Pricing interest rate, dividend, and equity risk

Thèse N° 9592

Présentée le 22 novembre 2019

au Collège du management de la technologie
Chaire Swissquote en finance quantitative
Programme doctoral en finance

pour l'obtention du grade de Docteur ès Sciences

par

Sander Félix M WILLEMS

Acceptée sur proposition du jury

Prof. E. Morellec, président du jury
Prof. D. Filipovic, directeur de thèse
Prof. P. Carr, rapporteur
Prof. C. Cuchiero, rapporteuse
Prof. P. Collin-Dufresne, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Acknowledgements

I am indebted to many people for guiding me through my doctoral studies at EPFL. First and foremost, I would like to thank my thesis advisor Damir Filipović, who taught me the ropes of academic research. His constant support and encouragement have allowed me to accomplish goals that once seemed far out of reach. I thank Peter Carr, Pierre Collin-Dufresne, Christa Cuchiero, and Erwan Morellec for accepting to be on my thesis committee and taking time out of their busy schedules to examine my thesis. I also thank Michèle Vanmaele for encouraging me to pursue a doctoral degree.

I would furthermore like to express gratitude to Peter Carr for the opportunity to visit New York University and to his colleagues at the Finance and Risk Engineering department for their warm welcome and excellent restaurant recommendations. My stay in New York has been one of the most enriching experiences of my doctoral studies.

Life does not stop at the end of the working day and I was fortunate to be surrounded with many people to spend time with outside the office. Let me start with a shout-out to some of the people I had the pleasure to meet in Switzerland: Alexey, Jakub, Julien, Francesco, Erik, Sebastian, Magdalena, Paweł, Damien, Alex, Sandra; thanks for all the great memories in Lausanne. To my amazing group of friends back home: returning to Belgium was always a pleasure thanks to you. My parents also deserve a dedicated acknowledgement, they have spared no effort to allow me to pursue my goals in life. Last and most important, thanks to my girlfriend Laurien for standing by my side all these years.

Lausanne, July 2019 S. W.

Abstract

This thesis studies the valuation and hedging of financial derivatives, which is fundamental for trading and risk-management operations in financial institutions. The three chapters in this thesis deal with derivatives whose payoffs are linked to interest rates, equity prices, and dividend payments.

The first chapter introduces a flexible framework based on polynomial jump-diffusions (PJD) to jointly price the term structures of dividends and interest rates. Prices for dividend futures, bonds, and the dividend paying stock are given in closed form. Option prices are approximated efficiently using a moment matching technique based on the principle of maximum entropy. An extensive calibration exercise shows that a parsimonious model specification has a good fit with Euribor interest rate swaps and swaptions, Euro Stoxx 50 index dividend futures and dividend options, and Euro Stoxx 50 index options.

The second chapter revisits the problem of pricing a continuously sampled arithmetic Asian option in the classical Black-Scholes setting. An identity in law links the integrated stock price to a one-dimensional polynomial diffusion, a particular instance of the PJD encountered in the first chapter. The Asian option price is approximated by a series expansion based on polynomials that are orthogonal with respect to the log-normal distribution. All terms in the series are fully explicit and no numerical integration nor any special functions are involved. The moment indeterminacy of the log-normal distribution introduces an asymptotic bias in the series, however numerical experiments show that the bias can safely be ignored in practice.

The last chapter presents a non-parametric method to construct a maximally smooth discount curve from observed market prices of linear interest rate products such as swaps, forward rate agreements, or coupon bonds. The discount curve is given in closed form and only requires basic linear algebra operations. The method is illustrated with several practical examples.

Keywords: mathematical finance, derivative pricing, term structure models, interest rates, polynomial processes, orthogonal polynomials, dividend derivatives

Résumé

Cette thèse étudie la valorisation et la couverture des produits dérivés en finance, activité essentielle dans une salle de marchés et un bureau de gestion des risques d'une institution financière. Les trois chapitres de cette thèse traitent des produits dérivés sur taux d'intérêts, actions et dividendes.

Le premier chapitre introduit un cadre flexible basé sur les processus stochastiques polynomiaux avec sauts afin de valoriser conjointement les courbes des taux de dividende et d'intérêt. Les prix des contrats à terme sur dividendes, obligations, et dividendes payant des actions possèdent une expression fermée. Les prix des options sont approximés de manière efficace à l'aide d'une technique de correspondance basée sur le principe d'entropie maximale. Un important travail de calibration montre que les paramètres d'un modèle parcimonieux permettent d'obtenir des valeurs qui concordent avec celles du taux d'intérêt Euribor swap et swaptions, l'indice Euro Stoxx 50 contrats à terme sur dividendes, l'indice Euro Stoxx 50 options sur dividendes et l'indice Euro Stoxx 50 options.

Le deuxième chapitre revisite le problème de la valorisation de l'option asiatique dans le modèle classique de Black-Scholes. Une identité en loi permet de faire le lien entre l'intégrale du prix de l'action et une diffusion polynomiale unidimensionnelle, un cas particulier des processus stochastiques polynomiaux avec sauts vus dans le premier chapitre. Le prix de l'option asiatique est approximé par un développement en série basé sur des polynômes qui sont orthogonaux sous la mesure log-normale. Tous les termes de la série possèdent une expression explicite et aucun calcul d'intégrale ni l'utilisation de fonctions spéciales n'est nécessaire. La loi log-normale n'étant pas caractérisée par ses moments, un biais asymptotique apparaît dans la série. Toutefois, des résultats numériques montrent que ce biais est négligeable en pratique.

Le dernier chapitre présente une méthode non-paramétrique pour construire une courbe de taux à partir des prix de produits à taux d'intérêt linéaire tels que le swap, l'accord à taux futur ou le coupon d'une obligation. La courbe de taux a une expression fermée et requiert seulement des opérations élémentaires d'algèbre linéaire. Cette méthode est illustrée avec plusieurs exemples pratiques.

Mots clefs : mathématiques financières, valorisation des produits dérivés, modèles de courbe de taux, taux d'intérêts, processus polynomiaux, polynômes orthogonaux, produits dérivés sur dividendes

Contents

List of Figures

List of Tables

Introduction

Mathematical finance is a relatively young research field that is concerned with modeling of financial markets. One of the major topics within this field, and also the focus of my thesis, is pricing and hedging of financial derivatives. Research questions in this topic are often rooted in problems faced by the financial industry, who employ derivatives in their day-to-day business. The constantly evolving product offering and regulations of the derivatives market provide ample opportunity for new research directions. In this introduction, I give an overview of the three chapters of my thesis and how they are related to each other. A more detailed introduction, including extensive literature reviews, can be found at the beginning of each chapter.

When I started my doctoral studies, one of the recent innovations in the derivatives market were the exchange traded dividend derivatives. These products have payoffs linked to the dividends paid by a company, or a basket of companies, over a certain period of time. The development of the market for dividend derivatives was mainly driven by issuers of retail equity structured products (e.g., autocallables) who wanted to offload their structurally long dividend exposure. From an investor's point of view, dividend derivatives offer an opportunity to have equity exposure with relatively low volatility, since dividend payments tend to be smoother over time than the prices of the stock paying the dividends. From a modeling perspective, the main difficulty with dividend derivatives stems from the fact that dividend payments also have a direct impact on the price of the dividend paying stock, and therefore also on prices of derivatives on the stock. The relation between dividend payments and stock prices is further complicated by changes in interest rates. Combining all three dimensions, i.e., dividends, stock prices, and interest rates, together in a tractable derivative pricing framework is a major challenge that has not been addressed in the literature to date. The first chapter of my thesis takes on this challenge by making use of polynomial jump-diffusions (PJD), as studied in Cuchiero et al. (2012) and Filipović and Larsson (2016, 2017). This class of stochastic processes is characterized by the fact that their generator maps polynomials to polynomials of the same degree or less. As a consequence, all conditional moments of any order can be computed in closed form, up to a matrix exponential. The first chapter develops a framework to jointly price the term structures of dividends and interest rates using PJD as building stones. The available conditional moments of the PJD lead to closed form prices for dividend futures, bonds, and the dividend paying stock. Option prices are approximated using a moment

matching technique. An extensive calibration exercise, using market data on dividend, stock, and interest rate derivatives over a five year period, illustrates the flexibility of the framework. The research in this chapter is based on Filipović and Willems (2019).

The second chapter continues with another application of PJD to derivative pricing. This chapter focuses on Asian style equity derivatives, where Asian style refers to the fact that the derivative's payoff depends on the average stock price over a period of time. These type of derivatives are popular, for example, when there are concerns about potential price manipulation, since it is more difficult to manipulate the price over a period of time than at one point in time. The stock price is assumed to follow a geometric Brownian motion and the averaging is assumed to be arithmetic and computed continuously. Furthermore, interest rates are assumed to be constant and the stock is assumed to pay no dividends. Even though this is arguably the most basic setup for modeling Asian style derivatives, no closed form solutions exists for Asian call and put options. The distribution of the stock price at one point in time is well-known to be log-normal, but the distribution of the average stock price is unknown. There exist many different methods to approximate Asian option prices, ranging from simulation to partial differential equation techniques. In the second chapter, I show how an identity in law can be used to relate the distribution of the average stock price to the distribution of a one-dimensional polynomial diffusion, which is a special instance of the PJD used in the first chapter. Therefore, all conditional moments of the average stock price are available in closed form. In a subsequent step, an orthogonal polynomial series expansion is used to approximate Asian option prices. The terms in the series can be calculated explicitly thanks to the available conditional moments. The series has an asymptotic bias, however numerical experiments show that it can safely be ignored for reasonable parameterizations. The research in this chapter is based on Willems (2019a).

In the last chapter of my thesis, I revisit the problem of bootstrapping a curve of zero-coupon bond prices, also known as the discount curve. A zero-coupon bond is a financial instrument that delivers a unit of currency at a future date, which is referred to as the maturity date. Prices of zero-coupon bonds therefore reflect the time-value of money. Derivative pricing models often assume that prices of zero-coupon bonds are observable for any maturity. In reality, however, zero-coupon bonds are not actively traded and in any case not for a continuum of maturities. In practice, prices of zero-coupon bonds are derived from more actively traded fixed-income instruments, such as coupon bonds, interest rate swaps, or forward rate agreements. The prices of these instruments depend on zero-coupon bond prices with a finite number of maturities and are therefore not sufficient to determine zero-coupon bond prices for the continuum of maturities. This is an under-determined problem in the sense that there exist infinitely many discount curves that perfectly price the observed instruments. In order to pin down one particular curve, an additional objective has to be imposed. The objective used in the last chapter is related to the smoothness of the discount curve. Specifically, the optimal curve is found by searching in an infinite-dimensional Hilbert function space for a discount curve with minimal integrated squared second derivative, subject to perfectly reproducing the prices of a given set of fixed-income instruments. The optimal

discount curve is given in closed form as the solution to a convex variational optimization problem. This chapter also investigates the sensitivities of the optimal discount curve with respect to perturbations in the input prices, which is important from a hedging perspective. Several practical examples using data on coupon bonds and interest rate swaps illustrate how the method works in practice. The method can also easily be adapted for use outside of the interest rate context. For example, the appendix of the first chapter describes how to bootstrap a curve of expected dividend payments from observed dividend futures prices. The research in this chapter is based on Filipović and Willems (2018).

Statement of Originality

I hereby declare that the content of this thesis is my own work, where some parts are the result of collaborations with my thesis supervisor Prof. Damir Filipović. No other person's work has been used without due acknowledgement.

