{-\kappa(S-u)}\big)^2 \mathrm{d}u\Big\}\Bigg) - 1\Bigg).
\end{aligned}$$

From equation (A.2), one can also derive analytical expression for option on ZC bonds. For instance consider a call option of maturity T on a ZC bond of maturity $S > T$ and of strike K.

The spot price of the option writes

$$\begin{aligned}
\pi_0 &= \mathbb{E}^*\left[\exp\left(-\int_0^T r_u \mathrm{d}u\right)(P(T,S)-K)_+\right] \\
&= \mathbb{E}^*\left[\exp\left(-\int_0^T r_u \mathrm{d}u\right)P(T,S)\mathbb{1}_{P(T,S)\geq K}\right] - K\mathbb{E}^*\left[\exp\left(-\int_0^T r_u \mathrm{d}u\right)\mathbb{1}_{P(T,S)\geq K}\right] \\
&= \mathbb{E}^*\left[\exp\left(-\int_0^T r_u \mathrm{d}u\right)\mathbb{E}^*\left[e^{-\int_T^S r_u \mathrm{d}u}\middle|\mathcal{F}_T\right]\mathbb{1}_{P(T,S)\geq K}\right] - KP(0,T)\mathbb{E}^T\left[\mathbb{1}_{P(T,S)\geq K}\right] \\
&= \mathbb{E}^*\left[\exp\left(-\int_0^S r_u \mathrm{d}u\right)\mathbb{1}_{P(T,S)\geq K}\right] - KP(0,T)\mathbb{E}^T\left[\mathbb{1}_{P(T,S)\geq K}\right] \\
&= P(0,S)\mathbb{E}^S\left[\mathbb{1}_{P(T,S)\geq K}\right] - KP(0,T)\mathbb{E}^T\left[\mathbb{1}_{P(T,S)\geq K}\right],
\end{aligned}$$

where we used that $\exp\left(-\int_0^T r_u \mathrm{d}u\right)\mathbb{1}_{P(T,S)\geq K}$ is $\mathcal{F}_T$-measurable and the forward pricing relation (A.20) to obtain last equality. We first inspect the event in the indicator functions:

$$\begin{aligned}
P(T,S)\geq K &\iff \ln P(T,S)\geq \ln K \\
&\iff r_T \leq r^*
\end{aligned}$$

where $r^* = \frac{-\kappa}{1-e^{-\kappa(S-T)}}\left(\ln K + \int_T^S \left(1-e^{-\kappa(S-u)}\right)\theta(u)\mathrm{d}u - \frac{\sigma^2}{2\kappa^2}\int_T^S \left(1-e^{-\kappa(S-u)}\right)^2 \mathrm{d}u\right)$. We are thus led to study the distribution of the short rate under forward measure $\mathbb{P}^S$ and $\mathbb{P}^T$. Girsanov's theorem provides that, under $\mathbb{P}^T$, the short rate process expresses as

$$r_t = e^{-\kappa t}r_0 + \kappa\int_0^t e^{-\kappa(t-s)}\theta(s)\mathrm{d}s - \frac{\sigma^2}{\kappa}\int_0^t \left(1-e^{-\kappa(T-s)}\right)e^{-\kappa(t-s)}\mathrm{d}s + \sigma e^{-\kappa t}\int_0^t e^{\kappa s}\mathrm{d}W_s.$$

It is still a Gaussian process characterized by its mean process $M_T(t) = e^{-\kappa t}r_0 + \kappa\int_0^t e^{-\kappa(t-s)}\theta(s)\mathrm{d}s - \frac{\sigma^2}{\kappa}\int_0^t (1-e^{-\kappa(T-s)})e^{-\kappa(t-s)}\mathrm{d}s$ and variance process that remains unchanged by the change of probability $\Sigma_T(t) = \Sigma(t)$. Under $\mathbb{P}^S$, we similarly get that the only mean process is modified as $M_S(t) = e^{-\kappa t}r_0 + \kappa\int_0^t e^{-\kappa(t-u)}\theta(u)\mathrm{d}u - \frac{\sigma^2}{\kappa}\int_0^t (1-e^{-\kappa(S-u)}e^{-\kappa(t-u)}\mathrm{d}u$. The option price is

$$\pi_0 = P(0,S)\Phi\Big(\frac{r^*-M_S(T)}{\sqrt{\Sigma(T)}}\Big) - KP(0,T)\Phi\Big(\frac{r^*-M_T(T)}{\sqrt{\Sigma(T)}}\Big).$$

Fitting the initial term-structure The function θ is used to fit the initial term-structure of ZC bonds $t \mapsto P^{Mkt}(0,t)$ observed on the market. We assume that this map is twice differentiable. We denote by $\frac{\partial f(P^{Mkt}(0,t))}{\partial T}$ the derivative evaluated in t of the function $t \mapsto f(P^{Mkt}(0,t))$ for some smooth function f.

Proposition A.1. *Setting for all $t \leq T$,*

$$\theta(t) = -\frac{1}{\kappa}\frac{\partial^2 \ln P^{Mkt}(0,t)}{\partial T^2} - \frac{\partial \ln P^{Mkt}(0,t)}{\partial T} + \frac{\sigma^2}{2\kappa^2}(1-e^{-2\kappa t})$$

allows to perfectly reproduce the initial zero-coupon bond graph:

$$P(0,t) = P^{Mkt}(0,t),\ t\leq T.$$

Proof. To replicate the market prices, one should impose, for all $t \leq T$,

$$\begin{aligned} &\ln P^{Mkt}(0,t) = \ln P(0,t) \\ &= -r_0 \frac{e^{-\kappa t} - 1}{k} - \int_0^t (1 - e^{-\kappa(t-u)})\theta(u)\mathrm{d}u + \frac{\sigma^2}{2\kappa^2} \int_0^t (1 - e^{-\kappa(t-u)})^2 \mathrm{d}u. \end{aligned} \tag{A.4}$$

Using that $\partial_t \Big(\int_0^t (1 - e^{-\kappa(t-u)})\theta(u)\mathrm{d}u \Big) = \kappa e^{-\kappa t} \int_0^t e^{\kappa u}\theta(u)\mathrm{d}u$, $\partial_t \Big(\int_0^t (1 - e^{-\kappa(t-u)})^2 \mathrm{d}u \Big) = (1 - e^{-\kappa t})^2$ and $\partial_t \Big(\kappa e^{-\kappa t} \int_0^t e^{\kappa u}\theta(u)\mathrm{d}u \Big) = -\kappa^2 e^{-\kappa t} \int_0^t e^{\kappa u}\theta(u)\mathrm{d}u + \kappa\theta(t)$ allows to differentiate twice (A.4), which gives the result. □

Limitations So far, the market hazard is modelled by one-dimensional Brownian motions. Consequently, the correlation between rates of different maturities is perfect: it can be viewed in the expression (A.3) that randomness ZC bonds prices is only due to the short rate process which does to depend on the maturity of the considered ZC bonds. This behaviour is not observed on market as illustrated in Figure A.1.

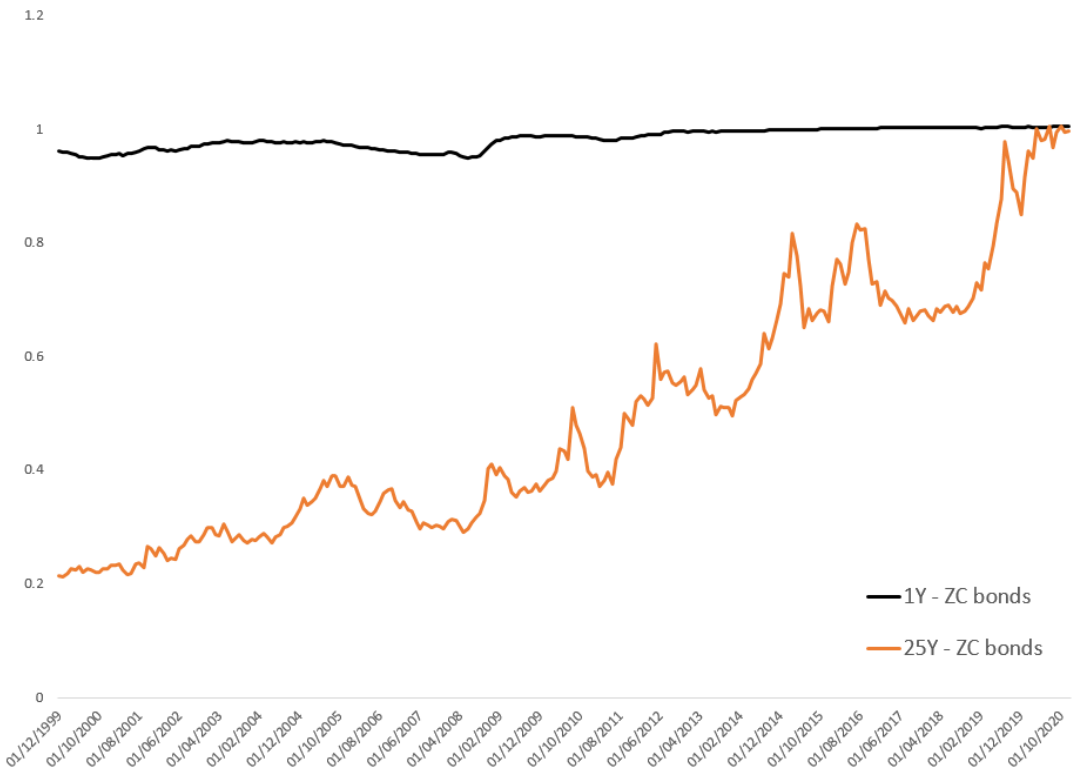


Figure A.1: ZC bond prices of maturity 1 year (black) and 25 years (red) between 12/01/1999 and 12/31/2020. Estimated correlation between the two series: $\hat{\rho} = 0.865$; 95% confidence interval: $\hat{\rho} \in [0.830, 0.893]$.

Such a framework is thus not suited to reproduce some empirical observations and to insurers needs. As insurance policies are generally associated with long-term guarantees, the joint distribution of interest rates of different maturities impact their investment strategy. Multidimensional models are appropriated to overcome this difficulty.

A.1.2 G2++

The G2++ model describes the short rate dynamics as being the sum of two Gaussian processes adjusted by a deterministic function to replicate the spot term-structure, in a similar fashion as described above.

Namely, under the risk-neutral measure $\mathbb{P}^*$, the short-rate is assumed to be given by

$$r_t = x_t + y_t + \varphi(t), \ r_0 = r^0$$

where stochastic processes $(x_t)_{t\geq 0}$ and $(y_t)_{t\geq 0}$ are determined by the equations

$$\begin{aligned} \mathrm{d}x_t &= -ax_t\mathrm{d}t + \sigma \mathrm{d}Z_t^1, \\ \mathrm{d}y_t &= -by_t\mathrm{d}t + \eta \mathrm{d}Z_t^2, \end{aligned} \tag{A.5}$$

with $(x_0, y_0) = (0, 0)$ and where Z^1, Z^2 are two Brownian motions whose joint distribution is characterized by the quadratic variation $\langle Z_\cdot^1, Z_\cdot^2 \rangle_t = \rho t$ for some $\rho \in [-1, 1]$, r^0, a, b, σ, η are positive constants and φ is deterministic function defined over a finite time lapse $[0, \bar{T}]$; in particular, $\varphi(0) = r^0$. Integration of (A.5) can also be done such that for each $t \leq \bar{T}$,

$$r_t = \varphi(t) + \sigma \int_0^t e^{-a(t-u)}\mathrm{d}Z_u^1 + \eta \int_0^t e^{-b(t-u)}\mathrm{d}Z_u^2. \tag{A.6}$$

The short rate is thus also a Gaussian process in the G2++ specification.

Negative rates By modelling the short rate through a Gaussian process, the G2++ allows to generate non-positive values. From (A.6), one deduces that under $\mathbb{P}^*$, r_t is distributed as a Gaussian variable of mean $\varphi(t)$ and variance $v(t) := \frac{\sigma^2}{2a}(1-e^{-2at}) + \frac{\eta^2}{2b}(1-e^{-2bt}) + 2\frac{\eta\sigma\rho}{a+b}(1-e^{(a+b)t})$ and one gets

$$\mathbb{P}^*(r_t \leq x) = \Phi\left(\frac{x - \varphi(t)}{\sqrt{v(t)}}\right) > 0, \ \forall x \in \mathbb{R}.$$

Pricing of ZC bonds and derivatives As in the previous section, ZC pricing is based on the study of the integrated process $\int_0^\cdot r_u \mathrm{d}u$; computations are similar so we do not provide them here and simply bring in the results. The interested reader is referred to [BM07] in which computations can be found in detail.

As a Lebesgue integral of a Gaussian process, $\int_0^\cdot r_u \mathrm{d}u$ is still a Gaussian process. As for the Hull & White model, it is enough to determine mean and variance processes.

Proposition A.2 ([BM07]). *For $t \leq T$, conditionally to the sigma-field generated by the pair $(Z_\cdot^1, Z_\cdot^2)$ up to time t, the random variable $\int_t^T r_u \mathrm{d}u$ is a Gaussian variable of mean*

$$M_{t,T} = \frac{1 - e^{-a(T-t)}}{a} x_t + \frac{1 - e^{-b(T-t)}}{b} y_t,$$

and of variance

$$\Sigma_{t,T} = \frac{\sigma^2}{a^2}\left((T-t) - \frac{2}{a}\left(1-e^{-a(T-t)}\right) + \frac{1}{2a}\left(1-e^{-2a(T-t)}\right)\right)$$
$$+ \frac{\eta^2}{b^2}\left((T-t) - \frac{2}{b}\left(1-e^{-b(T-t)}\right) + \frac{1}{2b}\left(1-e^{-2b(T-t)}\right)\right)$$
$$+ 2\rho\frac{\sigma\eta}{ab}\left(T-t+\frac{e^{-a(T-t)}-1}{a}+\frac{e^{-b(T-t)}-1}{b}-\frac{e^{-(a+b)(T-t)}-1}{a+b}\right).$$

This is useful to derive a ZC price formula: as a direct application of fact that $\mathbb{E}[e^{-G}] = e^{-m+s^2/2}$ when $G \sim \mathcal{N}(m, s^2)$ combined with Proposition A.2 we get that, in the G2++ specification, the ZC bond price expresses as

$$P(t,T) = \exp\left(-\int_t^T \varphi(s)\mathrm{d}s - \frac{1-e^{-a(T-t)}}{a}x_t - \frac{1-e^{-b(T-t)}}{b}y_t + \frac{1}{2}\Sigma_{t,T}\right). \tag{A.7}$$

The expression of forward rates follows, for $t \leq T < S$,

$$F(t,T,S) = \frac{1}{S-T}\left[\exp\left(\int_T^S \varphi(u)\mathrm{d}u + \frac{e^{-a(T-t)}-e^{-a(S-t)}}{a}x_t + \frac{e^{-b(T-t)}-e^{-b(S-t)}}{b}y_t\right.\right.$$
$$\left.\left.+\frac{1}{2}(\Sigma_{t,T}-\Sigma_{t,S})\right)-1\right]$$

Fitting the initial term-structure The deterministic function φ is used to exactly reproduce the initial ZC bond curve.