1 A Term Structure Model for Dividends and Interest Rates

Over the last decade, dividends have become a standalone asset class instead of a mere side product of an equity investment. In this chapter, we introduce a framework based on polynomial jump-diffusions to jointly price the term structures of dividends and interest rates. Prices for dividend futures, bonds, and the dividend paying stock are given in closed form. We present an efficient moment based approximation method for option pricing. In a calibration exercise we show that a parsimonious model specification has a good fit with Euribor interest rate swaps and swaptions, Euro Stoxx 50 index dividend futures and dividend options, and Euro Stoxx 50 index options.

1.1 Introduction

In recent years there has been an increasing interest in trading derivative contracts with a direct exposure to dividends. Brennan (1998) argues that a market for dividend derivatives could promote rational pricing in stock markets. In the over-the-counter (OTC) market, dividends have been traded since 2001 in the form of dividend swaps, where the floating leg pays the dividends realized over a predetermined period of time. The OTC market also accommodates a wide variety of more exotic dividend related products such as knock-out dividend swaps, dividend yield swaps and swaptions. Dividend trading gained significant traction in late 2008, when Eurex launched exchange traded futures contracts referencing the dividends paid out by constituents of the Euro Stoxx 50. The creation of a futures market for other major indices (e.g., the FTSE 100 and Nikkei 225) followed shortly after, as well as the introduction of exchange listed options on realized dividends with maturities of up to ten years. Besides the wide variety of relatively new dividend instruments, there is another important dividend derivative that has been around since the inception of finance: a simple dividend paying stock. Indeed, a share of stock includes a claim to all the dividends paid over the stock's lifetime. Any pricing model for dividend derivatives should therefore also be capable of efficiently pricing derivatives on the stock paying the dividends. What's more, the existence of interest rate-dividend hybrid products, the relatively long maturities of dividend options, and the long duration nature of the stock all motivate the use of stochastic interest rates. Despite

its apparent desirability, a tractable joint model for the term structures of interest rates and dividends, and the corresponding stock, has been missing in the literature to date.

We fill this gap and develop an integrated framework to efficiently price derivatives on dividends, stocks, and interest rates. We first specify dynamics for the dividends and discount factor, and in a second step we recover the stock price in closed form as the sum of the fundamental stock price (present value of all future dividends) and possibly a residual bubble component. The instantaneous dividend rate is a linear function of a multivariate factor process. The interest rates are modeled by directly specifying the discount factor to be linear in the factors, similarly as in Filipović et al. (2017). The factor process itself is specified as a general polynomial jump-diffusion, as studied in Filipović and Larsson (2017). Such a specification makes the model tractable because all the conditional moments of the factors are known in closed form. In particular, we have closed form expressions for the stock price and the term structures of dividend futures and interest rate swaps. Any derivative whose discounted payoff can be written as a function of a polynomial in the factors is priced through a moment matching method. Specifically, we find the unique probability density function with maximal Boltzmann-Shannon entropy matching a finite number of moments of the polynomial, as in Mead and Papanicolaou (1984). We then obtain the price of the derivative by numerical integration. In particular, this allows us to price swaptions, dividend options, and options on the dividend paying stock. We show that our polynomial framework also allows to incorporate seasonal behavior in the dividend dynamics.

Within our polynomial framework, we introduce the *linear jump-diffusion* (LJD) model. We show that the LJD model allows for a flexible dependence structure between the factors. This is useful to model a dependence between dividends and interest rates, but also to model the dependence within the term structure of interest rates or dividends. We calibrate a parsimonious specification of the LJD model to market data on Euribor interest rate swaps and swaptions, Euro Stoxx 50 index dividend futures and dividend options, and Euro Stoxx 50 index options. Our model reconciles the relatively high implied volatility of the index options with the relatively low implied volatility of dividend options and swaptions through a negative correlation between dividends and interest rates. The successful calibration of the model to three different classes of derivatives (interest rates, dividends, and equity) illustrates the high degree of flexibility offered by our framework.

This chapter is related to various strands of literature. In the literature on stock option pricing, dividends are often assumed to be either deterministic (e.g., Bos and Vandermark (2002), Bos et al. (2003), Vellekoop and Nieuwenhuis (2006)), a constant fraction of the stock price (e.g., Merton (1973), Korn and Rogers (2005)), or a combination of the two (e.g., Kim (1995), Overhaus et al. (2007)). Geske (1978) and Lioui (2006) model dividends as a stochastic fraction of the stock price. They derive Black-Scholes type of equations for European option prices, however dividends are not guaranteed to be non-negative in both setups. Chance et al. (2002) directly specify log-normal dynamics for the T-forward price of the stock, with T the maturity of the option. Closed form option prices are obtained as in Black (1976), assuming that today's

T-forward price is observable. This approach is easy to use since it does not require any modeling assumptions on the distribution of the dividends. However, it does not produce consistent option prices for different maturities. Bernhart and Mai (2015) take a similar approach, but suggest to fix a time horizon T long enough to encompass all option maturities to be priced. The T-forward price is modeled with a non-negative martingale and the stock price is defined as the T-forward price plus the present value of dividends from now until time T. As a consequence, prices of options with maturity smaller than T will depend on the joint distribution between future dividend payments and the T-forward price, which is not known in general. Bernhart and Mai (2015) resort to numerical tree approximation methods in order to price options. The dependence of their model on a fixed time horizon still leads to time inconsistency, since the horizon will necessarily have to be extended at some point in time. We contribute to this literature by building a stock option pricing model that guarantees non-negative dividends, is time consistent, and remains tractable.

Another strand of literature studies stochastic models to jointly price stock and dividend derivatives. Buehler et al. (2010) assumes that the stock price jumps at known dividend payment dates and follows log-normal dynamics in between the payment dates. The jump amplitudes are driven by an Ornstein-Uhlenbeck process such that the stock price remains log-normally distributed and the model has closed form prices for European call options on the stock. The high volatility in the stock price is reconciled with the low volatility in dividend payments by setting the correlation between the Ornstein-Uhlenbeck process and the stock price extremely negative (-95%). A major downside of the model is that dividends can be negative. Moreover, although the model has a tractable stock price, the dividends themselves are not tractable and Monte-Carlo simulations are required to price the dividend derivatives. In more recent work, Buehler (2015) decomposes the stock price in a fundamental component and a residual bubble component. The dividends are defined as a function of a secondary driving process that mean reverts around the residual bubble component. This model has closed form expressions for dividend futures, but Monte-Carlo simulations are still necessary to price nonlinear derivatives. Guennoun and Henry-Labordère (2017) consider a stochastic local volatility model for the pricing of stock and dividend derivatives. Their model guarantees a perfect fit to observed option prices, however all pricing is based on Monte-Carlo simulations. Tunaru (2018) proposes two different models to value dividend derivatives. The first model is similar to the one of Buehler et al. (2010), but models the jump amplitudes with a beta distribution. This guarantees positive dividend payments. However, the diffusive noise of the stock is assumed independent of the jump amplitudes in order to have tractable expressions for dividend futures prices. Smoothing the dividends through a negative correlation between stock price and jump amplitudes, as in Buehler et al. (2010), is therefore not possible. In a second approach, Tunaru (2018) directly models the cumulative dividends with a diffusive logistic growth process. This process has, however, no guarantee to be monotonically increasing, meaning that negative dividends can occur frequently. Willems (2019b) jointly specifies dynamics for the stock price and the dividend rate such that the stock price is positive and the dividend rate is a non-negative process mean-reverting around a

constant fraction of the stock price. The model of Willems (2019b) is in fact a special case of the general framework introduced in this chapter, although it is different from the LJD model and does not incorporate stochastic interest rates. We add to this literature by allowing for stochastic interest rates, which is important for the valuation of interest rate-dividend hybrid products or long-dated dividend derivatives (e.g., the dividend paying stock). Our model produces closed form prices for dividend futures and features efficient approximations for option prices which are significantly faster than Monte-Carlo simulations. The low volatility in dividends and interest rates is reconciled with the high volatility in the stock price through a negative correlation between dividends and interest rates.

Our work also relates to literature on constructing an integrated framework for dividends and interest rates. Previous approaches were mainly based on affine processes, see e.g. Bekaert and Grenadier (1999), Mamaysky et al. (2002), d'Addona and Kind (2006), Lettau and Wachter (2007, 2011), and Lemke and Werner (2009). In more recent work, Kragt et al. (2018) extract investor information from dividend derivatives by estimating a two-state affine state space model on stock index dividend futures in four different stock markets. Instead of modeling dividends and interest rates separately, they choose to model dividend growth, a risk-free discount rate, and a risk premium in a single variable called the 'discounted risk-adjusted dividend growth rate'. Yan (2014) uses zero-coupon bond prices and present value claims to dividend extracted from the put-call parity relation to estimate an affine term structure model for interest rates and dividends. Suzuki (2014) uses a Nelson-Siegel approach to estimate the fundamental value of the Euro Stoxx 50 using dividend futures and Euribor swap rates. We add to this literature by building an integrated framework for dividends and interest rates using the class of polynomial processes, which contains the traditional affine processes as a special case.

Finally, our work also relates to literature on moment based option pricing. Jarrow and Rudd (1982), Corrado and Su (1996b), and Collin-Dufresne and Goldstein (2002b) use Edgeworth expansions to approximate the density function of the option payoff from the available moments. Closely related are Gram-Charlier expansions, which are used for option pricing for example by Corrado and Su (1996a), Jondeau and Rockinger (2001), and Ackerer et al. (2018). Although these series expansions allow to obtain a function that integrates to one and matches an arbitrary number of moments by construction, it has no guarantee to be positive. In this chapter, we find the unique probability density function with maximal Boltzmann-Shannon entropy. subject to a finite number of moment constraints. Option prices are then obtained by numerical integration. A similar approach is taken by Fusai and Tagliani (2002) to price Asian options. The principle of maximal entropy has also been used to extract the risk-neutral distribution from option prices, see e.g. Buchen and Kelly (1996), Jackwerth and Rubinstein (1996), Avellaneda (1998), and Rompolis (2010). There exist many alternatives to maximizing the entropy in order to find a density function satisfying a finite number of moment constraints. For example, one can maximize the smoothness of the density function (see e.g., Jackwerth and Rubinstein (1996)) or directly maximize (minimize) the option price itself to obtain an upper (lower) bound on the price (see e.g., Lasserre et al. (2006)). A comparison of different

approaches is beyond the scope of this chapter.

The remainder of this chapter is structured as follows. Section 1.2 introduces the factor process and discusses the pricing of dividend futures, bonds, and the dividend paying stock. In Section 1.3 we explain how to efficiently approximate option prices using maximum entropy moment matching. Section 1.4 describes the LJD model. In Section 1.5 we calibrate a parsimonious model specification to real market data. Section 1.6 discusses some extensions of the framework. Section 1.7 concludes. All proofs and technical details can be found in Appendix A.