Proposition A.3. *Setting, for all $t \leq T$,*

$$\varphi(t) = -\frac{\partial \ln P(0,t)}{\partial T} + \frac{\sigma^2}{2a^2}(1-e^{-at})^2 + \frac{\eta^2}{2b^2}(1-e^{-bt})^2 - \rho\frac{\sigma\eta}{ab}(1-e^{-at})(1-e^{-bt})$$

allows to reproduce the market price of ZC bonds, that is:

$$P(0,t) = P^{Mkt}(0,t),\ t \leq T.$$

Proof. Using (A.7) and imposing $\ln P(0,t) = \ln P^{Mkt}(0,t)$ yields the result. □

A.1.3 Two-factor Hull-White model

The two-factor Hull-White model introduces an additional source of risk in the short rate through its mean-reversion level; the assumed dynamics for the short rate under the risk-neutral measure writes

$$\begin{cases} \mathrm{d}r_t &= k\left(\theta(t)+m_t-r_t\right)\mathrm{d}t + \sigma_1\mathrm{d}Z_t^1, \\ \mathrm{d}m_t &= -b'm_t\mathrm{d}t + \sigma_2\mathrm{d}Z_t^2, \end{cases} \tag{A.8}$$

with $(r_0, m_0) = (r^0, 0)$ and where (Z^1, Z^2) is a two-dimensional Brownian motion such that their quadratic variation writes $\left\langle Z_\cdot^1, Z_\cdot^2\right\rangle_t = \rho' t$ for some $-1 \leq \rho' \leq 1$ and $k, \sigma_1, b, \sigma_2, r^0$ are

positive constants. An application of the Itô's formula allows to integrate (A.8) to get

$$\begin{aligned} r_t &= r_0 e^{-kt} + k\int_0^t \theta(u)e^{-k(t-u)}\mathrm{d}u + k\int_0^t m_u e^{-k(t-u)}\mathrm{d}u + \sigma_1\int_0^t e^{-k(t-u)}\mathrm{d}Z_u^1, \\ m_t &= \sigma_2\int_0^t e^{-b'(t-u)}\mathrm{d}Z_u^2. \end{aligned} \tag{A.9}$$

Injecting the expression of the mean-reversion level m into the second integral in the expression of the short rate and an integration by part yield, for $k \neq b'$,

$$r_t = r_0 e^{-kt} + k\int_0^t \theta(u)e^{-k(t-u)}\mathrm{d}u + \sigma_1\int_0^t e^{-k(t-u)}\mathrm{d}Z_u^1 + \frac{\sigma_2 k}{k-b'}\int_0^t \left(e^{-b'(t-u)} - e^{-k(t-u)}\right)\mathrm{d}Z_u^2. \tag{A.10}$$

Analogy with the G2++ framework The previous expression (A.10) can be related to (A.6): in case where $k > b'$, we can write

$$r_t = r_0 e^{-kt} + \int_0^t \theta(u)e^{-k(t-u)}\mathrm{d}u + \sigma_3\int_0^t e^{-k(t-u)}\mathrm{d}Z_u^3 + \sigma_4\int_0^t e^{-b'(t-u)}\mathrm{d}Z_u^2, \tag{A.11}$$

where we introduced the parameters $\sigma_3 = \sqrt{\sigma_1^2 + \frac{\sigma_2^2 k^2}{(k-b')^2} + 2\rho'\frac{\sigma_1\sigma_2 k}{b'-k}}$, $\sigma_4 = \frac{\sigma_2 k}{k-b'}$ and defined the Brownian motion $Z_t^3 = \frac{1}{\sigma_3}\left(\sigma_1 Z_t^1 - \frac{\sigma_2 k}{k-b'}Z_t^2\right)$. The case $k < b'$ can be treated similarly. Writing the short rate in this form allows to display the link between (A.6) and (A.11). Indeed, by choosing $k = a$, $b' = b$, $\sigma_1 = \sqrt{\sigma^2 + \eta^2 + 2\rho\eta\sigma}$, $\sigma_2 = \frac{1}{k}(k-b')\eta$, $\rho' = \frac{\sigma\rho+\eta}{\sqrt{\sigma^2\eta^2+\sigma\eta\rho}}$, $\theta(t) = \frac{1}{k}\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} + \varphi(t)$ in (A.11), one can write

$$r_t = \varphi(t) + \sigma\int_0^t e^{-k(t-u)}\mathrm{d}Z_u^3 + \eta\int_0^t e^{-b'(t-u)}\mathrm{d}Z_u^2,$$

which is exactly (A.6) with the proper correlation structure between the two risks factors. Conversely, setting $a = k$, $b = b'$, $\sigma = \sigma_3$, $\eta = \sigma_4$, $\rho = \frac{\sigma_1\rho'-\sigma_4}{\sigma_3}$ and $\varphi(t) = r_0 e^{-kt} + \int_0^t \theta(u)e^{-k(t-u)}\mathrm{d}u$ in (A.6) and one recovers (A.11). Properties of the two-factor Hull-White model are thus deduced from that of the G2++.

A.1.4 CIR2++

Another two-factor model encountered among insurers is the CIR2++: it models the short rate process as the sum of two independent square-root processes (also named Cox-Ingersoll-Ross processes) adjusted by a deterministic level.

Namely, under the risk-neutral measure, the short rate writes

$$r_t = \varphi(t) + x_t + y_t,\ t \leq T, \tag{A.12}$$

where

$$\begin{aligned} \mathrm{d}x_t &= k_1(\theta_1 - x_t)\mathrm{d}t + \epsilon_1\sqrt{x_t}\mathrm{d}Z_t^1, \\ \mathrm{d}y_t &= k_2(\theta_2 - y_t)\mathrm{d}t + \epsilon_2\sqrt{y_t}\mathrm{d}Z_t^2, \end{aligned}$$

where Z^1 and Z^2 are two independent Brownian motions, $k_1, k_2, \theta_1, \theta_2, \epsilon_1, \epsilon_2$ are positive constants satisfying the Feller's condition, that is $2k_i\theta_i \geq \epsilon_i^2$ for $i = 1, 2$ and φ is a deterministic

function built to replicate initial ZC bonds curve.

Negative rates The two square-processes are positive under the Feller's condition, and always non-negative. The ability of (A.12) to produce non-negative rates depends on the shape of the deterministic function φ. Indeed,

$$\mathbb{P}^*(r_t \leq 0) = \mathbb{P}^*\Big(x_t + y_t \leq -\varphi(t)\Big),$$

which is a positive quantity as long as $\varphi(t) \leq 0$.

ZC bonds pricing The Laplace function of integrated processes $\int_0^\cdot x_s \mathrm{d}s$ and $\int_0^\cdot y_s \mathrm{d}s$ can be derived in closed-form formula using Itô's lemma to deduce a PDE satisfied of which it is solution: we refer to Exercise 1.2.7 in [Alf15]. $(x_t)_{t\leq T}$ and $(y_t)_{t\leq T}$ being independent, the Laplace function of the integrated process $\int_0^\cdot (x_u + y_u)\mathrm{d}u$ writes as the product of the Laplace function of the respective integrated processes $\int_0^\cdot x_s \mathrm{d}s$ and $\int_0^\cdot y_s \mathrm{d}s$. The ZC price under (A.12) can be obtained in closed-form formula as

$$P(t,T) = \exp\left(-\int_t^T \varphi(s)\mathrm{d}s\right) \times A_1(t,T)e^{-B_1(t,T)x_t} \times A_2(t,T)e^{-B_2(t,T)y_t}, \tag{A.13}$$

where for $i = 1, 2$,

$$\begin{aligned} A_i(t,T) &= \left(\frac{2h_i \exp\big((T-t)(k_i+h_i)/2\big)}{2h_i + (k_i+h_i)[\exp\big((T-t)h_i\big) - 1]}\right)^{2k_i\theta_i/\epsilon_i^2}, \\ B_i(t,T) &= \frac{2[\exp\big((T-t)h_i\big) - 1]}{2h_i + (k_i+h_i)[\exp\big((T-t)h_i\big) - 1]}, \\ h_i &= \sqrt{k_i^2 + 2\epsilon_i^2}. \end{aligned}$$

Fitting the initial term-structure Similarly as for the previous models, the CIR2++ allows to exactly replicate the initial term-structure of ZC bonds: the deterministic function φ is built to that extent.

Proposition A.4. *Setting, for $t \leq T$,*

$$\varphi(t) = -\frac{\partial \ln P^{Mkt}(0,t)}{\partial T} - x_0\frac{\partial B_1(0,t)}{\partial T} - y_0\frac{\partial B_2(0,t)}{\partial T} + \frac{\partial \ln A_1(0,t)}{\partial T} + \frac{\partial \ln A_2(0,t)}{\partial T},$$

allows to replicate the spot market prices, that is:

$$P(0,t) = P^{Mkt}(0,t), \;\; t \leq T.$$

Proof. Using (A.13) and imposing $\ln P(0,t) = \ln P^{Mkt}(0,t)$ yields the result. □

A.2 Pricing of interest rates derivatives

A.2.1 Pricing measures and hedging

In the previous section, all the introduced basic interest-rates derive from the fundamental assets that are the Zero-Coupon bonds. In this paragraph, to set some ideas, we are interested in the problem of hedging and pricing in a market rates in which risky assets are ZC bonds

of different maturities. The setup is the following: consider a finite time horizon $T > 0$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a right-continuous filtration $(\mathcal{F}_t)_{0\leq t\leq T}$. The market we are modelling is composed of risk-free asset B whose time-t value is

$$B(t) = \exp\left(\int_0^t r_s \mathrm{d}s\right),$$

in accordance with (1.19). $N \in \mathbb{N}^*$ Zero-Coupon bonds (risky assets) of different maturities are (continuously) traded on the market. Their prices are represented by adapted semi-martingales and are denoted by $\Big(\big(P(t,T_k)\big)_{t\leq T}\Big)_{k\in\{1,\dots,N\}}$. The discounted prices are denoted by $\tilde{P}(t,T_k) := P(t,T_k)/B(t)$ for all $t \in [0,T_k]$ and all $k \in \{1,\dots,N\}$. We first wonder how can the No-Arbitrage Assumption (NAA) be characterized in this market.

Theorem A.5 ([Bjö09])**.** *NAA holds if, and only if, there exists a probability measure $\mathbb{P}^*$, equivalent to $\mathbb{P}$, such that* ***all discounted ZC prices are martingales****, i.e. for all $k \in \{1,\dots,N\}$, $\Big(\tilde{P}(t,T_k)\Big)_{0\leq t\leq T}$ is a $\mathbb{P}^*$-martingale.*

Proof. This a particular case of a the general Theorem 10.14 stated in [Bjö09]. □

This result that is common to all arbitrage free market, whether or not complete, means that there exists a market price of risk that is **common to all risky assets**.
As a positive semi-martingale, the dynamics of the discounted ZC bond prices can be written in the following form, for all $k \in \{1,\dots,N\}$,

$$\mathrm{d}\tilde{P}(t,T_k) = \tilde{P}(t,T_k)\Big(\big(b(t,T_k) - r_t\big)\mathrm{d}t + \boldsymbol{\sigma}(t,T_k)\cdot \mathrm{d}\boldsymbol{W}_t\Big),$$

for some multidimensional Brownian motion $\big(\boldsymbol{W}_t\big)_{0\leq t\leq T}$ under $\mathbb{P}$, and some adapted measurable processes $\big(b(t,T_k)\big)_{0\leq t\leq T}$ and $\big(\boldsymbol{\sigma}(t,T_k)\big)_{0\leq t\leq T}$. Girsanov's theorem yields first that the Radon-Nikodym derivative $L_t := \frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}_{|\mathcal{F}_t}$ can be written as a stochastic exponential: $L_t = \exp\Big(\int_0^t \boldsymbol{\lambda}(s)\cdot \mathrm{d}\boldsymbol{W}_s - \frac{1}{2}\int_0^t \|\boldsymbol{\lambda}(s)\|^2\mathrm{d}s\Big)$ for some multidimensional adapted process $\big(\boldsymbol{\lambda}(t)\big)_{0\leq t\leq T}$; secondly, the theorem states that for any $k \in \{1,\dots,N\}$, the dynamics under $\mathbb{P}^*$ of discounted ZC bond of maturity T_k writes

$$\mathrm{d}\tilde{P}(t,T_k) = \tilde{P}(t,T_k)\Big((b(t,T_k) - r_t - \boldsymbol{\sigma}(t,T_k)\cdot\boldsymbol{\lambda}(t))\mathrm{d}t + \boldsymbol{\sigma}(t,T_k)\cdot \mathrm{d}\boldsymbol{W}_t^*\Big).$$

with $\big(\boldsymbol{W}_t\big)_{0\leq t\leq T}$ a multidimensional Brownian motion under $\mathbb{P}^*$. The process $\big(\boldsymbol{\lambda}(t)\big)_{0\leq t\leq T}$ is the mentioned market price of risk, also referred to as the *risk premium process* and should satisfy, according to Theorem A.5, for all $k \in \{1,\dots,N\}$ and all $t \leq T_k$:

$$b(t,T_k) - r_t = \boldsymbol{\sigma}(t,T_k)\cdot\boldsymbol{\lambda}(t). \tag{A.14}$$

The fact that the process $\boldsymbol{\lambda}$ is defined up to the time horizon T and does not depend on the maturities $(T_k)_{k\in\{1,\dots,N\}}$ is fundamental: it accounts for the consistency between market ZC prices of different maturities allowing to satisfy the arbitrage-free requirement.