1.2 Polynomial Framework

We consider a financial market modeled on a filtered probability space $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{Q})$ where $\mathbb{Q}$ is a risk-neutral pricing measure. Henceforth $\mathbb{E}_t[\cdot]$ denotes the $\mathscr{F}_t$-conditional expectation. We model the uncertainty in the economy through a factor process X_t taking values in some state space $E \subseteq \mathbb{R}^d$.[1] We assume that X_t is a polynomial jump-diffusion (cfr. Filipović and Larsson (2017)) with dynamics

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \mathrm{d}M_t, \tag{1.1}$$

for some parameters $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$, and some d-dimensional martingale M_t such that the generator $\mathscr{G}$ of X_t maps polynomials to polynomials of the same degree or less. One of the main features of polynomial jump-diffusions is the fact that they admit closed form conditional moments. For $n \in \mathbb{N}$, denote by $\mathrm{Pol}_n(E)$ the space of of polynomials on E of degree n or less and denote its dimension by N_n.[2] Let $h_1, \ldots, h_{N_n}$ form a polynomial basis for $\mathrm{Pol}_n(E)$ and denote $H_n(x) = (h_1(x), \ldots, h_{N_n}(x))^\top$. Since $\mathscr{G}$ leaves $\mathrm{Pol}_n(E)$ invariant, there exists a unique matrix $G_n \in \mathbb{R}^{N_n\times N_n}$ representing the action of $\mathscr{G}$ on $\mathrm{Pol}_n(E)$ with respect to the basis $H_n(x)$. Without loss of generality we assume to work with the monomial basis.

Example 1.2.1. *If $n = 1$, then we have $H_1(x) = (1, x_1, \ldots, x_d)^\top$ and G_1 becomes*

$$G_1 = \begin{pmatrix} 0 & 0 \\ \kappa\theta & -\kappa \end{pmatrix}. \tag{1.2}$$

From the invariance property of $\mathscr{G}$, one can derive the moment formula (Theorem 2.4 in Filipović and Larsson (2017))

$$\mathbb{E}_t[H_n(X_T)] = \mathrm{e}^{G_n(T-t)} H_n(X_t), \tag{1.3}$$

for all $t \le T$. Many efficient algorithms exist to numerically compute the matrix exponential

[1] We assume that E has non-empty interior.

[2] Since the interior of E is assumed to be non-empty, $\mathrm{Pol}_n(E)$ can be identified with $\mathrm{Pol}_n(\mathbb{R}^d)$ and therefore $N_n = \binom{n+d}{d}$.

(e.g., Al-Mohy and Higham (2011)).

1.2.1 Dividend Futures

Consider a stock that pays a continuous dividend stream to its owner at an instantaneous rate D_t, which varies stochastically over time. We model the cumulative dividend process $C_t = C_0 + \int_0^t D_s \, \mathrm{d}s$ as:

$$C_t = \mathrm{e}^{\beta t} p^\top H_1(X_t), \tag{1.4}$$

for some parameters $\beta \in \mathbb{R}$ and $p \in \mathbb{R}^{d+1}$ such that C_t is a positive, non-decreasing, and absolutely continuous (i.e., drift only) process. This specification for C_t implicitly pins down D_t, which is shown in the following proposition.

Proposition 1.2.2. *The instantaneous dividend rate D_t implied by* (1.4) *is given by*

$$D_t = \mathrm{e}^{\beta t} p^\top (\beta \mathrm{Id} + G_1) H_1(X_t), \tag{1.5}$$

where Id *denotes the identity matrix.*

Remark that both the instantaneous dividend rate and the cumulative dividends load linearly on the factor process. The exponential scaling of C_t with parameter β can be helpful to guarantee a non-negative instantaneous dividend rate. Indeed, if

$$\lambda = \sup_{x \in E} -\frac{p^\top G_1 H_1(x)}{p^\top H_1(x)} \tag{1.6}$$

is finite, then it follows from (1.5) that $D_t \geq 0$ if and only if $\beta \geq \lambda$.[3] Moreover, when all eigenvalues of κ have positive real parts, it follows from the moment formula (1.3) that

$$\lim_{T \to \infty} \frac{1}{T-t} \log\left(\frac{\mathbb{E}_t[D_T]}{D_t}\right) = \beta.$$

The parameter β therefore controls the asymptotic risk-neutral expected growth rate of the dividends.

The time-t price of a continuously marked-to-market futures contract referencing the dividends to be paid over a future time interval $[T_1, T_2]$, $t \leq T_1 \leq T_2$, is given by:

$$\begin{aligned} D_{fut}(t, T_1, T_2) &= \mathbb{E}_t\left[\int_{T_1}^{T_2} D_s \, \mathrm{d}s\right] \\ &= \mathbb{E}_t\left[C_{T_2} - C_{T_1}\right] \\ &= p^\top \left(\mathrm{e}^{\beta T_2} \mathrm{e}^{G_1(T_2 - t)} - \mathrm{e}^{\beta T_1} \mathrm{e}^{G_1(T_1 - t)}\right) H_1(X_t), \end{aligned} \tag{1.7}$$

where we have used the moment formula (1.3) in the last equality. Hence, the dividend futures

[3] We calculate λ explicitly for the linear jump-diffusion model studied in Section 1.4.

price is linear in the factor process. Note that the dividend futures term structure (i.e., the dividend futures prices for varying T_1 and T_2) does not depend on the specification of the martingale part of X_t.

1.2.2 Bonds and Swaps

Denote the risk-neutral discount factor by ζ_t. It is related to the short rate r_t as follows

$$\zeta_T = \zeta_t \mathrm{e}^{-\int_t^T r_s \mathrm{d}s}, \quad 0 \le t \le T.$$

We directly specify dynamics for the risk-neutral discount factor:

$$\zeta_t = \mathrm{e}^{-\gamma t} q^\top H_1(X_t), \tag{1.8}$$

for some parameters $\gamma \in \mathbb{R}$ and $q \in \mathbb{R}^{d+1}$ such that ζ_t is a positive and absolutely continuous process. This is similar to the specification (1.4) of C_t but, in order to allow for negative interest rates, we do not require ζ_t to be monotonic (non-increasing). Filipović et al. (2017) follow a similar approach and specify linear dynamics for the state price density with respect to the historical probability measure $\mathbb{P}$. Their specification pins down the market price of risk. It turns out that the polynomial property of the factor process is not preserved under the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ in this case. However, as seen in (1.7), the polynomial property (in particular the linear drift) under $\mathbb{Q}$ is important for pricing the dividend futures contracts.

The time-t price of a zero-coupon bond paying one unit of currency at time $T \ge t$ is given by:

$$P(t,T) = \frac{1}{\zeta_t} \mathbb{E}_t\left[\zeta_T\right].$$

Using the moment formula (1.3) we get a linear-rational expression for the zero-coupon bond price

$$P(t,T) = \mathrm{e}^{-\gamma(T-t)} \frac{q^\top \mathrm{e}^{G_1(T-t)} H_1(X_t)}{q^\top H_1(X_t)}. \tag{1.9}$$

Remark that the term structure of zero-coupon bond prices depends only on the drift of X_t. Similarly as in Filipović et al. (2017), one can introduce exogenous factors feeding into the martingale part of X_t to generate unspanned stochastic volatility (see e.g., Collin-Dufresne and Goldstein (2002a)), however we do not consider this.

Using the relation $r_t = -\partial_T \log P(t,T)|_{T=t}$, we obtain the following linear-rational expression for the short rate:

$$r_t = \gamma - \frac{q^\top G_1 H_1(X_t)}{q^\top H_1(X_t)}.$$

When all eigenvalues of κ have positive real parts, it follows that

$$\lim_{T\to\infty} -\frac{\log(P(t,T))}{T-t} = \gamma,$$

so that γ can be interpreted as the yield on the zero-coupon bond with infinite maturity.

Ignoring differences in liquidity and credit characteristics between discount rates and IBOR rates, we can value swap contracts as linear combinations of zero-coupon bond prices. The time-t value of a payer interest rate swap with first reset date $T_0 \geq t$, fixed leg payment dates $T_1 < \cdots < T_n$, and fixed rate K is given by:

$$\pi_t^{swap} = P(t,T_0) - P(t,T_n) - K\sum_{k=1}^{n} \delta_k P(t,T_k), \tag{1.10}$$

with $\delta_k = T_k - T_{k-1}$, $k = 1\ldots,n$. The forward swap rate is defined as the fixed rate K which makes the right hand side of (1.10) equal to zero. Note that the discounted swap value $\zeta_t \pi_t^{swap}$ becomes a linear function of X_t, which will be important for the purpose of pricing swaptions.

1.2.3 Dividend Paying Stock

Denote by S_t^* the *fundamental price* of the stock, which we define as the present value of all future dividends:

$$S_t^* = \frac{1}{\zeta_t}\mathbb{E}_t\left[\int_t^\infty \zeta_s D_s \,\mathrm{d}s\right]. \tag{1.11}$$

In order for S_t^* to be finite in our model, we must impose parameter restrictions. The following proposition provides sufficient conditions on the parameters, together with a closed form expression for S_t^*. The latter is derived using the fact that $\zeta_t D_t$ is quadratic in X_t, hence we are able to calculate its conditional expectation through the moment formula (1.3).

Proposition 1.2.3. *If the real parts of the eigenvalues of G_2 are bounded above by $\gamma - \beta$, then S_t^* is finite and given by*

$$S_t^* = \mathrm{e}^{\beta t}\frac{w^\top H_2(X_t)}{q^\top H_1(X_t)}, \tag{1.12}$$

where $w = \left[(\gamma-\beta)\,\mathrm{Id} - G_2^\top\right]^{-1} v$ and $v \in \mathbb{R}^{N_2}$ is the unique coordinate vector satisfying

$$v^\top H_2(x) = p^\top(\beta\mathrm{Id} + G_1)H_1(x)\, q^\top H_1(x).$$

Proposition 1.2.3 shows that the discounted fundamental stock price $\zeta_t S_t^*$ is quadratic in X_t, which means in particular that we have all moments of $\zeta_t S_t^*$ in closed form. Loosely speaking, the fundamental stock price will be finite as long as the dividends are discounted at a sufficiently high rate (by choosing γ sufficiently large). Henceforth we will assume that the assumption of Proposition 1.2.3 is satisfied.

The following proposition shows how the price of the dividend paying stock, which we denote by S_t, is related to the fundamental stock price.[4]

[4] This relationship has been highlighted in particular by Buehler (2010, 2015) in the context of derivative pricing.

Proposition 1.2.4. *The market is arbitrage free if and only if S_t is of the form*

$$S_t = S_t^* + \frac{L_t}{\zeta_t}, \tag{1.13}$$

with L_t a non-negative local martingale.

The process L_t can be interpreted as a bubble in the sense that it drives a wedge between the fundamental stock price and the observed stock price. If X_t is continuous, then applying Itô's lemma to (1.13) and using the fact that ζ_t is assumed to be absolutely continuous, we obtain the following risk-neutral stock price dynamics

$$\mathrm{d}S_t = (r_t S_t - D_t)\,\mathrm{d}t + \mathrm{e}^{\beta t}\frac{w^\top \mathrm{J}_{H_2}(X_t)}{q^\top H_1(X_t)}\,\mathrm{d}M_t + \frac{1}{\zeta_t}\mathrm{d}L_t, \tag{1.14}$$

where $\mathrm{J}_{H_2}(x)$ denotes the Jacobian of $H_2(x)$.[5] Remark that S_t has the correct risk-neutral drift, by construction. Given dynamics for r_t and D_t, an alternative approach to model S_t for derivative pricing purposes would have been to directly specify its martingale part, see for instance Willems (2019b). With such an approach, however, it is not straightforward to guarantee a positive stock price. Indeed, the downward drift of the instantaneous dividend rate could push the stock price in negative territory.[6] Moreover, it is clear that by directly specifying the martingale part of the stock price, we risk implicitly modeling a bubble in the stock price. In contrast, our approach implies a martingale part (the second term in (1.14)) that guarantees a positive stock price. This martingale part is completely determined by the given specification for dividends and interest rates. In case this is too restrictive for the stock price dynamics, one can always adjust accordingly through the specification of the non-negative local martingale L_t. For example, Buehler (2015) considers a local volatility model on top of the fundamental stock price that is separately calibrated to equity option prices.