Remark 36. *Fact that $\boldsymbol{\lambda}$ does not depend on $T_k \in \{1,\dots,N\}$ can be recovered using the term structure equation, as depicted in Section 23.2 of [Bjö09].*

We are now able to derive *the* price of derivatives traded on the market. Let us consider a contract delivering a payoff H at maturity time T. H is assumed to be an *attainable* contingent

claim in the sense that the existence of an *admissible*[1] trading strategy whose value at time T is H is assumed. Let us consider such a strategy and denote by $(V_t(H))_{t\leq T}$ the time-t value of the associated portfolio. It is built using the available assets on the market, that are the ZC bonds $\Big(P(t,T_k)\Big)_{t\leq T, k\in\{1,\dots,N\}}$ along with the risk-free asset B so that $V_t(H) = \phi_t^0 B(t) + \sum_{k=1}^N \phi_t^k P(t,T_k)$ where the processes $(\phi_t^0, \phi_t^k)_{t\leq T, k\in\{1,\dots,N\}}$ represent the position of the considered strategy in each asset. First, as an admissible strategy, we get that $V_t(H) \geq 0$ at any date t and that the strategy is *self-financing* meaning

$$\mathrm{d}V_t(H) = \phi_t^0 \mathrm{d}B(t) + \sum_{k=1}^N \phi_t^k \mathrm{d}P(t,T_k).$$

It is common to work with the discounted value of the portfolio that should satisfy under $\mathbb{P}^*$:

$$\mathrm{d}\tilde{V}_t(H) = \sum_{k=1}^N \phi_t^k \mathrm{d}\tilde{P}(t,T_k) = \sum_{k=1}^N \phi_t^k \tilde{P}(t,T_k)\boldsymbol{\sigma}(t,T_k)\cdot \mathrm{d}\boldsymbol{W}_t^*. \tag{A.15}$$

Under technical assumptions, we claim that (A.15) is a $\mathbb{P}^*$-martingale, whose final value is $\tilde{H} = H/B(T)$. Thus, at any time $t \leq T$, we get

$$\tilde{V}_t(H) = \mathbb{E}^*\left[\tilde{V}_T(H)|\mathcal{F}_t\right] = \mathbb{E}^*\left[\tilde{H}|\mathcal{F}_t\right],$$

meaning that **all discounted price processes are $\mathbb{P}^*$-martingales**. Since the bank account process $B(t)$ is an $\mathcal{F}_t$-measurable process, we finally get

$$V_t(H) = \mathbb{E}^*\left[D(t,T)H|\mathcal{F}_t\right]. \tag{A.16}$$

The price of the derivative of final payoff H is given by (A.16): it is the price one should disburse at time t to obtain the final wealth H at T. When $H = 1$, we obtain the price of a Zero-Coupon bond of maturity T:

$$P(t,T) = \mathbb{E}^*\left[D(t,T)\big|\mathcal{F}_t\right] = \mathbb{E}^*\left[\exp\left(-\int_t^T r_s \mathrm{d}s\right)\Big|\mathcal{F}_t\right]. \tag{A.17}$$

A non trivial application of the martingale representation theorem allows to get the existence of a process $(\boldsymbol{M}_t)_{0\leq t\leq T}$ such that $\int_0^T \|\boldsymbol{M}_s\|^2 \mathrm{d}s < \infty$ almost surely and

$$\tilde{V}_t(H) = \tilde{V}_0(H) + \int_0^T \boldsymbol{M}_s \cdot \mathrm{d}\boldsymbol{W}_s^*.$$

A possible hedging strategy consists then in choosing the weights ϕ^k so that

$$\boldsymbol{M}_t^i = \sum_{k=1}^N \phi_t^k \tilde{P}(t,T_k)\boldsymbol{\sigma}^i(t,T_k), t\leq T, 1\leq i\leq d, \text{ and } \phi_t^0 = \mathbb{E}^*\left[D(t,T)H|\mathcal{F}_t\right] - \sum_{k=1}^N \phi_t^k \tilde{P}(t,T_k).$$

To ensure the existence of an admissible strategy, observe that $d \leq N$ is required.

Remark 37. *It is common to use a multidimensional Brownian motion $(\boldsymbol{W}_t^*)_{0\leq t\leq T}$ to model*

[1] An admissible strategy is a self-financing strategy associated with a portfolio whose value is always non-negative. Some technical requirements are sometimes added to the definition, see for instance Definition 4.3.1 in [LL11].

the market hazard in interest rates market to account for the different time horizon involved, as we will discuss it in the next Section 1.5.2.

Remark 38. *All vector processes* $\big(\boldsymbol{\lambda}(t)\big)_{0\leq t\leq T}$ *satisfying (A.14) allow to define a risk-neutral probability measure as* $\mathrm{d}\mathbb{P}^{\boldsymbol{\lambda}} = L_T \mathrm{d}\mathbb{P}$. *All these probability measures can be used for pricing of ZC bonds derivatives.*

A.2.2 Change of *numéraire*

When the short rate is stochastic, so is the discount factor and pricing of interest-rates derivatives requires to be able to derive the distribution of the pair (r_t, H) according to the pricing formula (A.16). To get ride of this difficulty, it is useful to work under alternative probability measure equivalent to risk-neutral one $\mathbb{P}^*$.

A *numéraire* is defined to be any positive and non-dividend-paying asset; it is used to normalize all other assets. Under the risk-neutral measure, we work with discounted assets: the *numéraire* naturally used here is the bank account B. Let us consider an alternative *numéraire* whose value at time t is denoted by $U(t)$. We can consider the probability measure $\mathbb{P}^U$ whose Radon-Nikodym density with respect to $\mathbb{P}^*$ over $\mathcal{F}_t$ is given by

$$\frac{\mathrm{d}\mathbb{P}^U}{\mathrm{d}\mathbb{P}^*}\Big|_{\mathcal{F}_t} = \frac{U(t)B(0)}{U(0)B(t)}.$$

U being a price process, the discounted price process $\Big(U(t)/B(t)\Big)_{t\leq T}$ is a $\mathbb{P}^*$-martingale and it is straightforward to check that $\frac{U(t)B(0)}{U(0)B(t)}$ is non-negative and that $\mathbb{E}^*\left[\frac{U(t)B(0)}{U(0)B(t)}\right] = 1$. Let π_t be the price process of a tradable asset. From the martingale property of discounted prices under the risk-neutral measure, we know that

$$\frac{\pi_t}{B(t)} = \mathbb{E}^*\left[\frac{\pi_T}{B(T)}\Big|\mathcal{F}_t\right].$$

This is equivalent to say that

$$\begin{aligned}\frac{\pi_t}{U(t)} &= \mathbb{E}^*\left[\frac{B(t)U(T)}{B(T)U(t)} \times \frac{\pi_T}{U(T)}\Big|\mathcal{F}_t\right] \\ &= \frac{1}{\mathbb{E}^*\left[\frac{U(T)B(0)}{B(T)U(0)}\Big|\mathcal{F}_t\right]}\mathbb{E}^*\left[\frac{U(T)B(0)}{B(T)U(0)} \times \frac{\pi_T}{U(T)}\Big|\mathcal{F}_t\right] \\ &= \mathbb{E}^U\left[\frac{\pi_T}{U(T)}\Big|\mathcal{F}_t\right].\end{aligned} \tag{A.18}$$

It is an important property: **the price process of any asset divided by a given *numéraire* is a martingale under the probability measure associated with this *numéraire***. This is crucial for some pricing purpose.

Particular choices of *numéraire*

We introduced in (1.25) the spot forward rate. Following the preceding discussion, it is natural to consider the probability measure under which the forward rate is a martingale. Let $\theta \leq T$ be a maturity date. The **$\boldsymbol{\theta}$-forward probability measure** is associated to the numéraire

$(P(t,\theta))_{t\leq\theta}$ and is defined by the Radon-Nikodym derivative with respect to $\mathbb{P}^*$:

$$\frac{\mathrm{d}\mathbb{P}^\theta}{\mathrm{d}\mathbb{P}^*}\Big|_{\mathcal{F}_t} = \frac{P(t,\theta)B(0)}{P(0,\theta)B(t)} = \frac{P(t,\theta)}{P(0,\theta)B(t)},\ t\leq\theta. \tag{A.19}$$

Combining (A.18) and (A.16) yields, for any price process $(\pi_t)_{t\leq\theta}$,

$$\pi_t = P(t,\theta)\mathbb{E}^\theta\big[\pi_T\big|\mathcal{F}_t\big] \tag{A.20}$$

since $P(\theta,\theta)=1$. Main advantage of this formula is that it only requires to know the distribution of the price process $(\pi_t)_{t\leq\theta}$ under the forward measure whereas pricing formula under $\mathbb{P}^*$ requires to know the joint distribution of $(\int_0^\cdot r_t\mathrm{d}t, \pi_\cdot)$.

A second probability measure key for us comes from the forward swap rates (1.29). We can take as numéraire the value of the portfolio composed of zero-coupon bonds of different maturities that writes

$$\sum_{i=m}^{n-1}(T_{i+1}-T_i)P(t,T_{i+1}) =: B^S(t). \tag{A.21}$$

B^S is usually named the *annuity* of the associated swap contract. The probability measure $\mathbb{P}^S$ associated to this numéraire is often called **forward-swap measure** and is such that

$$\frac{\mathrm{d}\mathbb{P}^S}{\mathrm{d}\mathbb{P}^*}\Big|_{\mathcal{F}_t} = \frac{B^S(t)B(0)}{B^S(0)B(t)},\ t\leq T_m. \tag{A.22}$$

Appendix B

Detailed results on Monte-Carlo *frozen* volatilities

Table B.1 gathers *freezing* errors for both Rebonato and Hull & White approximations with parameters respectively obtained as outputs of calibration process. It is associated to Figure 2.2. Table B.2 presents results obtained under a fixed set of parameters and is associated to Figure 2.3. Tables B.3 and B.4 present similar quantities in Black environment.

B.1 Bachelier environment

ATM swaptions volatilities (in %) - Rebonato freezing					
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Closed-form
5	5	0.5264	0.5199	0.5330	0.5350
	10	0.5862	0.5794	0.5931	0.6024
	20	0.5654	0.5597	0.5710	0.6024
15	5	0.4292	0.4246	0.4337	0.4454
	10	0.5039	0.4991	0.5086	0.5328
	20	0.4852	0.4813	0.4891	0.5425
25	5	0.5642	0.5596	0.5687	0.5796
	10	0.5276	0.5236	0.5316	0.5445
	20	0.4397	0.4366	0.4427	0.4612
ATM swaptions volatilities (in %) - Hull & White freezing					
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Closed-form
5	5	0.5231	0.5166	0.5295	0.5304
	10	0.5913	0.5844	0.5982	0.6043
	20	0.5917	0.5858	0.5976	0.6089
15	5	0.4303	0.4257	0.4348	0.4460
	10	0.5118	0.5070	0.5166	0.5291
	20	0.5066	0.5026	0.5106	0.5272
25	5	0.5676	0.5632	0.5721	0.5818
	10	0.5346	0.5306	0.5385	0.5477
	20	0.4535	0.4505	0.4566	0.4672

Table B.1: Details on the *freezing* error for both approximations.

ATM swaptions volatilities (in %)						
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Rebonato	Hull & White
5	5	0.5264	0.5199	0.5330	0.5350	0.5339
	10	0.5862	0.5794	0.5931	0.6024	0.5988
	20	0.5654	0.5597	0.5710	0.6024	0.5810
15	5	0.4292	0.4246	0.4337	0.4454	0.4442
	10	0.5039	0.4991	0.5086	0.5328	0.5200
	20	0.4852	0.4813	0.4891	0.5425	0.5039
25	5	0.5642	0.5596	0.5687	0.5796	0.5780
	10	0.5276	0.5236	0.5316	0.5445	0.5401
	20	0.4397	0.4366	0.4427	0.4612	0.4525

Table B.2: *Freezing* error under a fixed set of parameters.

B.2 Black environment

ATM swaptions volatilities (in %) - Rebonato freezing					
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Closed-form
5	5	0.5274	0.5223	0.5326	0.5255
	10	0.5937	0.5881	0.5993	0.5952
	20	0.5835	0.5785	0.5886	0.6067
15	5	0.5060	0.5015	0.5104	0.5083
	10	0.5399	0.5355	0.5443	0.5549
	20	0.5035	0.4998	0.5071	0.5473
25	5	0.5552	0.5511	0.5592	0.5597
	10	0.5310	0.5273	0.5347	0.5392
	20	0.4596	0.4566	0.4625	0.4752
ATM swaptions volatilities (in %) - Hull & White freezing					
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Closed-form
5	5	0.5243	0.5192	0.5295	0.5213
	10	0.5999	0.5942	0.6056	0.5979
	20	0.6120	0.6067	0.6172	0.6135
15	5	0.5106	0.5061	0.5151	0.5119
	10	0.5515	0.5470	0.5560	0.5538
	20	0.5280	0.5242	0.5318	0.5349
25	5	0.5592	0.5552	0.5631	0.5623
	10	0.5379	0.5343	0.5414	0.5418
	20	0.4730	0.4701	0.4760	0.4801

Table B.3: Details on the *freezing* error for both approximations.

ATM swaptions volatilities (in %)						
Maturity	Tenor	Monte-Carlo	M.-C. inf.	M.-C. sup.	Rebonato	Hull & White
5	5	0.5274	0.5223	0.5326	0.5255	0.5076
	10	0.5937	0.5881	0.5993	0.5952	0.5826
	20	0.5835	0.5785	0.5886	0.6067	0.6057
15	5	0.5060	0.5015	0.5104	0.5083	0.4868
	10	0.5399	0.5355	0.5443	0.5549	0.5299
	20	0.5035	0.4998	0.5071	0.5473	0.5195
25	5	0.5552	0.5511	0.5592	0.5597	0.5618
	10	0.5310	0.5273	0.5347	0.5392	0.5491
	20	0.4596	0.4566	0.4625	0.4752	0.4969

Table B.4: *Freezing* error under a fixed set of parameters.

Appendix C

About density function of Itô's integrals

Lemma C.1. *Let $(B_t)_{t\geq 0}$ be a Brownian motion. Let us denote by $\mathcal{F}_t^B = \sigma(B_s, s \leq t)$ the filtration associated with B. Let $\varphi := (\varphi_t)_{t\geq 0}$ be a stochastic process adapted to $\mathcal{F}^\varphi := (\mathcal{F}_t^\varphi)_{t\geq 0}$, independent of $\mathcal{F}_T^B$ and such as $\mathbb{E}\big[\int_0^T \varphi_s^2 \mathrm{d}s\big] < \infty$. Let $x_0 \in \mathbb{R}$. Let*

$$X := x_0 + \int_0^T \varphi_s \mathrm{d}B_s$$

and consider f a continuous bounded function. Then

$$\mathbb{E}[f(X)|\mathcal{F}_T^B] = \int_{\mathbb{R}} f(x)\mathbb{E}\Big[\frac{1}{\sqrt{2\pi\Sigma(T)^2}} \exp\Big(-\frac{(x-x_0)^2}{2\Sigma(T)^2}\Big)\Big]\mathrm{d}x$$

where $\Sigma(T)^2 := \int_0^T \varphi_s^2 \mathrm{d}s$.