Remark 1.2.5. *Bubbles are usually associated with strict local martingales, see e.g. Cox and Hobson (2005). In fact, for economies with a finite time horizon, a bubble is only possible if the deflated gains process is a strict local martingale, which corresponds to a bubble of Type 3 according to the classification of Jarrow et al. (2007). For economies with an infinite time horizon, which is the case in our setup, bubbles are possible even if the deflated gains process is a true martingale. Such bubbles are of Type 1 and 2 in the classification Jarrow et al. (2007). Specifically, a (uniformly integrable) martingale L_t corresponds to a bubble of Type 2 (Type 1).*

We finish this section with a result on the duration of the stock. We define the stock duration as

$$Dur_t = \frac{\int_t^\infty (s-t)\,\mathbb{E}_t[\zeta_s D_s]\,\mathrm{d}s}{\zeta_t S_t^*}. \tag{1.15}$$

[5] A similar, but lengthier, expression can be derived in case there are jumps in X_t. We choose to omit it since it does not add much value to the discussion that follows.

[6] Instead of starting from dynamics for D_t, we could have specified dynamics for the dividend yield D_t/S_t. This would help to keep the stock price positive, but it does typically not produce a tractable distribution for D_t. This is problematic since dividend derivatives reference notional dividend payments paid out over a certain time period.

The stock duration represents a weighted average of the time an investor has to wait to receive his dividends, where the weights are the relative contribution of the present value of the dividends to the fundamental stock price. This definition is the continuous time version of the one used by Dechow et al. (2004) and Weber (2018). The following proposition gives a closed form expression for stock duration in our framework.

Proposition 1.2.6. *The stock duration is given by*

$$Dur_t = \frac{w^\top \left[(\gamma-\beta)\,\mathrm{Id} - G_2\right]^{-1} H_2(X_t)}{w^\top H_2(X_t)}. \tag{1.16}$$

1.3 Option Pricing

In this section we address the problem of pricing derivatives with discounted payoff functions that are not polynomials in the factor process. The polynomial framework no longer allows to price such derivatives in closed form. However, we can accurately approximate the prices using the available moments of the factor process.

1.3.1 Maximum Entropy Moment Matching

In all examples encountered below, we consider a derivative maturing at time T whose discounted payoff is given by $F(g(X_T))$, for some $g \in \mathrm{Pol}_n(E)$, $n \in \mathbb{N}$, and some function $F\colon \mathbb{R} \to \mathbb{R}$. The time-$t$ price π_t of this derivative is given by

$$\pi_t = \mathbb{E}_t\left[F\big(g(X_T)\big)\right]. \tag{1.17}$$

If the conditional distribution of the random variable $g(X_T)$ were available in closed form, we could compute π_t by integrating F over the real line. In general, however, we are only given all the conditional moments of the random variable $g(X_T)$. We thus aim to construct an approximative probability density function f matching a finite number of these moments. In a second step we approximate the option price through numerically integrating F with respect to f. Given that a function is an infinite dimensional object, finding such a function f is clearly an underdetermined problem and we need to introduce additional criteria to pin down one particular function. A popular choice in the engineering and physics literature is to choose the density function with maximum entropy:

$$\begin{aligned} \max_f \quad & -\int_R f(x)\ln f(x)\,\mathrm{d}x \\ \text{s.t.} \quad & \int_R x^n f(x)\,\mathrm{d}x = M_n, \qquad n = 0,\dots,N, \end{aligned} \tag{1.18}$$

where $R \subseteq \mathbb{R}$ denotes the support and $M_0 = 1, M_1, \dots, M_N$ denote the first $N+1$ moments of $g(X_T)$. Jaynes (1957) motivates such a choice by noting that maximizing entropy incorporates

the least amount of prior information in the distribution, other than the imposed moment constraints. In this sense it is maximally noncommittal with respect to unknown information about the distribution.

Straightforward functional variation with respect to f gives the following solution to this optimization problem:

$$f(x) = \exp\left(-\sum_{i=0}^{N} \lambda_i x^i\right), \quad x \in R,$$

where the Lagrange multipliers $\lambda_0, \ldots, \lambda_N$ have to be solved from the moment constraints:

$$\int_R x^n \exp\left(-\sum_{i=0}^{N} \lambda_i x^i\right) \mathrm{d}x = M_n, \quad n = 0, \ldots, N. \tag{1.19}$$

If $N = 0$ and $R = [0,1]$, then we recover the uniform distribution. For $N = 1$ and $R = (0,\infty)$ we obtain the exponential distribution, while for $N = 2$ and $R = \mathbb{R}$ we obtain the Gaussian distribution. For $N \geq 3$, one needs to solve the system in (1.19) numerically, which involves evaluating the integrals numerically.[7] We refer to the existing literature for more details on the implementation of maximum entropy densities, see e.g. Agmon et al. (1979), Mead and Papanicolaou (1984), Rockinger and Jondeau (2002), and Holly et al. (2011).

Remark 1.3.1. *By subsequently combining the law of iterated expectations and the moment formula* (1.3)*, we are also able to compute the conditional moments of the finite dimensional distributions of X_t. In particular, the method described in this section can also be applied to price path-dependent derivatives whose discounted payoff depends on the factor process at a finite number of future time points. One example of such products are the dividend options, which will be discussed below.*

[7] Directly trying to find the roots of this system might not lead to satisfactory results. A more stable numerical procedure is obtained by introducing the following potential function: $P(\lambda_0, \ldots, \lambda_N) = \int_R \exp(-\sum_{i=0}^{N} \lambda_i x^i)\,\mathrm{d}x + \sum_{i=0}^{N} \lambda_i M_i$. This function can easily be shown to be everywhere convex (see e.g., Mead and Papanicolaou (1984)) and its gradient corresponds to the vector of moment conditions in (1.19). In other words, the Lagrange multipliers can be found by minimizing the potential function $P(\lambda_0, \ldots, \lambda_N)$. This is an unconstrained convex optimization problem where we have closed form (up to numerical integration) expressions for the gradient and hessian, which makes it a prototype problem to be solved with Newton's method.

1.3.2 Swaptions, Stock and Dividend Options

The time-t price $\pi_t^{swaption}$ of a payer swaption with expiry date T_0, which gives the owner the right to enter into a (spot starting) payer swap at T_0, is given by:

$$\begin{aligned}\pi_t^{swaption} &= \frac{1}{\zeta_t}\mathbb{E}_t\left[\zeta_{T_0}\left(\pi_{T_0}^{swap}\right)^+\right]\\ &= \frac{1}{\zeta_t}\mathbb{E}_t\left[\left(\zeta_{T_0}-\zeta_{T_0}P(T_0,T_n)-K\sum_{k=1}^{n}\delta_k\zeta_{T_0}P(T_0,T_k)\right)^+\right]\\ &= \frac{\mathrm{e}^{-\gamma(T_0-t)}}{q^\top H_1(X_t)}\mathbb{E}_t\left[\left(q^\top\left(\mathrm{Id}-\mathrm{e}^{(G_1-\gamma\mathrm{Id})(T_n-T_0)}-K\sum_{k=1}^{n}\delta_k\mathrm{e}^{(G_1-\gamma\mathrm{Id})(T_k-T_0)}\right)H_1(X_{T_0})\right)^+\right],\end{aligned}$$

where we have used (1.9) in the last equality. Observe that the discounted payoff of the swaption is of the form in (1.17) with $F(\cdot)=\max(\cdot,0)$ and g is a polynomial of degree one in X_{T_0}.

The time-t price π_t^{stock} of a European call option on the dividend paying stock with strike K and expiry date T is given by

$$\begin{aligned}\pi_t^{stock} &= \frac{1}{\zeta_t}\mathbb{E}_t\left[\zeta_T(S_T-K)^+\right]\\ &= \frac{1}{\zeta_t}\mathbb{E}_t\left[(L_T+\zeta_T S_T^*-\zeta_T K)^+\right]\\ &= \frac{\mathrm{e}^{-\gamma(T-t)}}{q^\top H_1(X_t)}\mathbb{E}_t\left[\left(\mathrm{e}^{\gamma T}L_T+\mathrm{e}^{\beta T}w^\top H_2(X_T)-q^\top H_1(X_T)K\right)^+\right],\end{aligned}\tag{1.20}$$

where we have used (1.12) in the last equality. If (L_t,X_t) is jointly a polynomial jump-diffusion, we can compute all moments of the random variable $\mathrm{e}^{\gamma T}L_T+\mathrm{e}^{\beta T}w^\top H_2(X_T)-q^\top H_1(X_T)K$ and proceed as explained in Section 1.3.1.

Remark 1.3.2. *If one assumes independence between the processes L_t and X_t, then the assumption that (L_t,X_t) must jointly be a polynomial jump-diffusion is not necessarily needed. Indeed, suppose L_t is specified such that we can compute $F(k)=\mathrm{e}^{-\gamma(T-t)}\mathbb{E}_t[(\mathrm{e}^{\gamma T}L_T-k)^+]$ efficiently. By the law of iterated expectations we have*

$$\pi_t^{stock}=\frac{\mathbb{E}_t\left[F(g(X_T))\right]}{q^\top H_1(X_t)},$$

where we define $g(x)=-\mathrm{e}^{\beta T}w^\top H_2(x)+q^\top H_1(x)K\in\mathrm{Pol}_2(E)$. The numerator in the above expression is now of the form in (1.17) *and we proceed as before.*

Consider next a European call option on the dividends realized in $[T_1,T_2]$, expiry date T_2, and strike price K. This type of options are actively traded on the Eurex exchange where the Euro

Stoxx 50 dividends serve as underlying. The time-t price π_t^{div} of this product is given by

$$\begin{aligned}\pi_t^{div} &= \frac{1}{\zeta_t}\mathbb{E}_t\left[\zeta_{T_2}\left(\int_{T_1}^{T_2} D_s\,\mathrm{d}s - K\right)^+\right] \\ &= \frac{1}{\zeta_t}\mathbb{E}_t\left[\left(\zeta_{T_2}(C_{T_2} - C_{T_1} - K)\right)^+\right] \\ &= \frac{\mathrm{e}^{-\gamma(T_2-t)}}{q^\top H_1(X_t)}\mathbb{E}_t\left[\left(q^\top H_1(X_{T_1})\left(\mathrm{e}^{\beta T_2} p^\top H_1(X_{T_2}) - \mathrm{e}^{\beta T_1} p^\top H_1(X_{T_1}) - K\right)\right)^+\right].\end{aligned}$$

We can compute in closed form all the moments of the scalar random variable

$$q^\top H_1(X_{T_2})\left(\mathrm{e}^{\beta T_2} p^\top H_1(X_{T_2}) - \mathrm{e}^{\beta T_1} p^\top H_1(X_{T_1}) - K\right)$$

by subsequently applying the law of iterated expectations and the moment formula (1.3), see Remark 1.3.1. We then proceed as before by finding the maximum entropy density corresponding to these moments and computing the option price by numerical integration.

1.4 The Linear Jump-Diffusion Model

In this section we give a worked-out example of a factor process that fits in the polynomial framework of Section 1.2. In the following, if $x \in \mathbb{R}^d$ then $\mathrm{diag}(x)$ denotes the diagonal matrix with $x_1, \ldots, x_d$ on its diagonal. If $x \in \mathbb{R}^{d\times d}$, then we denote $\mathrm{diag}(x) = (x_{11}, \ldots, x_{dd})^\top$.