Proof. Consider first the case when $\mathbb{P}(\Sigma(T)^2 > 0) = 1$. Since $\varphi \in \mathbb{H}^2 = \mathbb{L}^2\big(\mathrm{Prog}([0,T]\times\Omega), \mathrm{d}t \times \mathrm{d}\mathbb{P}\big)$, there exists a sequence of elementary processes $(\varphi^n)_{n\in\mathbb{N}}$ belonging to $(\mathbb{H}^2)^{\mathbb{N}}$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_0^T (\varphi_s^n - \varphi_s)^2 \mathrm{d}s\Big] = 0.$$

By definition, for each $n \in \mathbb{N}$, there exists $p_n \in \mathbb{N}$, $0 = t_0 < t_1 < \cdots < t_{p_n} \leq T$ and $(\varphi_i^n)_{0\leq i\leq p_n}$ that are square integrable, independent of $\mathcal{F}^B$ and being such that

$$\forall t \in [0,T],\ \varphi_t^n = \sum_{i=0}^{p_n-1} \varphi_i^n \mathbb{1}_{]t_i,t_{i+1}]}(t).$$

where φ_i^n is $\mathcal{F}_{t_i}^\varphi$-measurable for each i. For each $n \in \mathbb{N}$,

$$\int_0^T \varphi_s^n \mathrm{d}B_s = \sum_{i=0}^{p_n-1} \varphi_i^n (B_{t_{i+1}} - B_{t_i})$$

writes as the sum of products of independent random variables. Consequently,

$$\mathbb{E}\Big[f\Big(x_0 + \int_0^T \varphi_s^n \mathrm{d}B_s\Big)|\mathcal{F}^\varphi\Big] = \psi\big(\varphi_0^n, \ldots, \varphi_{p_n-1}^n\big), \text{ almost surely,}$$

for all $n \in \mathbb{N}$ where we set $\psi(y_0, \dots, y_{p_n-1}) = \mathbb{E}\Big[f\Big(x_0 + \sqrt{\sum_{i=0}^{p_n-1} y_i^2} \times G\Big)\Big]$ with G a standardized Gaussian variate independent of $\mathcal{F}^\varphi$. And thus

$$\mathbb{E}\Big[f\Big(x_0 + \int_0^T \varphi_s^n \mathrm{d}B_s\Big)|\mathcal{F}^\varphi\Big] \overset{a.s.}{=} \int_{\mathbb{R}} f(x) \frac{e^{-(x-x_0)^2/2\Sigma(T,n)^2}}{\sqrt{2\pi\Sigma(T,n)^2}} \mathrm{d}x \tag{C.1}$$

where $\Sigma(T,n)^2 = \sum_{i=0}^{p_n-1} (\varphi_i^n)^2 = \int_0^T (\varphi_s^n)^2 \mathrm{d}s$. Furthermore, recall that the sequence of processes $(\varphi^n)_{n\in\mathbb{N}}$ converges in the Hilbert space $\mathbb{H}^2$; in particular,

$$\|\varphi^n - \varphi\|_{\mathbb{H}^2}^2 = \mathbb{E}\int_0^T (\varphi_s^n - \varphi_s)^2 \mathrm{d}s \xrightarrow[n\to+\infty]{} 0.$$

Cauchy-Schwarz inequality twice and Minkowski one then imply

$$\mathbb{E}\Big[\int_0^T \big((\varphi_s^n)^2 - \varphi_s^2\big)\,\mathrm{d}s\Big] \le \|\varphi^n - \varphi\|_{\mathbb{H}^2}\|\varphi^n + \varphi\|_{\mathbb{H}^2} \le \|\varphi^n - \varphi\|_{\mathbb{H}^2}\big(\|\varphi^n\|_{\mathbb{H}^2} + \|\varphi\|_{\mathbb{H}^2}\big)$$

hence the $\mathbb{L}^1$ convergence of the sequence of integrated processes $\Big(\int_0^T (\varphi_s^n)^2 \mathrm{d}s\Big)_{n\in\mathbb{N}}$. One can extract a subsequence $\Big(\int_0^T (\varphi_s^{n_k})^2 \mathrm{d}s\Big)_{k\in\mathbb{N}}$ converging almost surely. In particular,

$$\Sigma(T,n_k)^2 := \int_0^T (\varphi_s^{n_k})^2 \mathrm{d}s \xrightarrow[k\to\infty]{\text{a.s.}} \int_0^T \varphi_s^2 \mathrm{d}s = \Sigma(T)^2.$$

Yet, since we have assumed the almost sure positivity of $\Sigma(T)$, we can assumed that for k large enough, $\Sigma(T,n_k)$ is also almost surely positive: there exists $K \in \mathbb{N}$ such that for all $k \ge K$, $\mathbb{P}\big(\Sigma(T,n_k)^2 > 0\big) = 1$. For such k, a change of variable in the right hand side of (C.1) is possible and yields

$$\mathbb{E}\Big[f\Big(x_0 + \int_0^T \varphi_s^n \mathrm{d}B_s\Big)|\mathcal{F}^\varphi\Big] \overset{a.s.}{=} \int_{\mathbb{R}} f\big(x_0 + \Sigma(T,n_k)u\big) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u.$$

On the one hand, f being continuous, $f\big(x_0 + \Sigma(T,n_k)u\big)\frac{e^{-u^2/2}}{\sqrt{2\pi}} \xrightarrow[k\to\infty]{\text{a.s.}} f(x_0 + \Sigma(T)u)\frac{e^{-u^2/2}}{\sqrt{2\pi}}$ for all $u \in \mathbb{R}$; on the other hand, f being bounded one has for all $u \in \mathbb{R}$, $|f\big(x_0 + \Sigma(T,n_k)u\big)\frac{e^{-u^2/2}}{\sqrt{2\pi}}| \le \|f\|_\infty \frac{e^{-u^2/2}}{\sqrt{2\pi}}$ and Lebesgue convergence theorem implies

$$\int_{\mathbb{R}} f\big(x_0 + \Sigma(T,n_k)u\big) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u \xrightarrow[k\to\infty]{\text{a.s.}} \int_{\mathbb{R}} f\big(x_0 + \Sigma(T)u\big) \frac{e^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u.$$

Moreover, the extracted subsequence $\Big(\int_0^T (\varphi_s^{n_k})^2 \mathrm{d}s\Big)_{k\in\mathbb{N}}$ also converges in $\mathbb{L}^2\big(\mathrm{Prog}([0,T] \times \Omega), \mathrm{d}t \times \mathrm{d}\mathbb{P}\big)$. Itô's isometry implies

$$\mathbb{E}\Big[\Big(\int_0^T \varphi_s^{n_k} \mathrm{d}B_s - \int_0^T \varphi_s \mathrm{d}B_s\Big)^2\Big] = \mathbb{E}\Big[\int_0^T (\varphi_s^{n_k} - \varphi_s)^2 \mathrm{d}s\Big] \xrightarrow[k\to+\infty]{} 0$$

and one can again extract a subsequence $(\varphi^{n_{k,p}})_{p\in\mathbb{N}}$ such that the following almost sure convergence holds:

$$\int_0^T \varphi_s^{n_{n,p}} \mathrm{d}B_s \xrightarrow[p\to\infty]{\text{a.s.}} \int_0^T \varphi_s \mathrm{d}B_s.$$

Then, thanks to the continuity of f, one has

$$f\Big(x_0 + \int_0^T \varphi_s^{n_{n,p}} \mathrm{d}B_s\Big) \xrightarrow[p\to\infty]{\text{a.s.}} f\Big(x_0 + \int_0^T \varphi_s \mathrm{d}B_s\Big).$$

Lebesgue convergence theorem applied to the left hand side in (C.1) provides that

$$\mathbb{E}\left[f\left(x_0 + \int_0^T \varphi_s^{n_{n,p}} \mathrm{d}B_s\right) |\mathcal{F}^{\varphi}\right] \xrightarrow[p\to\mathbb{E}]{} \left[f\left(x_0 + \int_0^T \varphi_s \mathrm{d}B_s\right) |\mathcal{F}^{\varphi}\right].$$

The claim is proved by unicity of the limit.

The case when $\mathbb{P}\left(\Sigma(T)^2 = 0\right) = 1$ is more direct. In that case, $\mathbb{P}\left(\text{for almost every } t \geq 0 : \varphi_t = 0\right) = 1$ implying that $X \overset{a.s.}{=} x_0$ and X is constant equal x_0 (almost surely). The formula given in the statement of the lemma holds in this case by considering the following convergence understood in term of distributions

$$\frac{e^{-(x-x*)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \xrightarrow[\sigma\to 0]{} \delta_{x*}(x)$$

where δ_{x*} is the Dirac measure of atom $x*$. □

Appendix D

Regularity of the price function in Heston model

For the sake of completeness, we provide here an adaptation of the proof of the regularity of the price function used in Theorem 4.11 given in [ET10] in the particular case of the standard Heston model.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with the filtration $(\mathcal{F}_t)_{t\geq 0}$. Consider the stock price (or other financial quantity) process X whose spot value is $X_0 = x$ and dynamics is modelled by

$$\mathrm{d}X_t = \sigma\sqrt{V_t}\mathrm{d}W_t \tag{D.1}$$

where the volatility process is defined as the solution of

$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \epsilon\sqrt{V_t}\mathrm{d}B_t,\ V_0 = v. \tag{D.2}$$

The processes $W_.$ and $B_.$ are two Brownian motions under the measure $\mathbb{P}$ with constant correlation: $\mathrm{d}\langle W_., B_.\rangle_t = \rho\mathrm{d}t$ for some $\rho \in [-1, 1]$. In particular, for all $t > 0$,

$$X_t \overset{a.s.}{=} \sigma \int_0^t \sqrt{V_u}\mathrm{d}W_u. \tag{D.3}$$

Let $T > 0$ be a finite time horizon and let φ be a payoff function such that $\mathbb{E}[\varphi(X_T)] < \infty$. The time-$t$ price of the European option delivering $\varphi(X_T)$ in T is given by $u(t, X_t, V_t) = \mathbb{E}[\varphi(X_T)\big|\mathcal{F}_t]$ where $u(t, x, v) = \mathbb{E}[\varphi(X_T)|(X_t, V_t) = (x, v)]$. The associated partial differential equation is

$$\partial_t f + \mathcal{L}f = 0 \tag{D.4}$$

where we set

$$\mathcal{L}f = \frac{1}{2}v\frac{\partial^2}{\partial x^2}f + \kappa(\theta - v)\frac{\partial}{\partial v}f + \frac{1}{2}\epsilon^2 v\frac{\partial^2}{\partial v^2}f + \rho\epsilon x v\frac{\partial^2}{\partial x\partial v}f.$$

Assumption: The payoff function φ is bounded, twice differentiable on $\mathbb{R}_+$ and is such that $\times \mapsto x\varphi'(x)$ and $x \mapsto x^2\varphi''(x)$ are bounded. The purpose is to show that the price function u satisfies the equation (D.4).

Proposition D.1. $(t, x, v) \mapsto u(t, x, v)$ *is continuous over* $[0, T] \times \left(\mathbb{R}_+\right)^2$.

Proof. Let $(t, x, v) \in [0, T] \times \left(\mathbb{R}_+\right)^2$. Let (τ_n, x_n, v_n) be a sequence converging towards (τ, x, v)

where $\tau = T - t$. Define the sequences of approximating processes as

$$\begin{aligned} \mathrm{d}X_t^n &= \sigma\sqrt{V_t^n}\mathrm{d}W_t \\ \mathrm{d}V_t^n &= \kappa(\theta - V_t^n)\mathrm{d}t + \epsilon\sqrt{V_t^n}\mathrm{d}B_t, \end{aligned}$$

starting from (x, v) for all $n \in \mathbb{N}$. Following [BMO98], one has

$$\mathbb{E}\big[\sup_{0\le t\le \tau+\gamma}(X_t^n - X_t)^2\big] \xrightarrow[n\to+\infty]{} 0$$

where $\gamma > 0$ is such that $\tau_n \le \tau + \gamma$. Consequently, with Jensen's inequality, Burkholder-Davis-Gundy's one and Hölder regularity of the square-root function, we get

$$\begin{aligned} &\Big(\mathbb{E}\Big[\int_0^{\tau_n}\sqrt{V_t^n}\mathrm{d}W_t\Big] - \mathbb{E}\Big[\int_0^{\tau}\sqrt{V_t}\mathrm{d}W_t\Big]\Big)^2 \\ &\quad\le 2\Big(\mathbb{E}\Big[\int_0^{\tau_n}(\sqrt{V_t^n} - \sqrt{V_t})\mathrm{d}W_t\Big]^2 + \mathbb{E}\Big[\int_{\tau_n}^{\tau}\sqrt{V_t}\mathrm{d}W_t\Big]^2\Big) \\ &\quad\le 2\Big(\int_0^{\tau_n}\mathbb{E}\big[|V_t^n - V_t|\big]\mathrm{d}t + \int_{\tau_n}^{\tau}\mathbb{E}\big[V_t\big]\mathrm{d}t\Big]\Big) \xrightarrow[n\to+\infty]{} 0. \end{aligned}$$

Using (D.3), it follows that $(X_{\tau_n}^n)_{n\in\mathbb{N}}$ converges towards X_T in probability. Thus, $\mathbb{E}\big[\varphi(X_{\tau_n})|(X_0^n, V_0^n) = (x_n, v_n)\big] \xrightarrow[n\to\infty]{} \mathbb{E}\big[\varphi(X_\tau)|(X_0, V_0) = (x, v)\big]$ with Lebesgue's theorem since φ is bounded. The result follows from fact that the point (x, v, τ) is arbitrary. □

Proposition D.2. *The option price u is $\mathcal{C}^{1,2,2}\big([0,T)\times\mathbb{R}_+^*\times\mathbb{R}_+^*\big)$ and satisfies $\partial_t u + \mathcal{L}u = 0$ at all points $(t, x, v) \in [0,T)\times\mathbb{R}_+^*\times\mathbb{R}_+^*$.*

Proof. Let $(t, x, v) \in [0,T)\times\mathbb{R}_+^*\times\mathbb{R}_+^*$ and let $\mathcal{D} = (0,T)\times(x_1, x_2)\times(v_1, v_2)$ be a rectangle such that $(t, x, v) \in \mathcal{D}$. Since u is continuous thanks to the previous Proposition D.1, there exists a unique solution $\tilde{u}$ to the boundary value problem

$$\begin{aligned} \partial_t\tilde{u} + \mathcal{L}\tilde{u} &= 0 \text{ for } (t, x, v) \in \mathcal{D}, \\ \tilde{u} &= u \text{ for } (t, x, v) \in \partial_0\mathcal{D} \end{aligned}$$

where $\partial_0\mathcal{D} = \partial\mathcal{D}\setminus\{0\}\times(x_1, x_2)\times(v_1, v_2)$. Thanks to Itô's formula, one gets that the process $Y_t = \tilde{u}(t, X_t, V_t)$ is martingale over $[s, \tau_{\mathcal{D}})$ where

$$\tau_{\mathcal{D}} = \inf\big\{t \ge s:\ (t, X_t, V_t) \ni \mathcal{D}\big\}.$$

Thus,

$$\begin{aligned} \tilde{u}(t, x, v) = Y_t &= \mathbb{E}\big[Y_{T\wedge\tau_{\mathcal{D}}}|(X_t, V_t) = (x, v)\big] \\ &= \mathbb{E}\big[u\big(T\wedge\tau_{\mathcal{D}}, X_{T\wedge\tau_{\mathcal{D}}}, V_{T\wedge\tau_{\mathcal{D}}}\big)|(X_t, V_t) = (x, v)\big] = \tilde{u}(t, x, v) \end{aligned}$$

by the strong Markov property. Finally, $u = \tilde{u}$ over $\mathcal{D}$ and thus $u \in \mathcal{C}^{1,2,2}\big([0,T)\times\mathbb{R}_+^*\times\mathbb{R}_+^*\big)$ and satisfies $\partial_t u + \mathcal{L}u = 0$ on $\mathcal{D}$. Since it true for any (t, x, v), the claim is proved.