The *linear jump-diffusion* (LJD) model assumes the following dynamics for the factor process

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \mathrm{diag}(X_{t-})\,(\Sigma\,\mathrm{d}B_t + \mathrm{d}J_t)\,, \tag{1.21}$$

where B_t is a standard d-dimensional Brownian motion, $\Sigma \in \mathbb{R}^{d\times d}$ is a lower triangular matrix with non-negative entries on its main diagonal, J_t is a compensated compound Poisson process with arrival intensity $\xi \geq 0$ and a jump distribution $F(\mathrm{d}z)$ that admits moments of all orders.[8] Both the jump amplitudes and the Poisson jumps are assumed to be independent from the diffusive noise. The purely diffusive LJD specification (i.e., $\xi = 0$) has appeared in various financial contexts such as stochastic volatility (Nelson (1990), Barone-Adesi et al. (2005)), energy markets (Pilipović (1997)), interest rates (Brennan and Schwartz (1979)), and Asian option pricing (Linetsky (2004), Willems (2019a)). The extension with jumps has not received much attention yet.

The following proposition verifies that X_t is indeed a polynomial jump-diffusion and also shows how to choose parameters such that X_t has positive components.

Proposition 1.4.1. *Assume that matrix κ has non-positive off-diagonal elements, $(\kappa\theta)_i \geq 0$, $i = 1, \ldots, d$, and F has support $\mathscr{S} \subseteq (-1, \infty)^d$. Then for every initial value $X_0 \in (0,\infty)^d$ there*

[8] For simplicity we assume a compound Poisson process with a single jump intensity, however this can be generalized (see Filipović and Larsson (2017)).

exists a unique strong solution X_t to (1.21) *with values in* $(0,\infty)^d$*. Moreover, X_t is a polynomial jump-diffusion.*

We will henceforth assume that the assumptions of Proposition 1.4.1 are satisfied, as it allows to derive parameter restrictions to guarantee $C_t > 0$, $\zeta_t > 0$, and $D_t \geq 0$. In order to have $p^\top H_1(x) > 0$ and $q^\top H_1(x) > 0$ for all $x \in (0,\infty)^d$, the vectors p and q must have non-negative components with at least one component different from zero. The following proposition introduces a lower bound on β such that $D_t \geq 0$.

Proposition 1.4.2. *Let $p = (p_0, p_1, \ldots, p_d)^\top \in [0,\infty)^{1+d}$ and denote $\tilde{p} = (p_1, \ldots, p_d)^\top$. Assume that at least one of the $p_1, \ldots, p_d$ is non-zero, so that dividends are not deterministic. Without loss of generality we assume $p_1, \ldots, p_k > 0$ and $p_{k+1}, \ldots, p_d = 0$, for some $1 \leq k \leq d$. If we denote by κ_j the j-th column of κ, then we have $D_t \geq 0$ if and only if*

$$\beta \geq \begin{cases} \max\left\{\dfrac{\tilde{p}^\top \kappa_1}{p_1}, \ldots, \dfrac{\tilde{p}^\top \kappa_k}{p_k}\right\} & \text{if } p_0 = 0, \\ \max\left\{-\dfrac{\tilde{p}^\top \kappa \theta}{p_0}, \dfrac{\tilde{p}^\top \kappa_1}{p_1}, \ldots, \dfrac{\tilde{p}^\top \kappa_k}{p_k}\right\} & \text{if } p_0 > 0. \end{cases} \tag{1.22}$$

The LJD model allows a flexible instantaneous correlation structure between the factors through the matrix Σ. This is in contrast to non-negative affine jump-diffusions, a popular choice in term structure modeling when non-negative factors are required, see, e.g., Duffie et al. (2003). Indeed, as soon as one introduces a non-zero instantaneous correlation between the factors of a non-negative affine jump-diffusion, the affine (and polynomial) property is lost. Correlation between factors can be used to incorporate a dependence between the term structures of interest rates and dividends, but also to model a dependence within a single term structure. The LJD model also allows for state-dependent, positive and negative, jump sizes of the factors. This again is in contrast to non-negative affine jump-diffusions.

The following proposition provides the eigenvalues of the corresponding matrix G_2 under the assumption of a triangular form for κ. Combined with Proposition 1.2.3, this gives sufficient conditions to guarantee a finite stock price in the LJD model.

Proposition 1.4.3. *If κ is a triangular matrix, then the eigenvalues of the matrix G_2 are*

$$0, -\kappa_{11}, \ldots, -\kappa_{dd},$$
$$-\kappa_{ii} - \kappa_{jj} + (\Sigma\Sigma^\top)_{ij} + \xi \int_{\mathscr{S}} z_i z_j \, F(\mathrm{d}z), \quad 1 \leq i, j \leq d.$$

The eigenvalues of G_1 coincide with the values on the first line.

1.5 Numerical Study

In this section we calibrate a parsimonious LJD model specification to market data on weekly intervals (Wednesday to Wednesday) from May 2010 until December 2015. All the data is obtained from Bloomberg. On every day of the sample we minimize the squared difference between the model implied and market observed prices. The initial values of the factor process are included as free parameters, which brings the total number of parameters to be calibrated to 12. For the optimization we use the Nelder-Mead simplex algorithm. We use the outcome of every optimization as initial guess for the optimization on the next sample day.

1.5.1 Data Description

The dividend paying stock in our calibration study is the Euro Stoxx 50, the leading blue-chip stock index in the Eurozone. The index is composed of fifty stocks of sector leading companies from twelve Eurozone countries. We choose to focus on the European market because the dividend futures contracts on the Euro Stoxx 50 are the most liquid in the world and have been around longer than in any other market. Kragt et al. (2018) report an average daily turnover of more than EUR 150 million for all expiries combined, with the majority of the liquidity in the first five expiries. The Euro Stoxx 50 dividend futures contracts are traded on Eurex and reference the sum of the declared ordinary gross cash dividends (or cash-equivalent, e.g. stock dividends) on index constituents that go ex-dividend during a given calendar year, divided by the index divisor on the ex-dividend day. Corporate actions that cause a change in the index divisor are excluded from the dividend calculations, e.g. special and extraordinary dividends, return of capital, stock splits, etc. One every day of the sample there are ten annual contracts available for trading with maturity dates on the third Friday of December. For example, on September 1 2015, the k-th to expire contract, $k = 1,\ldots,10$, references the dividends paid between the third Friday of December $2014 + k - 1$ and the third Friday of December $2014 + k$. We interpolate adjacent dividend futures contracts using the approach of Kragt et al. (2018) to construct contracts with a constant time to maturity of 1 to 9 years.[9] In the calibration we use the contracts with maturities in 1, 2, 3, 5, 7, and 9 years, the remaining ones will be used for an out-of-sample exercise. Figure 1.1a plots the interpolated dividend futures prices with 1, 2, 5, and 9 years to maturity.

Next to the Euro Stoxx 50 dividend futures contracts, there also exist exchanged traded options on realized dividends. The maturity dates and the referenced dividends of the options coincide with those of the corresponding futures contracts. At every calibration date, we consider the Black (1976) implied volatility of an at-the-money (ATM) dividend option with 2 years to maturity. Since dividend option contracts have fixed maturity dates, we interpolate the implied volatility of the second and third to expire ATM option contract.[10] Figure 1.1b plots

[9] We could also calibrate the model without doing any interpolation of the data. However, in order to make the fitting errors of the sequential calibrations more comparable over time, we choose to interpolate all instruments such that they have a constant time to maturity.

[10] We linearly interpolate the total implied variance $\sigma^2_{Black}\tau$, where σ_{Black} denotes the implied volatility and τ

the implied volatilities of the dividend options over time.

The term structure of interest rates is calibrated to European spot-starting swap contracts referencing the six month Euro Interbank Offered Rate (Euribor) with tenors of 1, 2, 3, 5, 7, 10, and 20 years. Figure 1.1c plots the par swap rates of swaps with tenors of 1, 5, 10, and 20 years (other tenors have been left out of the plot for clarity). In addition, we also include ATM swaptions with time to maturity equal to 3 months and underlying swap with tenor 10 years. These are among the most liquid fixed-income instruments in the European market. The swaptions are quoted in terms of normal implied volatility and are plotted in Figure 1.1d.

We also consider Euro Stoxx 50 index options with ATM strike and a maturity of 3 months. Their prices are quoted in terms of Black-Scholes implied volatility and plotted in Figure 1.1b together with the dividend options implied volatility. Figure 1.1e plots the Euro Stoxx 50 index level over time.

1.5.2 Model Specification

We propose a parsimonious four-factor LJD specification without jumps for $X_t = (X^I_{0t}, X^I_{1t}, X^D_{0t}, X^D_{1t})^\top$

$$\begin{cases} \mathrm{d}X^I_{0t} &= \kappa^I_0 \left(X^I_{1t} - X^I_{0t}\right) \mathrm{d}t \\ \mathrm{d}X^I_{1t} &= \kappa^I_1 (\theta^I - X^I_{1t}) \,\mathrm{d}t \quad + \quad \sigma^I X^I_{1t} \,\mathrm{d}B_{1t} \\ \mathrm{d}X^D_{0t} &= \kappa^D_0 \left(X^D_{1t} - X^D_{0t}\right) \mathrm{d}t \\ \mathrm{d}X^D_{1t} &= \kappa^D_1 (\theta^D - X^D_{1t}) \,\mathrm{d}t \quad + \quad \sigma^D X^D_{1t} \left(\rho \,\mathrm{d}B_{1t} + \sqrt{1-\rho^2}\,\mathrm{d}B_{2t}\right) \end{cases}, \tag{1.23}$$

with $\rho \in [-1,1]$, $\kappa^I_0, \kappa^D_0, \kappa^I_1, \kappa^D_1, \theta^I, \theta^D, \sigma^I, \sigma^D > 0$, and $X_0 \in (0,\infty)^4$. By Proposition 1.4.1, X_t takes values in $(0,\infty)^4$. Since we only include options with ATM strike in the calibration, we choose not to include any jumps in the dynamics in order to keep the number of parameters small. We define the cumulative dividend process as

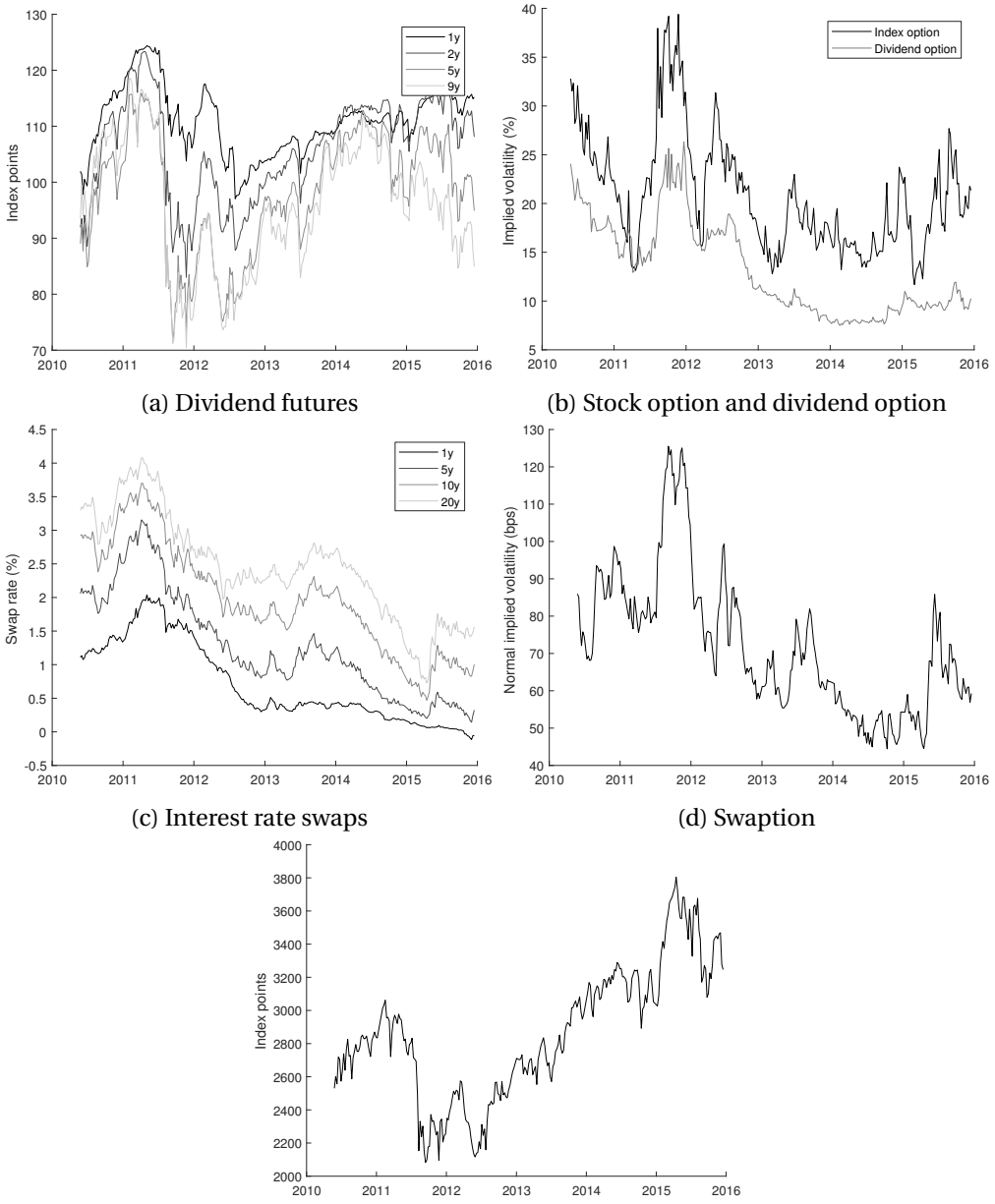
$$C_t = \mathrm{e}^{\beta t} X^D_{0t},$$

so that X^D_{0t} and X^D_{1t} are driving the term structure of dividends. The corresponding instantaneous dividend rate becomes

$$D_t = \mathrm{e}^{\beta t} \left(\left(\beta - \kappa^D_0\right) X^D_{0t} + \kappa^D_0 X^D_{1t} \right).$$

Using Proposition 1.4.2, we guarantee $D_t \geq 0$ by requiring $\beta \geq \kappa^D_0$. In order to further reduce the number of parameters, we set $\beta = \kappa^D_0$, so that $D_t = \mathrm{e}^{\beta t} \beta X^D_{1t}$ and X^D_{0t} no longer enters in the dynamics of D_t. We can thus normalize $C_0 = X^D_{00} = 1$.

the maturity of the option.



(a) Dividend futures

(b) Stock option and dividend option

(c) Interest rate swaps

(d) Swaption

(e) Index level

Figure 1.1 – Data used in the calibration exercise. Dates range from May 2010 until December 2015 at a weekly frequency. Figure 1.1a shows the interpolated Euro Stoxx 50 dividend futures prices with a constant time to maturity of 1, 2, 5, and 9 years. The contracts with time to maturity of 3, 4, and 7 years are not plotted for clarity. Figure 1.1c shows the par swap rate of Euribor spot starting swaps with tenors 1, 5, 10, and 20 years. The swap rates with tenors 2, 3, 4, and 7 years are not plotted for clarity. Figure 1.1b shows the Black-Scholes and Black implied volatility, respectively, of ATM Euro Stoxx 50 index and dividend options. The stock option has a time to maturity of 3 months and the dividend option 2 years. Figure 1.1d shows the normal implied volatility of swaptions with time to maturity 3 months and the underlying swap has a tenor of 10 years. Figure 1.1e shows the level of the Euro Stoxx 50.

The discount factor process is defined as

$$\zeta_t = \mathrm{e}^{-\gamma t} X^I_{0t},$$

so that X^I_{0t} and X^I_{1t} are driving the term structure of interest rates. The corresponding short rate becomes

$$r_t = (\gamma + \kappa^I_0) - \kappa^I_0 \frac{X^I_{1t}}{X^I_{0t}},$$

which is unbounded from below and bounded above by $\gamma + \kappa^I_0$.[11] This allows to capture the negative interest rates that occur in the sample. Dividing ζ_t by a positive constant does not affect model prices, so for identification purposes we normalize $\theta^I = 1$.[12]

The matrix κ is upper triangular and given by

$$\kappa = \begin{pmatrix} \kappa^I_0 & -\kappa^I_0 & 0 & 0 \\ 0 & \kappa^I_1 & 0 & 0 \\ 0 & 0 & \kappa^D_0 & -\kappa^D_0 \\ 0 & 0 & 0 & \kappa^D_1 \end{pmatrix}.$$

The diagonal elements, which coincide with the eigenvalues, of κ are all positive by assumption. We can therefore interpret γ as the asymptotic zero-coupon bond yield and β as the asymptotic risk-neutral expected dividend growth rate. Using Propositions 1.2.3 and 1.4.3, we introduce the following constraint on the model parameters in order to guarantee a finite stock price:

$$\gamma - \beta > \max\left\{0\,,\, (\sigma^I)^2 - 2\kappa^I_1\,,\, (\sigma^D)^2 - 2\kappa^D_1\,,\, \sigma^I\sigma^D\rho - \kappa^I_1 - \kappa^D_1\right\}.$$

The parameter $\rho \in [-1,1]$ controls the correlation between interest rates and dividends. Specifically, the instantaneous correlation between the dividend rate and the short rate is given by

$$\frac{\mathrm{d}[D,r]_t}{\sqrt{\mathrm{d}[D,D]_t}\sqrt{\mathrm{d}[r,r]_t}} = -\rho, \tag{1.24}$$

where $[\cdot,\cdot]_t$ denotes the quadratic covariation. The minus sign in front of ρ appears because the Brownian motion B_{1t} drives the discount factor, which is negatively related to the short rate.

[11] In the more general polynomial framework described in Section 1.2, it is possible to lower bound the short rate. For example, one can use compactly supported polynomial processes, similarly as in Ackerer and Filipović (2019a).

[12] For a constant $k > 0$, the dynamics of $(\tilde{X}^I_{0t}, \tilde{X}^I_{1t}) := (kX^I_{0t}, kX^I_{1t})$ is given by

$$\begin{cases} \mathrm{d}\tilde{X}^I_{0t} &= \kappa^I_0\left(\tilde{X}^I_{1t} - \tilde{X}^I_{0t}\right)\mathrm{d}t \\ \mathrm{d}\tilde{X}^I_{1t} &= \kappa^I_1(\tilde{\theta}^I - \tilde{X}^I_{1t})\,\mathrm{d}t \quad + \quad \sigma^I\tilde{X}^I_{1t}\,\mathrm{d}B_{1t} \end{cases},$$

with $\tilde{\theta}^I := k\theta^I$. The dynamics of $(\tilde{X}^I_{0t}, \tilde{X}^I_{1t})$ is therefore of the same form as that of (X^I_{0t}, X^I_{1t}).

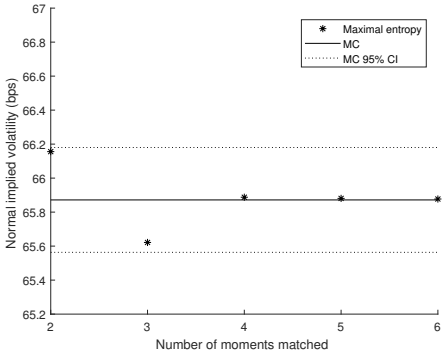
1.5.3 Calibration Results

Although the option pricing technique described in Seciton 1.3.1 works in theory for any finite number of moment constraints, there is a computational cost associated with computing the moments on the one hand, and solving the Lagrange multipliers on the other hand. In the calibration, we use moments up to order four to price swaptions, dividend options, and stock options. The number of moments needed for an accurate option price depends on the specific form of the payoff function and on the model parameters. As an example, Figure 1.2 shows prices of a swaption, dividend option, and stock option for different number of moments matched and using the calibrated parameters from an arbitrary day in the sample. As a benchmark, we perform a Monte-Carlo simulation of the model. We discretize (1.23) at a weekly frequency with a simple Euler scheme and simulate 10^5 trajectories.[13] For all three options, the maximum entropy method based on the first four moments produces an approximation within the 95%-confidence interval of the Monte-Carlo simulation.

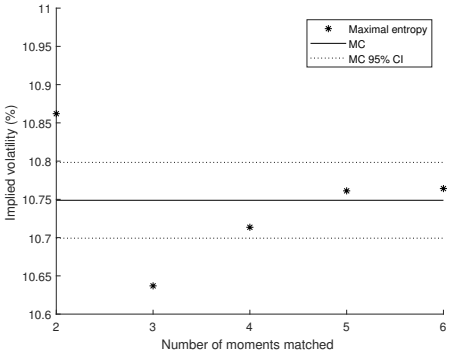
Table 1.1 shows the absolute pricing error over the sample period. Considering the relatively small number of parameters, the fit is surprisingly good. Dividend futures have a mean absolute relative error between 1 and 4%. The mean absolute error of the swap rates is in the order of basis points for all tenors. The fit with the dividend option, swaption, and stock option implied volatilities is close to perfect with a mean absolute pricing error of less than three basis points. The Eurostoxx 50 index level is matched with a mean relative error of less 0.1%. Figure 1.3 shows the evolution of the errors over time. The largest errors for the dividend futures occur in the 2011-2013 period, which corresponds to the peak of the European debt crisis.

Figure 1.4a plots the calibrated γ, which is the yield on the zero-coupon bond with infinite maturity. The plot shows a steady decline over time from approximately 6.5% to 1%. This reflects the drop in interest rates over the sample period as a consequence of quantitative easing by the European Central Bank. Figure 1.4b plots the calibrated β, which corresponds to the asymptotic risk-neutral expected growth rate of the dividends. It is always substantially lower than γ, which is required to keep the stock price finite. Figure 1.4c plots the calibrated ρ, which in view of (1.24) determines the correlation between the term structure of interest rates and dividends. Remarkably, ρ is positive over almost all of the sample period, with an average of around 80%. In view of (1.24), this indicates a highly negative correlation between interest rates and dividends. This negative correlation is a central ingredient in our model, since it increases the volatility of the stock price relative to the dividends and interest rates. This allows to reconcile the relatively high implied volatility of stock options with the relatively low implied volatility of dividend options and swaptions. For example, the large drop in ρ at the beginning of 2015 corresponds to the a period where the implied volatility of the stock option dropped sharply, but the dividend option was unaffected. Figure 1.5 plots the instantaneous correlation between S_t and r_t. The stock price is affected by interest rates through two channels: 1)

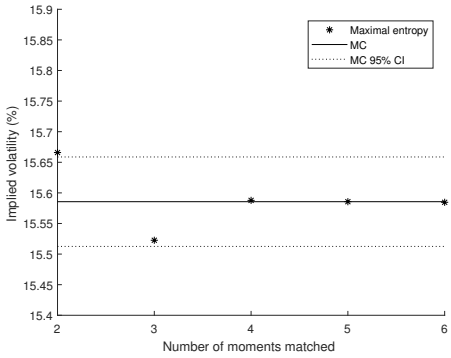
[13] In addition, we also use the forward contract price as a control variate. This variance reduction technique reduces the variance of the Monte-Carlo estimator approximately by a factor 4.