□

Appendix E

Gradient vector of swaptions prices in DD-SV-LMM and optimization routines

The vector $\boldsymbol{\chi}$ in Proposition 5.1 writes:

$$\boldsymbol{\chi}(\boldsymbol{\Theta}; z) := [\chi_a(z), \chi_b(z), \chi_c(z), \chi_d(z), \chi_\kappa(z), \chi_\theta(z), \chi_\epsilon(z), \chi_\rho(z)]^T$$

where χ_y denotes the partial derivative of φ with respect to the parameter y. Using that $\chi_y(x) = \frac{\partial}{\partial x}\Psi(x; t, S_0^{m,n}, V_0; \boldsymbol{\Theta})$ and since $\Psi(x; t, S_0^{m,n}, V_0; \boldsymbol{\Theta})$ is defined recursively with terminal value $\Psi(\boldsymbol{\Theta}; X_{m,n}(0), V_0, T_m; z)$, one needs to compute $\frac{\partial \Psi(x;t,S_0^{m,n},V_0;\boldsymbol{\Theta})}{\partial x}$ on each interval $(\tau_j, \tau_{j+1}]$. We will rather give the partial derivatives of Ψ at any time.

E.1 Partial derivatives of characteristic function

E.1.1 Partial derivative of Ψ with respect to θ

Since $A_j(\tau)$ and consequently also $B(\tau, z)$ are independant from θ for all t, the partial derivative of Ψ with respect to θ writes:

$$\frac{\partial \Psi(X_{m,n}(t), V(t), t; z)}{\partial \theta} = \frac{\partial A(\tau, z)}{\partial \theta}\Psi(X_{m,n}(t), V(t), t; z).$$

The partial derivative of $A(\tau, z)$ is given by:

$$\frac{\partial A(\tau, z)}{\partial \theta} = \frac{\partial A(\tau_j, z)}{\partial \theta} - \frac{\kappa\tilde{\rho}\lambda z(\tau - \tau_j)}{\epsilon} + \frac{2\kappa}{\epsilon^2}D_j(\tau)$$

since $D_j(\tau)$ is independant from θ.

E.1.2 Partial derivative of Ψ with respect to κ

The partial derivative of Ψ with respect to κ writes:

$$\frac{\partial \Psi(X_{m,n}(t), V(t), t; z)}{\partial \kappa} = \left[\frac{\partial A(\tau, z)}{\partial \kappa} + V_0\frac{\partial B(\tau, z)}{\partial \kappa}\right]\Psi(X_{m,n}(t), V(t), t; z)$$

with

$$\frac{\partial A(\tau,z)}{\partial\kappa} = \frac{\partial A(\tau_j,z)}{\partial\kappa} - \frac{\theta\tilde{\rho}\lambda z(\tau-\tau_j)}{\epsilon} + \frac{2\theta}{\epsilon^2}D_j(\tau) + \frac{2\kappa\theta}{\epsilon^2}\frac{\partial D_j(\tau)}{\partial\kappa},$$
$$\frac{\partial B(\tau,z)}{\partial\kappa} = \frac{\partial B(\tau_j,z)}{\partial\kappa} - \frac{1}{V_0}\frac{\partial A_j(\tau)}{\partial\kappa},$$

where we set

$$\begin{aligned}
\frac{\partial D_j(\tau)}{\partial\kappa} &= \frac{\mu}{\nu^2} + \frac{1}{2}\left(1-\frac{\mu}{\nu}\right)(\tau-\tau_j) - \frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial\kappa},\\
\frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial\kappa} &= \frac{1}{\nu V_0 A_j^2(\tau)}\left[\mu+\nu\left(1-\epsilon^2\frac{\partial B(\tau_j,z)}{\partial\kappa}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right],\\
&\quad - \frac{1}{E_j(\tau)}(\nu-\mu+\epsilon^2 B(\tau_j,z))\frac{\mu}{\nu}(\tau-\tau_j)e^{-\nu(\tau-\tau_j)},\\
\frac{\partial A_j(\tau)}{\partial\kappa} &= \frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^1(\tau)}{\partial\kappa} - \frac{A_j(\tau)}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^2(\tau)}{\partial\kappa},\\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^1(\tau)}{\partial\kappa} &= \frac{1}{A_j^2(\tau)}\left[2\tanh\frac{\nu(\tau-\tau_j)}{2}\left(\mu\frac{\partial B(\tau_j,z)}{\partial\kappa} - \epsilon^2 B(\tau_j,z))\frac{\partial B(\tau_j,z)}{\partial\kappa} + B(\tau_j,z)\right)\right.\\
&\quad \left. + \frac{\mu(\tau-\tau_j)}{2\nu}\left(B(\tau_j,z)(2\mu-\epsilon^2 B(\tau_j,z)) + \lambda^2(z-z^2)\right)\right],\\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^2(\tau)}{\partial\kappa} &= \frac{1}{V_0 A_j^2(\tau)}\left[\left(1+\frac{\mu(\tau-\tau_j)}{2} - \epsilon^2\frac{\partial B(\tau_j,z)}{\partial\kappa}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right.\\
&\quad \left. + \frac{\mu}{\nu}\left(\frac{\tau-\tau_j}{2}(\mu-\epsilon^2 B(\tau_j,z)) + 1\right)\right].
\end{aligned}$$

Note that in order to make the calculation of the partial derivative of A_j easier, we write A_j as the ratio of $\tilde{A}_j^1$ and $\tilde{A}_j^2$ instead of A_j^1 and A_j^2, where $\tilde{A}_j^1$ and $\tilde{A}_j^2$ are defined as:

$$\tilde{A}_j^1(\tau) = A_j^1(\tau)\cosh\frac{\nu(\tau-\tau_j)}{2},$$
$$\tilde{A}_j^2(\tau) = A_j^2(\tau)\cosh\frac{\nu(\tau-\tau_j)}{2}.$$

This trick will be re-employed for other derivatives.

E.1.3 Partial derivative of Ψ with respect to ϵ

The partial derivative of Ψ with respect to ϵ writes:

$$\frac{\partial\Psi(X_{m,n}(t),V(t),t;z)}{\partial\epsilon} = \left[\frac{\partial A(\tau,z)}{\partial\epsilon} + V_0\frac{\partial B(\tau,z)}{\partial\epsilon}\right]\Psi(X_{m,n}(t),V(t),t;z)$$

with

$$\frac{\partial A(\tau,z)}{\partial\epsilon} = \frac{\partial A(\tau_j,z)}{\partial\epsilon} + \frac{\kappa\theta\tilde{\rho}\lambda z(\tau-\tau_j)}{\epsilon^2} - \frac{4\kappa\theta}{\epsilon^3}D_j(\tau) + \frac{2\kappa\theta}{\epsilon^2}\frac{\partial D_j(\tau)}{\partial\epsilon},$$
$$\frac{\partial B(\tau,z)}{\partial\epsilon} = \frac{B(\tau_j,z)}{\epsilon} - \frac{1}{V_0}\frac{\partial A_j(\tau)}{\partial\epsilon},$$

where

$$
\begin{aligned}
\frac{\partial \xi}{\partial \epsilon} &= \frac{\xi - 1}{\epsilon}, \qquad \frac{\partial \mu}{\partial \epsilon} = \frac{\mu - \kappa}{\epsilon}, \qquad \frac{\partial \nu}{\partial \epsilon} = \frac{\nu^2 - \kappa\mu}{\epsilon\nu}, \\
\frac{\partial D_j(\tau)}{\partial \epsilon} &= \frac{1}{\nu}\frac{\partial \nu}{\partial \epsilon} + \frac{1}{2}\left(\kappa\frac{\partial \xi}{\partial \epsilon} - \frac{\partial \nu}{\partial \epsilon}\right)(\tau - \tau_j) - \frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial \epsilon}, \\
\frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial \epsilon} &= \frac{1}{V_0 A_j^2(\tau)}\left[\frac{\partial \nu}{\partial \epsilon} + \left(\frac{\partial \mu}{\partial \epsilon} - 2\epsilon B(\tau_j, z) - \epsilon^2\frac{\partial B(\tau_j, z)}{\partial \epsilon}\right)\tanh\frac{\nu(\tau - \tau_j)}{2}\right] \\
&\quad - \frac{1}{E_j(\tau)}(\nu - \mu + \epsilon^2 B(\tau_j, z))\frac{\partial \nu}{\partial \epsilon}(\tau - \tau_j)e^{-\nu(\tau - \tau_j)}, \\
\frac{\partial A_j(\tau)}{\partial \epsilon} &= \frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^1(\tau)}{\partial \epsilon} - \frac{A_j(\tau)}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^2(\tau)}{\partial \epsilon}, \\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^1(\tau)}{\partial \epsilon} &= \frac{1}{A_j^2(\tau)}\left[2\tanh\frac{\nu(\tau - \tau_j)}{2}\left(\mu\frac{\partial B(\tau_j, z)}{\partial \epsilon} - \epsilon^2 B(\tau_j, z)\frac{\partial B(\tau_j, z)}{\partial \epsilon} + \frac{\partial \mu}{\partial \epsilon}B(\tau_j, z)\right.\right. \\
&\quad \left.\left. - \epsilon(B(\tau_j, z))^2\right) + \frac{\partial \nu}{\partial \epsilon}\frac{\tau - \tau_j}{2}\left(B(\tau_j, z)(2\mu - \epsilon^2 B(\tau_j, z)) + \lambda^2(z - z^2)\right)\right], \\
\frac{1}{\tilde{A}_j^2}\frac{\partial \tilde{A}_j^2}{\partial \epsilon} &= \frac{1}{V_0 A_j^2}\left[\frac{\partial \nu}{\partial \epsilon}\left(1 + (\mu - \epsilon^2 B(\tau_j, z))\frac{\tau - \tau_j}{2}\right)\right. \\
&\quad \left. + \tanh\left(\frac{\nu(\tau - \tau_j)}{2}\right)\left(\nu\frac{\partial \nu}{\partial \epsilon}\frac{\tau - \tau_j}{2} + \frac{\partial \mu}{\partial \epsilon} - 2\epsilon B(\tau_j, z) - \epsilon^2\frac{\partial B(\tau_j, z)}{\partial \epsilon}\right)\right].
\end{aligned}
$$

E.1.4 Partial derivative of Ψ with respect to ρ

The partial derivative of Ψ with respect to ρ writes:

$$
\frac{\partial \Psi(X_{m,n}(t), V(t), t; z)}{\partial \rho} = \left[\frac{\partial A(\tau, z)}{\partial \rho} + V_0\frac{\partial B(\tau, z)}{\partial \rho}\right]\Psi(X_{m,n}(t), V(t), t; z)
$$

with

$$
\begin{aligned}
\frac{\partial A(\tau, z)}{\partial \rho} &= \frac{\partial A(\tau_j, z)}{\partial \rho} - \frac{\kappa\theta\lambda\tilde{\rho}z(\tau - \tau_j)}{\epsilon\rho} + \frac{2\kappa\theta}{\epsilon^2}\frac{\partial D_j(\tau)}{\partial \rho}, \\
\frac{\partial B(\tau, z)}{\partial \rho} &= \frac{\partial B(\tau_j, z)}{\partial \rho} - \frac{1}{V_0}\frac{\partial A_j(\tau)}{\partial \rho},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \mu}{\partial \rho} &= \frac{\mu-\kappa}{\rho}, \frac{\partial \nu}{\partial \rho} = \frac{\mu}{\nu}\frac{\partial \mu}{\partial \rho}, \\
\frac{\partial D_j(\tau)}{\partial \rho} &= \frac{1}{\nu}\frac{\partial \nu}{\partial \rho} + \frac{1}{2}\left(\kappa\frac{\xi-1}{\rho} - \frac{\partial \nu}{\partial \rho}\right)(\tau-\tau_j) - \frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial \rho}, \\
\frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial \rho} &= \frac{1}{V_0 A_j^2(\tau)}\left[\frac{\partial \nu}{\partial \rho} + \left(\frac{\partial \mu}{\partial \rho} - \epsilon^2\frac{\partial B(\tau_j,z)}{\partial \rho}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right] \\
&\quad - \frac{1}{E_j(\tau)}\frac{\partial \nu}{\partial \rho}(\tau-\tau_j)(\nu-\mu+\epsilon^2 B(\tau_j,z))e^{-\nu(\tau-\tau_j)}, \\
\frac{\partial A_j(\tau)}{\partial \rho} &= \frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^1(\tau)}{\partial \rho} - \frac{A_j(\tau)}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^2(\tau)}{\partial \rho}, \\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^1(\tau)}{\partial \rho} &= \frac{1}{A_j^2(\tau)}\left[2\tanh\frac{\nu(\tau-\tau_j)}{2}\left(\mu\frac{\partial B(\tau_j,z)}{\partial \rho} - \epsilon^2 B(\tau_j,z)\frac{\partial B(\tau_j,z)}{\partial \rho} + B(\tau_j,z)\frac{\partial \mu}{\partial \rho}\right)\right. \\
&\quad \left. + \frac{\partial \nu}{\partial \rho}\frac{\tau-\tau_j}{2}\left(B(\tau_j,z)(2\mu-\epsilon^2 B(\tau_j,z)) + \lambda^2(z-z^2)\right)\right], \\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial \tilde{A}_j^2(\tau)}{\partial \rho} &= \frac{1}{V_0 A_j^2(\tau)}\left[\left(1+(\mu-\epsilon^2 B(\tau_j,z))\frac{\tau-\tau_j}{2}\right)\frac{\partial \nu}{\partial \rho}\right. \\
&\quad \left. + \left(\nu\frac{\partial \nu}{\partial \rho}\frac{t-T_j}{2} + \frac{\partial \mu}{\partial \rho} - \epsilon^2\frac{\partial B(\tau_j,z)}{\partial \rho}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right].
\end{aligned}
$$