(a) Swaption



(b) Dividend option

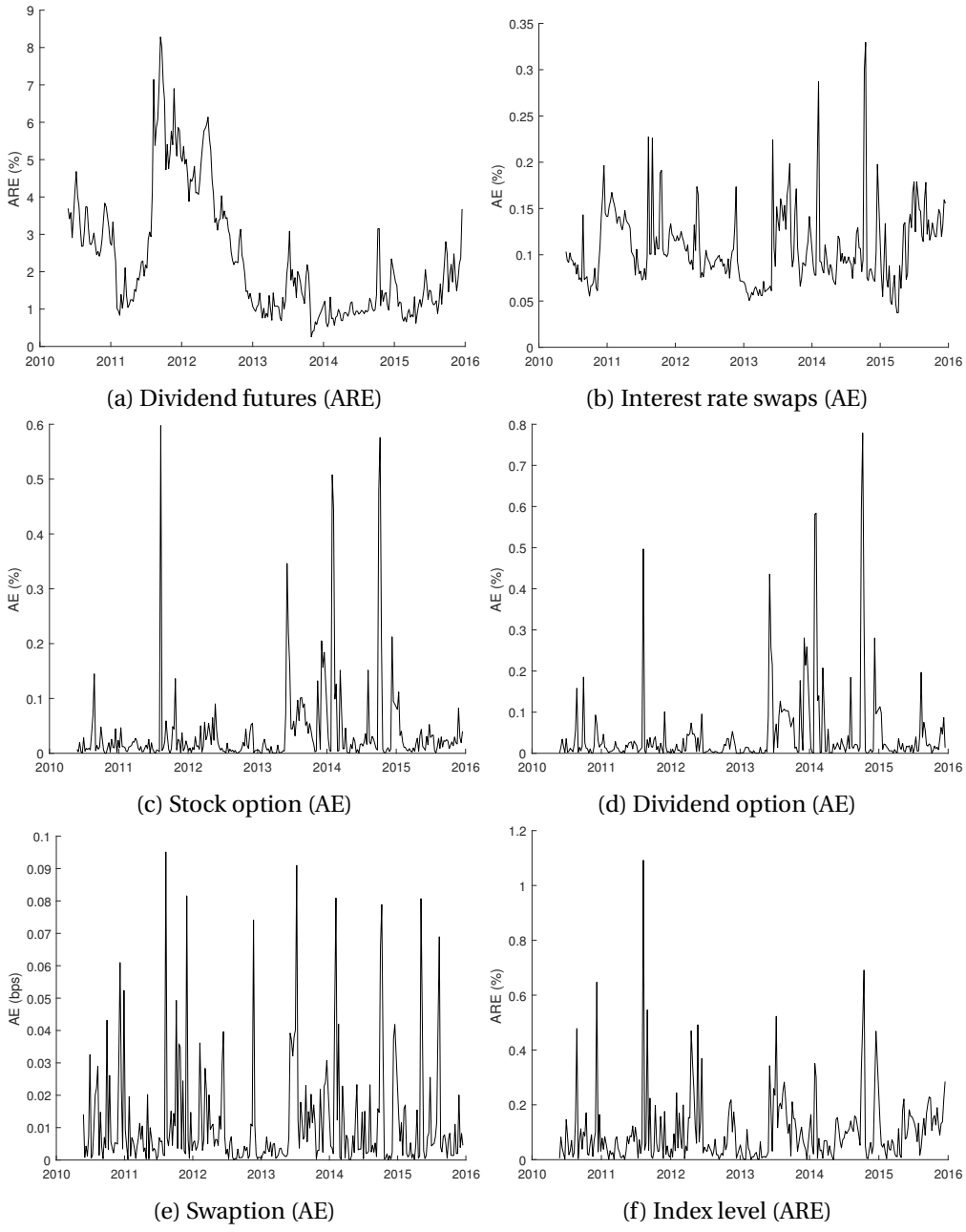


(c) Stock option

Figure 1.2 – Maximum entropy option prices for different number of moments matched. The swaption has maturity 3 months and underlying swap with tenor ten years, the dividend option has maturity 2 years, and the stock option has maturity 3 months. All options have ATM strike.

	Mean	Median	Std	Max
Dividend futures (ARE in %)				
1y	3.90	3.01	3.02	12.98
2y	1.12	0.93	1.00	5.42
3y	2.26	1.25	2.29	10.53
4y	2.59	1.54	2.41	10.80
5y	2.25	2.15	1.77	7.88
7y	1.18	0.76	1.10	5.34
9y	3.13	2.40	2.42	12.06
Interest rate swaps (AE in %)				
1y	0.12	0.11	0.08	0.41
2y	0.08	0.06	0.07	0.40
3y	0.09	0.09	0.05	0.38
4y	0.08	0.09	0.04	0.35
5y	0.06	0.05	0.05	0.28
7y	0.10	0.09	0.06	0.27
10y	0.16	0.15	0.09	0.38
20y	0.16	0.13	0.13	0.61
Dividend option (AE in %)	0.04	0.01	0.10	0.78
Swaption (AE in bps)	0.01	0.01	0.02	0.10
Stock option (AE in %)	0.04	0.01	0.08	0.60
Index level (ARE in %)	0.10	0.07	0.12	1.09

Table 1.1 – Statistics on Absolute Error (AE) and Absolute Relative Error (ARE) of instruments included in the calibration. The dividend option has a maturity of 2 years, the swaption has a maturity of 3 months and underlying swap of 10 years, and the stock option has a maturity of 3 months. All options have ATM strike.



(a) Dividend futures (ARE)

(b) Interest rate swaps (AE)

(c) Stock option (AE)

(d) Dividend option (AE)

(e) Swaption (AE)

(f) Index level (ARE)

Figure 1.3 – Absolute Error (AE) or Absolute Relative Error (ARE) of instruments included in the calibration. The errors of the dividend futures and interest rate swaps are averaged across all maturities. The dividend option has maturity 2 years, the swaption has maturity 3 months and underlying swap of 10 years, and the stock option has maturity 3 months. All options have ATM strike.

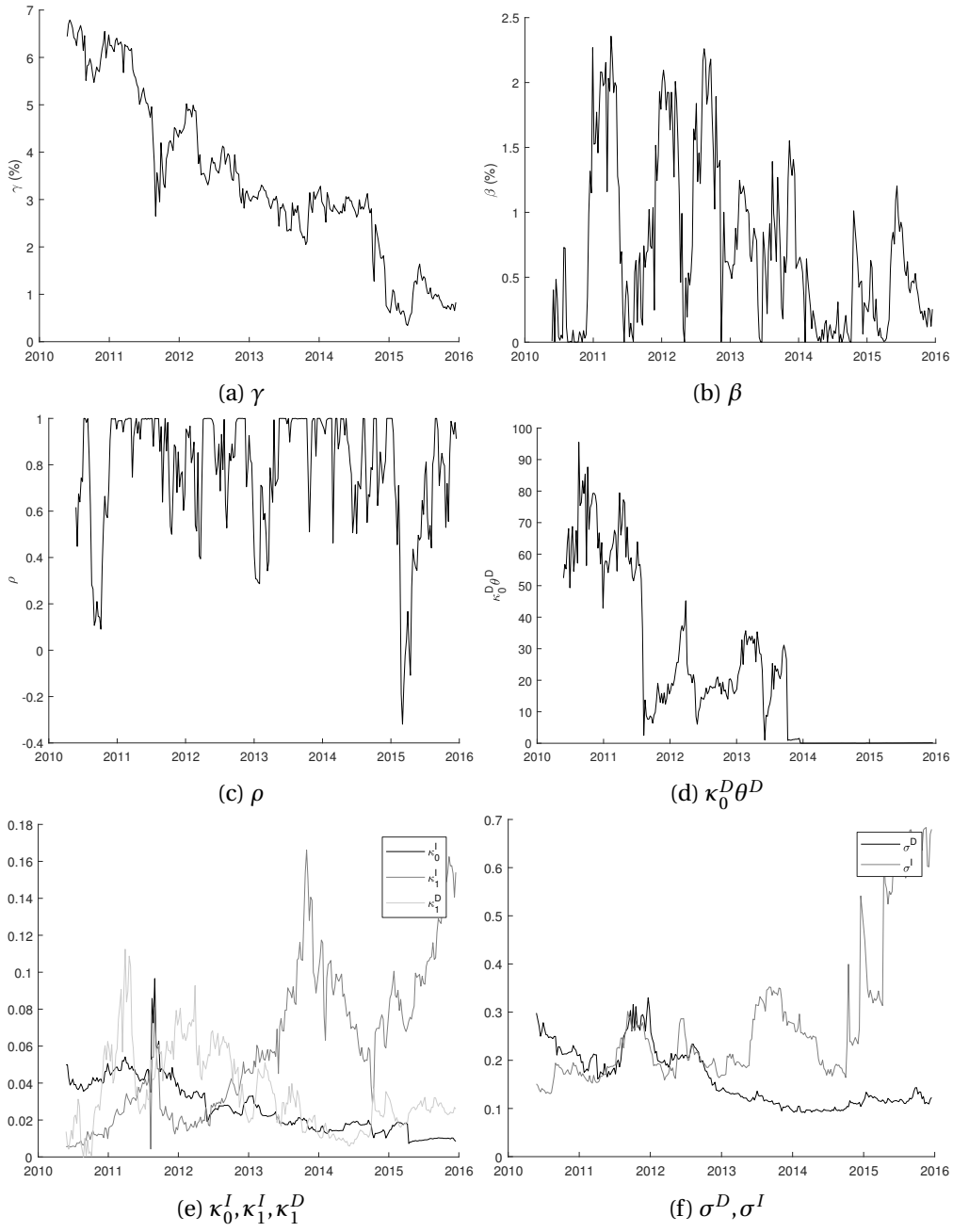
	Mean	Median	Std	Max
Dividend futures (ARE in %)				
6y	1.58	1.42	1.08	4.52
8y	1.98	1.45	1.66	9.60
Interest rate swaps (AE in %)				
6y	0.07	0.06	0.05	0.25
8y	0.12	0.12	0.07	0.32
Dividend option (AE in %)	2.40	2.42	0.85	5.28
Swaption (AE in bps)	3.85	3.71	2.17	8.84
Stock option (AE in %)	1.81	1.51	1.65	15.83

Table 1.2 – Statistics on Absolute Error (AE) or Absolute Relative Error (ARE) of instruments not included in the calibration. The dividend option has maturity 3 years, the swaption has maturity 6 months and underlying swap of 10 years, and the stock option has a maturity of 6 months. All options have ATM strike.

through the discounting of future dividends and 2) through the correlation between dividends and interest rates. Figure 1.5 shows a negative instantaneous correlation between S_t and r_t, except on a handful of days where the second channel marginally offsets the first one.

Figure 1.6a plots the scaled initial value X^D_{10}, which corresponds to the spot dividend rate D_0. Not surprisingly, it closely resembles the dynamics of the 1 year dividend futures price in Figure 1.1a. The term structure of dividend futures is downward sloping over almost the entire sample, which is reflected in the calibration by the fact that X^D_{10} is always well above its long-term mean θ^D. Figure 1.6b plots the initial values X^I_{00}, X^I_{10} of the interest rate factors, and Figure 1.6c plots the corresponding model implied short rate r_0. The increasing trend of the factor process over time illustrates the increasingly exceptional low interest rate situation in the Eurozone. Figure 1.6d plots the normal volatility $\sigma^I \kappa^I_0 \frac{X^I_{10}}{X^I_{00}}$ of the short rate, which looks similar in shape to the swaption implied volatility in Figure 1.1d.