E.1.5 Partial derivatives of Ψ with respect to a, b, c and d

One can observe that only $\gamma_k(\tau)$ depends on the parameters a, b, c et d, which means that the derivatives of the characteristic function with respect to these parameters are close from each other. Consequently, we group the four partial derivatives in this section. Let $x \in \{a,b,c,d\}$, the partial derivative of Ψ with respect to x writes

$$
\frac{\partial \Psi(X_{m,n}(t),V(t),t;z)}{\partial x} = \left[\frac{\partial A(\tau,z)}{\partial x} + V_0\frac{\partial B(\tau,z)}{\partial x}\right]\Psi(X_{m,n}(t),V(t),t;z)
$$

with

$$
\begin{aligned}
\frac{\partial A(\tau,z)}{\partial x} &= \frac{\partial A(\tau_j,z)}{\partial x} - \frac{\kappa\theta z(\tau-\tau_j)}{\epsilon}\frac{\partial(\tilde{\rho}\lambda)}{\partial x} + \frac{2\kappa\theta}{\epsilon^2}\frac{\partial D_j(\tau)}{\partial x}, \\
\frac{\partial B(\tau,z)}{\partial x} &= \frac{\partial B(\tau_j,z)}{\partial x} - \frac{1}{V_0}\frac{\partial A_j(\tau)}{\partial x},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial(\tilde{\rho}\lambda)}{\partial x} &= \sum_{k=m}^{n-1} w_k(0)\frac{\partial\|\gamma_k(\tau)\|}{\partial x}\rho_k(\tau),\\
\frac{\partial\xi}{\partial x} &= \frac{\epsilon}{\kappa}\sum_{k=m}^{n-1}\alpha_k(0)\sum_{l=m(\tau)}^{k}\frac{\Delta T_l(F_l(0)+\delta)\rho_l(\tau)\frac{\partial\|\gamma_l(\tau)\|}{\partial x}}{1+\Delta T_l F_l(0)},\\
\frac{\partial\mu}{\partial x} &= \kappa\frac{\partial\xi}{\partial x} - \epsilon z\frac{\partial(\tilde{\rho}\lambda)}{\partial x},\\
\frac{\partial\lambda^2}{\partial x} &= 2\left\langle \sum_{k=m}^{n-1} w_k(0)\frac{\partial\gamma_k(\tau)}{\partial x}, \sum_{k=m}^{n-1} w_k(0)\gamma_k(\tau)\right\rangle,\\
\frac{\partial\nu}{\partial x} &= \frac{1}{\nu}\left(\frac{\partial\mu}{\partial x}\mu + \frac{1}{2}\frac{\partial\lambda^2}{\partial x}\epsilon^2(z-z^2)\right),\\
\frac{\partial D_j(\tau)}{\partial x} &= \frac{1}{\nu}\frac{\partial\nu}{\partial x} + \frac{1}{2}\left(\kappa\frac{\partial\xi}{\partial x} - \frac{\partial\nu}{\partial x}\right)(\tau-\tau_j) - \frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial x},\\
\frac{1}{E_j(\tau)}\frac{\partial E_j(\tau)}{\partial x} &= \frac{1}{V_0 A_j^2(\tau)}\left[\frac{\partial\nu}{\partial x} + \left(\frac{\partial\mu}{\partial x} - \epsilon^2\frac{\partial B(\tau_j,z)}{\partial x}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right]\\
&\quad - \frac{1}{E_j(\tau)}\frac{\partial\nu}{\partial x}(\tau-\tau_j)(\nu-\mu+\epsilon^2 B(\tau_j,z))e^{-\nu(\tau-\tau_j)},\\
\frac{\partial A_j(\tau)}{\partial x} &= \frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^1(\tau)}{\partial x} - \frac{A_j(\tau)}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^2(\tau)}{\partial x},\\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^1(\tau)}{\partial x} &= \frac{1}{A_j^2(\tau)}\left[2\tanh\frac{\nu(\tau-\tau_j)}{2}\left(\mu\frac{\partial B(\tau_j,z)}{\partial x} - \epsilon^2 B(\tau_j,z)\frac{\partial B(\tau_j,z)}{\partial x} + B(\tau_j,z)\frac{\partial\mu}{\partial x}\right.\right.\\
&\quad \left.\left. + \frac{1}{2}\frac{\partial\lambda^2}{\partial x}(z-z^2)\right) + \frac{\partial\nu}{\partial x}\frac{\tau-\tau_j}{2}\left(B(\tau_j,z)(2\mu-\epsilon^2 B(\tau_j,z)) + \lambda^2(z-z^2)\right)\right],\\
\frac{1}{\tilde{A}_j^2(\tau)}\frac{\partial\tilde{A}_j^2(\tau)}{\partial x} &= \frac{1}{V_0 A_j^2(\tau)}\left[\left(1+(\mu-\epsilon^2 B(\tau_j,z))\frac{\tau-\tau_j}{2}\right)\frac{\partial\nu}{\partial x}\right.\\
&\quad \left. + \left(\nu\frac{\partial\nu}{\partial x}\frac{\tau-\tau_j}{2} + \frac{\partial\mu}{\partial x} - \epsilon^2\frac{\partial B(\tau_j,z)}{\partial x}\right)\tanh\frac{\nu(\tau-\tau_j)}{2}\right].
\end{aligned}
$$

E.2 On the Levenberg-Marquardt algorithm

In this section, we detail the classic Levenberg-Marquardt algorithm and the extended version handling bound constraints.

E.2.1 Standard Levenberg-Marquardt algorithm

In Algorithm 2, the function L corresponds to the value of the objective function F in $\boldsymbol{\Theta}_{k+1}$ when the residuals are approximated by a first order Taylor expansion. Mathematically, we have:

$$F(\boldsymbol{\Theta}_k + \boldsymbol{d}) \simeq L(\boldsymbol{d}) = \frac{1}{2}\left\|\boldsymbol{f}(\boldsymbol{\Theta}_k) + \boldsymbol{J}(\boldsymbol{\Theta}_k)\boldsymbol{d}\right\|^2.$$

Hence, $L(\mathbf{0}) - L(\boldsymbol{d})$ can be interpreted as the gain predicted by a linear model. It is easy to check that this quantity is always positive.

Algorithm 2: Levenberg-Marquardt algorithm

Input: $\boldsymbol{\Theta}_0$, F, $\boldsymbol{f}$, $\boldsymbol{J}$, L, ϵ_1, ϵ_2, ϵ_3, k_{max}, τ

1 **begin**
2 $\quad k \leftarrow 0$; $\nu \leftarrow 2$
3 $\quad \mathbf{A} \leftarrow \boldsymbol{J}(\boldsymbol{\Theta}_0)^T \boldsymbol{J}(\boldsymbol{\Theta}_0)$; $\boldsymbol{g} \leftarrow \boldsymbol{J}(\boldsymbol{\Theta}_0)^T \boldsymbol{f}(\boldsymbol{\Theta}_0)$
4 $\quad$ *found* $\leftarrow (F(\boldsymbol{\Theta}_k) \leq \epsilon_1$ or $\|\boldsymbol{g}\|_\infty \leq \epsilon_2)$; $\mu \leftarrow \tau \max_i\{a_{ii}\}$
5 $\quad$ **while** *!(found) and* $k < k_{max}$ **do**
6 $\quad\quad$ Solve $(\mathbf{A} + \mu \mathbf{I})\boldsymbol{d} = -\boldsymbol{g}$
7 $\quad\quad$ **if** $\|\boldsymbol{d}\|^2 \leq \epsilon_3^2 \|\boldsymbol{\Theta}_k\|^2$ **then**
8 $\quad\quad\quad$ *found* $\leftarrow$ **true**
9 $\quad\quad$ **end**
10 $\quad\quad$ **else**
11 $\quad\quad\quad$ $\boldsymbol{\Theta}_{k+1} \leftarrow \boldsymbol{\Theta}_k + \boldsymbol{d}$
12 $\quad\quad\quad$ **if** $F(\boldsymbol{\Theta}_k) - F(\boldsymbol{\Theta}_{k+1}) > 0$ *and* $L(\mathbf{0}) - L(\boldsymbol{d}) > 0$ **then**
13 $\quad\quad\quad\quad$ $\eta \leftarrow (F(\boldsymbol{\Theta}_k) - F(\boldsymbol{\Theta}_{k+1}))/(L(\mathbf{0}) - L(\boldsymbol{d}))$
14 $\quad\quad\quad\quad$ $\mathbf{A} \leftarrow \boldsymbol{J}(\boldsymbol{\Theta}_{k+1})^T \boldsymbol{J}(\boldsymbol{\Theta}_{k+1})$; $\boldsymbol{g} \leftarrow \boldsymbol{J}(\boldsymbol{\Theta}_{k+1})^T \boldsymbol{f}(\boldsymbol{\Theta}_{k+1})$
15 $\quad\quad\quad\quad$ *found* $\leftarrow (F(\boldsymbol{\Theta}_{k+1}) \leq \epsilon_1$ ou $\|\boldsymbol{g}\|_\infty \leq \epsilon_2)$
16 $\quad\quad\quad\quad$ $\mu \leftarrow \mu \max\{\frac{1}{3}, 1 - (2\eta - 1)^3\}; \nu \leftarrow 2$
17 $\quad\quad\quad$ **end**
18 $\quad\quad\quad$ **else**
19 $\quad\quad\quad\quad$ $\mu \leftarrow \mu\nu$; $\nu \leftarrow 2\nu$
20 $\quad\quad\quad$ **end**
21 $\quad\quad$ **end**
22 $\quad\quad$ $k \leftarrow k + 1$
23 $\quad$ **end**
24 **end**

The quantity η (appearing first in line 13 of the routine above) allows to measure how good the approximation of $F(\boldsymbol{\Theta}_k + \boldsymbol{d})$ by $L(\boldsymbol{d})$ is. A large value of η indicates that $L(\boldsymbol{d})$ is a good approximation of $F(\boldsymbol{\Theta}_k + \boldsymbol{d})$, whereas a small value of η indicates the contrary. In the first case, μ is decreased in order to imitate the Gauss-Newton algorithm behaviour; in the second case, μ is increased in order to imitate the behaviour of the steepest descent method.

As regards the updating strategy of the damped parameter, we use the one introduced in [N+99].

E.2.2 Extended Levenberg-Marquardt algorithm handling bound constraints

We present an extension of the classic Levenberg-Marquardt algorithm which can handle bound constraints. This extension was first proposed by [KFY02].

Let us denote by P_X the projection onto the feasible set X (which in the framework of the DDSVLMM is equal to $(\mathbb{R}_+)^4 \times (\mathbb{R}_+^*)^3 \times]-1;1[$). With respect to the Algorithm 2, only the lines from 11 to 20 must be modified. They must be replaced by the following ones.

The parameters γ, β and σ are empirically fixed parameters in $(0, 1)$.

Algorithm 3: Extended Levenberg-Marquardt algorithm

```
Θ_{k+1} ← P_X(Θ_k + d); d ← Θ_{k+1} − Θ_k
if F(P_X(Θ_{k+1})) ≤ γF(Θ_k) then
    if F(Θ_k) − F(Θ_{k+1}) > 0 and L(0) − L(d) > 0 then
        η ← (F(Θ_k) − F(Θ_{k+1}))/(L(0) − L(d))
        A ← J(Θ_{k+1})^T J(Θ_{k+1}); g ← J(Θ_{k+1})^T f(Θ_{k+1})
        found ← (F(Θ_{k+1}) ≤ ε_1 ou ‖g‖_∞ ≤ ε_2)
        μ ← μ max{1/3, 1 − (2η − 1)^3}; ν ← 2
    end
    else
        μ ← μν; ν ← 2ν
    end
end
else if ∇F(Θ_{k+1})^T d ≤ 0 then
    Perform a line search, i.e. look for α such that F(P_X(Θ_k + αd)) is reasonably lower
      than F(Θ_k)
end
else
    Apply a projected gradient step, i.e. compute t = max_{l∈ℕ} β^l such that
      F(P_X(Θ_k − tg)) ≤ F(Θ_k) + σ∇g^T(P_X(Θ_k − tg) − Θ_k)
end
```

Bibliography

[AA00] Leif Andersen and Jesper Andreasen. Volatility skews and extensions of the LIBOR market model. *Applied Mathematical Finance*, 7(1):1–32, 2000. 57, 65, 66

[AAB+20] Hervé Andres, Pierre-Edouard Arrouy, Paul Bonnefoy, Alexandre Boumezoued, and Sophian Mehalla. Fast calibration of the LIBOR Market Model with Stochastic Volatility based on analytical gradient. *arXiv preprint arXiv:2006.13521*, 2020. 9, 19, 205

[ABLM20] Pierre-Edouard Arrouy, Alexandre Boumezoued, Bernard Lapeyre, and Sophian Mehalla. Jacobi Stochastic Volatility factor for the Libor Market Model. *HAL preprint: hal-02468583*, 2020. 142, 205, 206

[ACA20a] Aurélien Alfonsi, Adel Cherchali, and Jose Arturo Infante Acevedo. Multilevel Monte-Carlo for computing the SCR with the standard formula and other stress tests. *arXiv preprint arXiv:2010.12651*, 2020. 30

[ACA20b] Aurélien Alfonsi, Adel Cherchali, and Jose Arturo Infante Acevedo. A synthetic model for asset-liability management in life insurance, and analysis of the SCR with the standard formula. *European Actuarial Journal*, 10(2):457–498, 2020. 21

[Adr14] Robert Adrain. *Research Concerning the Probabilities of the Errors which Happen in Making Observations, & C.* George Long, 1814. 122

[AF18] Damien Ackerer and Damir Filipović. Option pricing with orthogonal polynomial expansions. *Swiss Finance Institute Research Paper*, (17-41), 2018. 190, 193, 194, 217

[AFP17] Damien Ackerer, Damir Filipović, and Sergio Pulido. The Jacobi stochastic volatility model. *Finance and Stochastics*, pages 1–34, 2017. 20, 90, 91, 142, 143, 147, 148, 149, 150, 152, 174, 175, 177, 182, 184, 188, 190, 191, 194

[AG99] Dong-Hyun Ahn and Bin Gao. A parametric nonlinear model of term structure dynamics. *The Review of Financial Studies*, 12(4):721–762, 1999. 103, 104

[AGL16] Søren Asmussen, Pierre-Olivier Goffard, and Patrick J Laub. Orthonormal polynomial expansions and lognormal sum densities. *arXiv preprint arXiv:1601.01763*, 2016. 141, 142

[AJEE19] Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019. 110

[Akh20] Naum Il ich Akhiezer. *The classical moment problem and some related questions in analysis*. SIAM, 2020. 119

[Alf06] Aurélien Alfonsi. *Credit Risk modelling. Calibration and discretization of financial models.* PhD thesis, 2006. 98, 154

[Alf12] Aurélien Alfonsi. Strong convergence of some drift implicit Euler scheme. Application to the CIR process. *arXiv preprint arXiv:1206.3855*, 2012. 106, 107

[Alf15] Aurélien Alfonsi. *Affine Diffusions and Related Processes: Simulation, Theory and Applications*, volume 6. Springer, 2015. 91, 92, 99, 100, 101, 102, 105, 107, 108, 157, 165, 182, 184, 206, 230

[AMLB20] Pierre-Edouard Arrouy, Sophian Mehalla, Bernard Lapeyre, and Alexandre Boumezoued. Jacobi Stochastic Volatility factor for the Libor Market Model. 2020. 9, 16, 117

[AMST07] Hansjörg Albrecher, Philipp Mayer, Wim Schoutens, and Jurgen Tistaert. The little Heston trap. *Wilmott*, (1):83–92, 2007. 142, 206, 207, 216

[AP07] Leif BG Andersen and Vladimir V Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007. 65

[AS64] Milton Abramowitz and IA Stegun. Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables Applied mathematics series. *National Bureau of Standards, Washington, DC*, 1964. 92, 93, 96, 126, 134, 176

[AZ20] Boddele C. Bulpitt T. Darkiewicz-Moniuszko G. Ward R. Zandbergen F. Arrouy, P.-E. and S. Zhu. Setting discount rates under IFRS 17: Getting the job done. 2020. 21

[Bac00] Louis Bachelier. Théorie de la spéculation. 17:21–86, 1900. 46

[BBD08] Abdel Berkaoui, Mireille Bossy, and Awa Diop. Euler scheme for SDEs with non-lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics*, 12:1–11, 2008. 184

[BC16] Paul Bonnefoy-Cudraz. Implémentation et calibrage d'un Générateur de Scénarios Economiques: impact sur la volatilité du Solvency Capital Requirement. *Mémoire, EURIA*, 2016. 44

[BCOR09] Sergio Blanes, Fernando Casas, Jose-Angel Oteo, and José Ros. The Magnus expansion and some of its applications. *Physics reports*, 470(5-6):151–238, 2009. 116, 117

[BD04] Mireille Bossy and Awa Diop. An efficient discretization scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^\alpha$, $\alpha \in [1/2, 1)$. *Rapport de recherche, Institut National de Recherche en Informatique et en Automatique (INRIA)*, (5396), 2004. 106

[BDB01] Alan Brace, Tim Dun, and Geoff Barton. Towards a central interest rate model. *Option pricing, interest rates and risk management*, pages 278–313, 2001. 61, 75

[Ben10] H Bensussan. *Interest rate and longevity risks: dynamic modelling and applications to derivative products and life insurance.* PhD thesis, Ecole Polytechnique, 2010. 112

[BGaM97] Alan Brace, Dariusz G atarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical finance*, 7(2):127–155, 1997. 56, 57, 58, 59, 60, 180

[BGK17] Elia Berdin, Helmut Gründl, and Christian Kubitza. Rising interest rates, lapse risk, and the stability of life insurers. Technical report, ICIR Working Paper Series, 2017. 26

[BGM10] Eric Benhamou, Emmanuel Gobet, and Mohammed Miri. Time dependent Heston model. *SIAM Journal on Financial Mathematics*, 1(1):289–325, 2010. 108, 118, 143, 144

[BH05] Tomas Björk and Henrik Hult. A note on Wick products and the fractional Black-Scholes model. *Finance and Stochastics*, 9(2):197–209, 2005. 110

[Bjö09] Tomas Björk. *Arbitrage theory in continuous time*. Oxford university press, 2009. 50, 231

[BKKR06] Daniel Bauer, Rüdiger Kiesel, Alexander Kling, and Jochen Ruß. Risk-neutral valuation of participating life insurance contracts. *Insurance: Mathematics and Economics*, 39(2):171–183, 2006. 21

[BL12] Greg Blekherman and Jean-Bernard Lasserre. The truncated K-moment problem for closure of open sets. *Journal of Functional Analysis*, 263(11):3604–3616, 2012. 118

[Bla76] Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1-2):167–179, 1976. 47

[BLNZ95] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995. 206, 211

[BM98] Sergei Blinnikov and Richhild Moessner. Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1):193–205, 1998. 129, 137

[BM07] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007. 44, 50, 54, 59, 60, 61, 75, 180, 227

[BMO98] Khaled Bahlali, Brahim Mezerdi, and Youssef Ouknine. Pathwise uniqueness and approximation of solutions of stochastic differential equations. In *Séminaire de Probabilités XXXII*, pages 166–187. Springer, 1998. 242

[BMV19] Fabrice Borel-Mathurin and Julien Vedani. Market-consistent valuation: a step towards calculation stability. 2019. 21

[Bom13] Romain Bompis. *Stochastic expansion for the diffusion processes and applications to option pricing*. PhD thesis, Ecole Polytechnique X, 2013. 143

[BRS12] Daniel Bauer, Andreas Reuss, and Daniela Singer. On the calculation of the Solvency Capital Requirement based on nested simulations. *Astin Bulletin*, 42(02):453–499, 2012. 21, 29

[BS73] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973. 61

[BW09] Rabi N Bhattacharya and Edward C Waymire. *Stochastic processes with applications*, volume 61. Siam, 2009. 92, 93, 94

[Car] T Carleman. Les fonctions quasi-analytiques,(1926). *Gauthier-Villars, Paris*. 119

[CdBRG17] Yiran Cui, Sebastian del Baño Rollin, and Guido Germano. Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research*, 263(2):625–638, 2017. 205, 206, 207, 208

[Che58] PL Chebyshev. Sur les fractions continues. *J. Math. Pures Appl*, 3(2):289–323, 1858. 123, 124

[Che59] Pafnutii Lvovich Chebyshev. Sur le développement des fonctions à une seule variable. *Bull. Acad. Sci. St. Petersb*, 1:193–200, 1859. 124

[CKRT12] Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740, 2012. 109, 113, 143, 147

[CM99] Peter Carr and Dilip Madan. Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61–73, 1999. 142, 205

[CN+14] Marcus C Christiansen, Andreas Niemeyer, et al. Fundamental definition of the solvency capital requirement in Solvency ii. *Astin Bulletin*, 44(3):501–533, 2014. 26

[Cra28] Harald Cramér. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928. 122, 125, 127, 128, 138

[Cra46] Harald Cramér. *Mathematical methods of statistics (PMS-9)*, volume 9. Princeton university press, 1946. 122, 123

[Cra04] Harald Cramér. *Random variables and probability distributions*, volume 36. Cambridge University Press, 2004. 122

[CS07] Peter Carr and Jian Sun. A new approach for option pricing under stochastic volatility. *Review of Derivatives Research*, 10(2):87–150, 2007. 103, 104

[Cuc19] Christa Cuchiero. Polynomial processes in stochastic portfolio theory. *Stochastic processes and their applications*, 129(5):1829–1872, 2019. 143

[D+99] Laurent Decreusefond et al. Stochastic analysis of the fractional Brownian motion. *Potential analysis*, 10(2):177–214, 1999. 109

[D+14] Nguyen Tien Dung et al. Jacobi processes driven by fractional Brownian motion. *Taiwanese Journal of Mathematics*, 18(3):835–848, 2014. 110

[DABB17] Laurent Devineau, Pierre-Edouard Arrouy, Paul Bonnefoy, and Alexandre Boumezoued. Fast calibration of the Libor Market Model with Stochastic Volatility and Displaced Diffusion. *Journal of Industrial and Management Optimization*, 13:1, 2017. 142, 143, 144, 177, 193, 194

[dBRFCU10] Sebastian del Baño Rollin, Albert Ferreiro-Castilla, and Frederic Utzet. On the density of log-spot in the Heston volatility model. *Stochastic Processes and their Applications*, 120(10):2037–2063, 2010. 205, 206, 207, 216

[dCAH18] Maria Fernanda del Carmen Agoitia Hurtado. *Time-inhomogeneous polynomial processes in electricity spot price models*. PhD thesis, 2018. 109, 113, 115, 117

[dcpedr20] Autorité de contrôle prudentiel et de résolution. Générateurs de scénarios économiques : points d'attention et bonnes pratiques. 2020. 21, 22, 23

[DD+98] Griselda Deelstra, Freddy Delbaen, et al. Convergence of discretized stochastic (interest rate) processes with stochastic drift term. *Applied stochastic models and data analysis*, 14(1):77–84, 1998. 106, 184

[DFGT07] José Da Fonseca, Martino Grasselli, and Claudio Tebaldi. Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research*, 10(2):151–180, 2007. 112

[DHPD00] Tyrone E Duncan, Yaozhong Hu, and Bozenna Pasik-Duncan. Stochastic calculus for fractional Brownian motion I. Theory. *SIAM Journal on Control and Optimization*, 38(2):582–612, 2000. 109, 111

[Dir09] Directive 2009/138/EC of the European Parliament and the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency ii). Technical report, European Insurance and Occupational Pensions Authority, 2009. 24, 25, 26

[DJDW01] FCJM De Jong, FC Drost, and BJM Werker. A jump-diffusion model for exchange rates in a target zone. *Statistica Neerlandica*, 55(3):270–300, 2001. 90

[DM56] Abraham De Moivre. *The doctrine of chances: A method of calculating the probabilities of events in play*, volume 1. Chelsea Publishing Company, 1756. 121

[DNS12] Steffen Dereich, Andreas Neuenkirch, and Lukasz Szpruch. An Euler-type method for the strong approximation of the Cox-Ingersoll-Ross process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2140):1105–1115, 2012. 106

[Dou05] Yan Doumerc. *Matrices aléatoires, processus stochastiques et groupes de réflexions*. PhD thesis, Toulouse 3, 2005. 111, 112, 113

[DPS00] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000. 69

[DS02] Freddy Delbaen and Hiroshi Shirakawa. An interest rate model with upper and lower bounds. *Asia-Pacific Financial Markets*, 9(3-4):191–209, 2002. 89, 90, 92, 104, 105, 143

[Duf01] Daniel Dufresne. *The integrated square-root process*. Centre for Actuarial Studies, Department of Economics, University of Melbourne, 2001. 104, 167, 172

[EER+18] Omar El Euch, Mathieu Rosenbaum, et al. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018. 110

[EER19] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019. 110, 111

[EGG20] Ernst Eberlein, Christoph Gerhart, and Zorana Grbac. Multiple curve Lévy forward price model allowing for negative interest rates. *Mathematical Finance*, 30(1):167–195, 2020. 57

[EH13] Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013. 103

[Eio14] EIOPA Insurance stress test 2014. Technical report, European Insurance and Occupational Pensions Authority, 2014. 26

[Eio15] Guidelines on the valuation of technical provisions. Technical report, European Insurance and Occupational Pensions Authority, 2015. 22

[EKMH14] Nicole El Karoui, Mohamed Mrad, and Caroline Hillairet. Ramsey Rule with Progressive utility and Long Term Affine Yields Curves. 2014. 42

[EMOT55] A Erldelyi, W Magnus, F Oberhettinger, and FG Tricomi. Higher transcendental functions, vol. 2. 1955. 176

[EP11] Bjorn Eriksson and Martijn Pistorius. Method of moments approach to pricing double barrier contracts in polynomial jump-diffusion models. *International Journal of Theoretical and Applied Finance*, 14(07):1139–1158, 2011. 109, 113, 143

[ET10] Erik Ekström and Johan Tysk. The Black–Scholes equation in stochastic volatility models. *Journal of Mathematical Analysis and Applications*, 368(2):498–507, 2010. 169, 241

[EVDH03] Robert J Elliott and John Van Der Hoek. A general fractional white noise theory and applications to finance. *Mathematical Finance*, 13(2):301–330, 2003. 109

[Fan10] FANG Fang. The COS method: An efficient Fourier method for pricing financial derivatives. 2010. 71

[Fel] W Feller. An Introduction to Probability Theory and its Applications, vol. ii, new york, 1971. *Zbl0219*, 60003. 119, 121

[Fil09] Damir Filipovic. *Term-Structure Models. A Graduate Course.* Springer, 2009. 99

[FL16] Damir Filipović and Martin Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016. 100, 113, 114, 115, 117, 143, 147

[FL20] Damir Filipović and Martin Larsson. Polynomial jump-diffusion models. *Stochastic Systems*, 10(1):71–97, 2020. 113

[FLCM16] Anthony Floryszczak, Olivier Le Courtois, and Mohamed Majri. Inside the Solvency 2 black box: net asset values and solvency capital requirements with a least-squares Monte-Carlo approach. *Insurance: Mathematics and Economics*, 71:15–26, 2016. 29

[FMS13] Damir Filipović, Eberhard Mayerhofer, and Paul Schneider. Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111, 2013. 139, 140, 141, 142, 190

[Gau09] Carl Friedrich Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, volume 7. Perthes et Besser, 1809. 122

[Gil03] J. Ch. Gilbert. Optimisation Différentiable–Théorie et Algorithmes. Syllabus de cours à l'ENSTA, 2003. 219

[GJ06] Christian Gourieroux and Joann Jasiak. Multivariate Jacobi process with application to smooth transitions. *Journal of Econometrics*, 131(1-2):475–505, 2006. 111

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018. 110

[GKT16] Zorana Grbac, David Krief, and Peter Tankov. Approximate Option Pricing in the Lévy Libor Model. In *Advanced Modelling in Mathematical Finance*, pages 453–476. Springer, 2016. 143

[GP51] J Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38(3-4):481–482, 1951. 69, 70

[GPSS15] Zorana Grbac, Antonis Papapantoleon, John Schoenmakers, and David Skovmand. Affine LIBOR models with multiple curves: theory, examples and calibration. *SIAM Journal on Financial Mathematics*, 6(1):984–1025, 2015. 57

[GR11] István Gyöngy and Miklós Rásonyi. A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic processes and their applications*, 121(10):2189–2200, 2011. 105

[GS04] Christian Gouriéroux and Razvan Sufana. Derivative pricing with multivariate stochastic volatility: Application to credit risk. *Les Cahiers du CREF of HEC Montréal Working Paper No. CREF*, pages 04–09, 2004. 111

[GV19] Hilmar Gudmundsson and David Vyncke. On the calibration of the 3/2 model. *European Journal of Operational Research*, 2019. 103, 206, 209

[Gyö98] István Gyöngy. A note on Euler's approximations. *Potential Analysis*, 8(3):205–216, 1998. 106

[Hal98] Anders Hald. *A History of Mathematical Statistics from 1750 to 1930*, volume 2. Wiley New York, 1998. 123

[Hal00] Anders Hald. The Early History of the Cumulants and the Gram-Charlier series. *International Statistical Review*, 68(2):137–153, 2000. 122, 123, 124

[Hal13] Peter Hall. *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, 2013. 123, 135

[Hal15] Brian Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*, volume 222. Springer, 2015. 117

[Ham20] Hans Hamburger. Über eine erweiterung des stieltjesschen momentenproblems. *Mathematische Annalen*, 81(2):235–319, 1920. 118

[Ham44] Hans Ludwig Hamburger. Hermitian transformations of deficiency-index (1, 1), Jacobi matrices and undetermined moment problems. *American Journal of Mathematics*, 66(4):489–522, 1944. 118

[Hau23] Felix Hausdorff. Momentprobleme für ein endliches intervall. *Mathematische Zeitschrift*, 16(1):220–248, 1923. 118

[Hes93] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993. 205, 206, 207

[Hes97] Steven L Heston. A simple new formula for options with stochastic volatility. 1997. 61, 103

[Hil26] Einar Hille. A class of reciprocal functions. *Annals of Mathematics*, pages 427–464, 1926. 128

[HL08] Patrick Hagan and Andrew Lesniewski. LIBOR market model with SABR style stochastic volatility. *JP Morgan Chase and Ellington Management Group*, 32, 2008. 57

[HM05] Desmond J Higham and Xuerong Mao. Convergence of Monte-Carlo simulations involving the mean-reverting square root process. *Journal of Computational Finance*, 8(3):35–61, 2005. 106

[HMS02] Desmond J Higham, Xuerong Mao, and Andrew M Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2002. 105

[HØ03] Yaozhong Hu and Bernt Øksendal. Fractional white noise calculus and applications to finance. *Infinite dimensional analysis, quantum probability and related topics*, 6(01):1–32, 2003. 109

[HW90] John Hull and Alan White. Pricing interest-rate-derivative securities. *The review of financial studies*, 3(4):573–592, 1990. 223

[HW99a] Patrick S Hagan and Diana E Woodward. Equivalent black volatilities. *Applied Mathematical Finance*, 6(3):147–157, 1999. 63, 67

[HW99b] John Hull and A White. Forward rate volatilities, swap rate volatilities, and the implementation of the LIBOR market model, working paper, University of Toronto. 1999. 74

[Jam97] Farshid Jamshidian. LIBOR and swap market models and measures. *Finance and Stochastics*, 1(4):293–330, 1997. 46, 57, 143

[Jam99] Farshid Jamshidian. LIBOR market model with semimartingales. Technical report, Working Paper, NetAnalytic Ltd, 1999. 57

[JLP+19] Eduardo Abi Jaber, Martin Larsson, Sergio Pulido, et al. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019. 110, 111

[JR03] Mark Joshi and Riccardo Rebonato. A stochastic-volatility, displaced-diffusion extension of the LIBOR market model. *Quantitative Finance*, 3(6):458–469, 2003. 143

[K+88] Shuya Kanagawa et al. On the rate of convergence for Maruyama's approximate solutions of stochastic differential equations. 1988. 105

[Kap10] Alexander Kaplun. *Bounded short-rate models with Ehrenfest and Jacobi processes*. PhD thesis, 2010. 90, 92

[KFY02] Christian Kanzow, Masao Fukushima, and Nobuo Yamashita. *Levenberg-Marquardt methods for constrained nonlinear equations with strong local convergence properties*. Citeseer, 2002. 248

[Kie18] Hansjörg Kielhöfer. Calculus of variations. *Texts in Applied Mathematics*, 67, 2018. 85

[KM65] Samuel Karlin and James McGregor. Ehrenfest urn models. *Journal of Applied Probability*, 2(2):352–376, 1965. 90

[KN12] Dessislava Koleva and Elisa Nicolato. Option pricing under Heston and 3/2 stochastic volatility models: an approximation to the fast Fourier transform. *Masterks thesis, Aarhus University*, 2012. 103, 104

[KP92] Peter E Kloeden and Eckhard Platen. Stochastic differential equations. In *Numerical Solution of Stochastic Differential Equations*, pages 103–160. Springer, 1992. 49

[Kre44] Mark Grigorievich Krein. On one extrapolation problem of A.N. Kolmogorov. *Dokl. Akad. Nauk SSSR*, 1944. 120

[KS91] Ioannis Karatzas and Steven E Shreve. Brownian motion and stochastic calculus Springer-Verlag. *New York*, 1991. 90, 91, 94, 95, 109, 167, 169

[KT81] Samuel Karlin and Howard E Taylor. *A second course in stochastic processes*. Elsevier, 1981. 89, 143

[L+05] Manolis IA Lourakis et al. A brief description of the Levenberg-Marquardt algorithm implemented by levmar. *Foundation of Research and Technology*, 4(1):1–6, 2005. 212

[Lap10] Pierre-Simon Laplace. Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur applications aux probabilités. *Memoires de l'Academie des Sciences de Paris*, 1810. 122

[LBGO19] Shuaiqiang Liu, Anastasia Borovykh, Lech A Grzelak, and Cornelis W Oosterlee. A neural network-based framework for financial model calibration. *arXiv preprint: arXiv:1904.10523*, 2019. 205

[LDB20] P. Bonnefoy L. Devineau, P.-E. Arrouy and A. Boumezoued. Fast calibration of the libor market model with stochastic volatility and displaced diffusion. *Journal of Industrial and Management Optimization*, 16(4):1699, 2020. 20, 68, 205, 206, 210, 212, 216

[Leg05] Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805. 122

[Lew] AL Lewis. Option Valuation under Stochastic Volatility with mathematica code, 2000. 103

[LKD10] Roger Lord, Remmert Koekkoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010. 105, 183

[LL11] Damien Lamberton and Bernard Lapeyre. *Introduction to stochastic calculus applied to finance*. Chapman and Hall/CRC, 2011. 57, 232

[LL16] Andreas Lagerås and Mathias Lindholm. Issues with the Smith–Wilson method. *Insurance: Mathematics and Economics*, 71:93–102, 2016. 41

[LP17] Martin Larsson and Sergio Pulido. Polynomial diffusions on compact quadric sets. *Stochastic Processes and their Applications*, 127(3):901–926, 2017. 91

[LS07] Kristian Stegenborg Larsen and Michael Sørensen. Diffusion models for exchange rates in a target zone. *Mathematical Finance*, 17(2):285–306, 2007. 90

[Ma09] Jun Ma. Pricing Foreign Equity Options with Stochastic Correlation and Volatility. *Annals of Economics & Finance*, 10(2), 2009. 89, 92, 143

[Mar63] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. 206

[Mar78] William Margrabe. The value of an option to exchange one asset for another. *The journal of finance*, 33(1):177–186, 1978. 47

[Mea09] Measurement of Liabilities for Insurance Contracts: Current Estimates and Risk Margins. Technical report, International Actuarial Association, 2009. 31

[Mil29] WE Milne. On the degree of convergence of the Gram-Charlier series. *Transactions of the American Mathematical Society*, 31(3):422–443, 1929. 127

[MNT99] Kaj Madsen, Hans Bruun Nielsen, and Ole Tingleff. Methods for non-linear least squares problems. 1999. 211

[Mol30] Edward C Molina. The theory of probability: some comments on Laplace's théorie analytique. *Bulletin of the American Mathematical Society*, 36(6):369–392, 1930. 125

[MVN68] Benoit B Mandelbrot and John W Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968. 109

[MYT18] Yuliya Mishura and Anton Yurchenko-Tytarenko. Fractional Cox–Ingersoll–Ross process with non-zero mean. *arXiv preprint arXiv:1804.01677*, 2018. 110

[N+99] Hans Bruun Nielsen et al. *Damping parameter in Marquardt's method*. IMM, 1999. 248

[Nas79] JC Nash. Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation. Adam Hilger, Bristol. 1990. *Chap*, 3:30–48, 1979. 209

[NS14] Andreas Neuenkirch and Lukasz Szpruch. First order strong approximations of scalar SDEs defined in a domain. *Numerische Mathematik*, 128(1):103–136, 2014. 106, 107

[NV08] Syoiti Ninomiya and Nicolas Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, 15(2):107–121, 2008. 108

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006. 210, 211

[Øks03] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003. 113

[Pal15] Ernesto Palidda. *Multi factor stochastic volatility for interest rates modeling*. PhD thesis, Université Paris-Est, 2015. 112

[Par14] European Parlament. Delegated acts. *Official Journal of the European Union*, 2014. 21, 22, 48

[Pat95] Billingsley Patrick. Probability and measure. *A Wiley-Interscience Publication, John Wiley*, 1995. 118, 119

[PCC+16] Hal Pedersen, Mary Pat Campbell, Stephan L Christiansen, Samuel H Cox, Daniel Finn, Ken Griffin, Nigel Hooker, Matthew Lightwood, Stephen M Sonlin, and Chris Suchar. Economic scenario generators: a practical guide. *The Society of Actuaries*, 2016. 21, 22

[Ped98] Henrik L Pedersen. On krein's theorem for indeterminacy of the classical moment problem. *Journal of approximation theory*, 95(1):90–100, 1998. 120

[Pól20] Georg Pólya. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift*, 8(3-4):171–181, 1920. 124

[R+02] Aurel Rascanu et al. Differential equations driven by fractional Brownian motion. *Collectanea Mathematica*, 53(1):55–81, 2002. 109

[Reb98] Riccardo Rebonato. *Interest-rate option models*. Second Edition, Wiley, 1998. 73

[Reg14] Commission delegated regulation (EU) 2015/35. Technical report, European Parliament and Council, 2014. 31

[Rod15] Olinde Rodrigues. *De l'attraction des sphéroïdes*. PhD thesis, 1815. 126

[RS75] Michael Reed and Barry Simon. *II: Fourier Analysis, Self-Adjointness*, volume 2. Elsevier, 1975. 121

[RY13] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013. 109

[S+00] Jordan Stoyanov et al. Krein condition in probabilistic moment problems. *Bernoulli*, 6(5):939–949, 2000. 120

[Sch00] Philipp Schönbucher. A Libor market model with default risk. *Available at SSRN 261051*, 2000. 57

[Sch02] Erik Schlögl. Arbitrage-free interpolation in models of market observable interest rates. pages 197–218, 2002. 60

[Sch17] Konrad Schmüdgen. *The moment problem*, volume 14. Springer, 2017. 118

[Seb63] GAF Seber. The non-central chi-squared and beta distributions. *Biometrika*, 50(3/4):542–544, 1963. 101

[She14] Georgiy Shevchenko. Fractional Brownian motion in a nutshell. *arXiv preprint arXiv:1406.1956*, 2014. 109

[SN14] L Szpruch and A Neuenkirch. First order strong approximations of scalar SDEs with values in a domain. *Numerische Mathematik*, 128(1):103–136, 2014. 107

[ST43] James Alexander Shohat and Jacob David Tamarkin. *The problem of moments*. Number 1. American Mathematical Soc., 1943. 118, 119, 120

[Sti94] T-J Stieltjes. Recherches sur les fractions continues. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 8, pages J1–J122, 1894. 118

[Sto69] Hans R Stoll. The relationship between put and call option prices. *The Journal of Finance*, 24(5):801–824, 1969. 49

[Sto13] Jordan M Stoyanov. *Counterexamples in probability*. Courier Corporation, 2013. 118

[Stu04] Stephan Sturm. *Calculation of the greeks by Malliavin calculus*. na, 2004. 168

[SV03] Tommi Sottinen and Esko Valkeila. On arbitrage and replication in the fractional Black-Scholes pricing model. *Statistics & Decisions*, 21(2):93–108, 2003. 110

[SW01] Andrew Smith and Tim Wilson. Fitting yield curves with long term constraints. Technical report, Technical report, 2001. 39

[Sze26] G Szegö. Beiträge zur Theorie der Laguerreschen Polynome. i: Entwicklungssätze. *Mathematische Zeitschrift*, 25(1):87–115, 1926. 128

[Sze39] Gabor Szegö. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939. 141

[Tan03] Peter Tankov. *Financial modelling with jump processes*. CRC press, 2003. 113

[Tec] Technical Documentation of the methodology to derive EIOPA's risk-free interest rate term structures, year=2020, institution=European Insurance and Occupational Pensions Authority. Technical report. 31

[Tec19] Technical Documentation of the methodology to derive EIOPA's risk-free interest rate term structure. Technical report, European Insurance and Occupational Pensions Authority, 2019. 31, 35, 37, 38

[Thi73] Thorvald Nikolai Thiele. Om en tilnærmelsesformel. *Tidsskrift for mathematik*, 3:22–31, 1873. 125

[Thi89] TN Thiele. Almindelig Iagttagelseslaere: Sandsynlighedsregning og mindste Kvadraters Methode. *Copenhagen: CA Reitzel*, 1889. 125

[Tri47] Francesco Tricomi. Sulle funzioni ipergeometriche confluenti. *Annali di Matematica Pura ed Applicata*, 26(1):141–175, 1947. 134

[TT90] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990. 105

[uni09] Directive 2009/136/EC of the European parliament and of the council. Technical report, UNION, PEAN, 2009. 23, 26

[UO30] George E Uhlenbeck and Leonard S Ornstein. On the theory of the Brownian motion. *Physical review*, 36(5):823, 1930. 100

[Usp37] James Victor Uspensky. Introduction to mathematical probability. 1937. 125, 128, 134

[VD12] Julien Vedani and Laurent Devineau. Solvency assessment within the ORSA framework: issues and quantitative methodologies. *arXiv preprint arXiv:1210.6000*, 2012. 30

[VEKLP17] Julien Vedani, Nicole El Karoui, Stéphane Loisel, and Jean-Luc Prigent. Market inconsistencies of market-consistent European life insurance economic valuations: pitfalls and practical solutions. *European Actuarial Journal*, 7(1):1–28, 2017. 21, 23

[VV12] Almut ED Veraart and Luitgard AM Veraart. Stochastic volatility and stochastic leverage. *Annals of Finance*, 8(2-3):205–233, 2012. 89

[Wri31] Sewall Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97, 1931. 89

[WZ06] Lixin Wu and Fan Zhang. LIBOR market model with stochastic volatility. *Journal of industrial and management optimization*, 2(2):199, 2006. 54, 57, 62, 67, 68, 73, 143, 144, 180, 193, 205

[Y+02] Liqing Yan et al. The Euler scheme with irregular coefficients. *The Annals of Probability*, 30(3):1172–1194, 2002. 105

[YW+71] Toshio Yamada, Shinzo Watanabe, et al. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167, 1971. 90