As an out-of-sample exercise, we compute model implied prices of instruments not included in the calibration. Specifically, we consider dividend futures and interest rate swaps with maturity in 6 and 8 years, a dividend option with maturity in 3 years, a swaption with maturity in 6 months and underlying swap with tenor 10 years, and a stock option with maturity in 6 months. All options have ATM strike. The market and model implied prices are shown in Figure 1.7. The errors are reported in Table 1.2. The errors of the dividend futures and the interest rate swaps are comparable to their in-sample counterparts. There is however a clear deterioration in the fit with option prices out-of-sample, especially dividend options. This indicates that a richer volatility structure than the parsimonious one in (1.23) might be needed to fit the term structure of option prices.



(a) γ (b) β

(c) ρ (d) $\kappa_0^D\theta^D$

(e) $\kappa_0^I, \kappa_1^I, \kappa_1^D$ (f) σ^D, σ^I

Figure 1.4 – Calibrated model parameters.

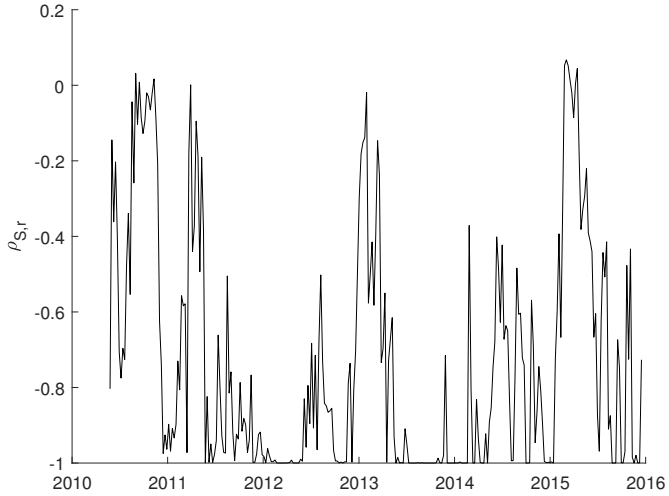
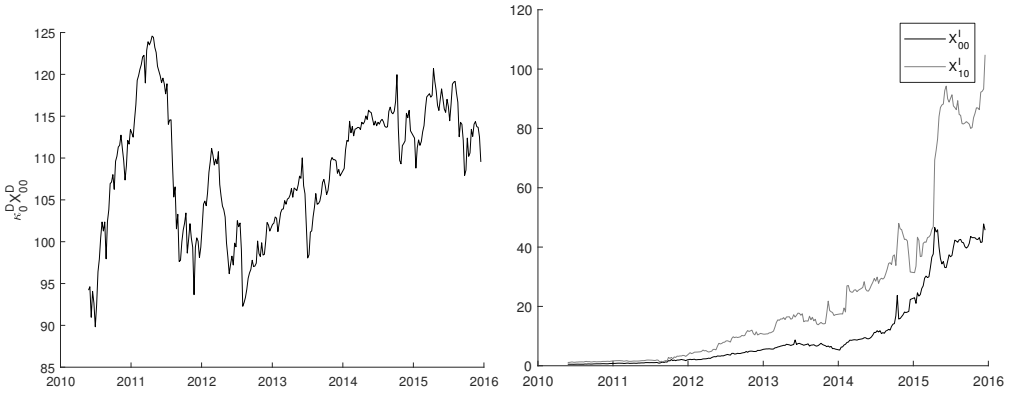
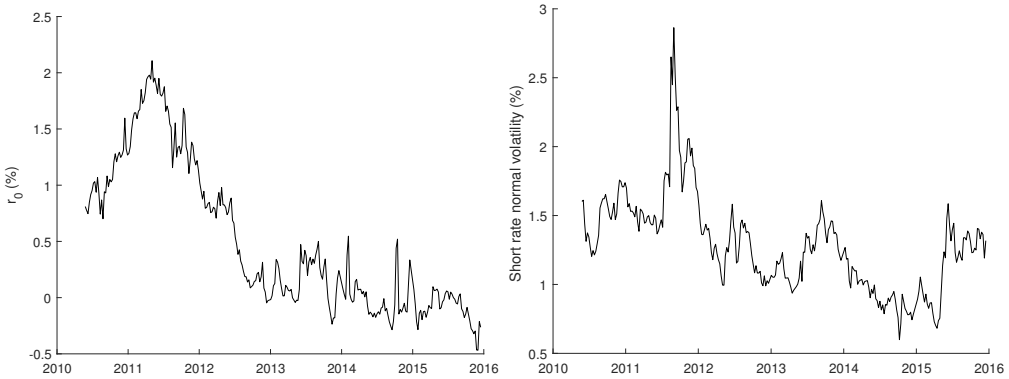


Figure 1.5 – Instantaneous correlation between S_t and r_t implied by calibrated parameters.



(a) Dividend rate $\kappa_0^D X_{10}^D$

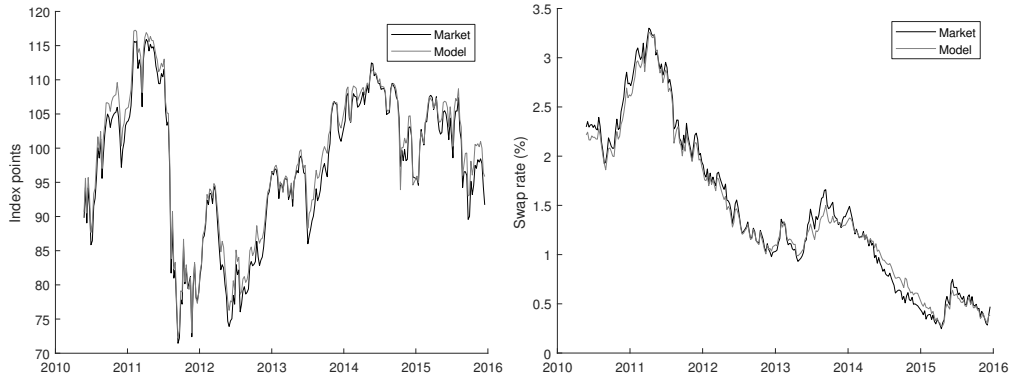
(b) X_{00}^I and X_{10}^I



(c) Short rate $r_0 = (\gamma + \kappa_0^I) - \kappa_0^I \frac{X_{10}^I}{X_{00}^I}$

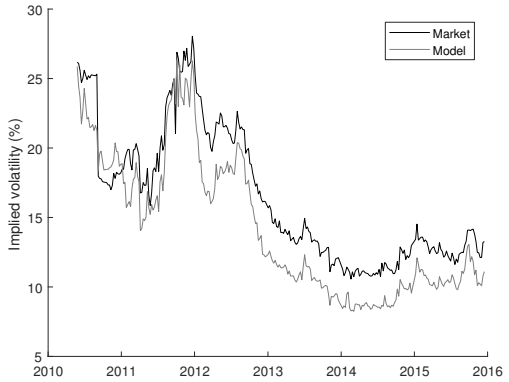
(d) Normal volatility short rate $\sigma^I \kappa_0^I \frac{X_{10}^I}{X_{00}^I}$.

Figure 1.6 – Calibrated interest rate factors X_{00}^I and X_{10}^I, scaled dividend factor $\kappa_0^D X_{10}^D$, the short rate $r_0 = (\gamma + \kappa_0^I) - \kappa_0^I \frac{X_{10}^I}{X_{00}^I}$, and the normal volatility of the short rate $\sigma^I \kappa_0^I \frac{X_{10}^I}{X_{00}^I}$ over time.



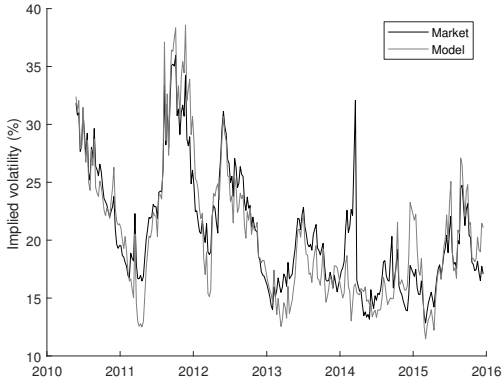
(a) Dividend futures 6y

(b) Interest rate swap 6y



(c) Dividend option 3y

(d) Swaption 6m×10y



(e) Stock option 6m

Figure 1.7 – Market and model implied prices of instruments not included in the calibration.

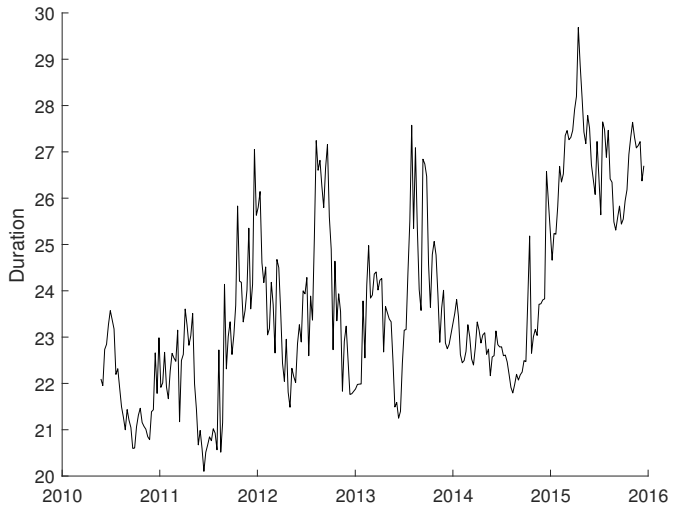


Figure 1.8 – Stock duration implied by calibrated parameters.

Figure 1.8 plots the stock duration (1.15) implied by the calibrated model parameters. The duration is quite stable over the calibration period, with an average around 24 years. Dechow et al. (2004) and Weber (2018) construct a stock duration measure based on balance sheet data and find an average duration of approximately 15 and 19 years, respectively, for a large cross-section of stocks. Table 1.3 contains computation times for calculating option prices. The bulk of the computation time is due to the computation of the moments of $g(X_T)$ in (1.17). The number of stochastic factors that drive a derivative's payoff and the degree of moments that have to be matched therefore strongly affect the computation time. We observe that all timings of the maximum entropy method are well below the time it took to run the benchmark Monte-Carlo simulation. The pricing of swaptions is much faster than the pricing of dividend and stock options, especially as the number of moments increases. This is because the discounted swaption payoff only depends on on the 2-dimensional process $(X^I_{0t}, X^I_{1t})^\top$, while the discounted payoff of the dividend and stock option depends on the entire 4-dimensional process $X_t = (X^I_{0t}, X^I_{1t}, X^D_{0t}, X^D_{1t})^\top$. In addition, the discounted payoff of the dividend and stock option is quadratic in the factors. Therefore, in order to compute moments up to degree N of the discounted payoff, we need to compute moments up to degree $2N$ of the factors. The computation of the dividend option is further complicated by the path-depedent nature of its payoff. Indeed, the dividend option payoff depends on the realization of the factors at T_1 *and* T_2. In order to compute the moments of $\zeta_{T_2}(C_{T_2} - C_{T_1})$, we have to apply the moment formula twice. Hence, it involves computing a matrix exponential twice, which causes an additional computation time compared to the stock option.

	$N=2$	$N=3$	$N=4$	$N=5$	$N=6$	MC
Swaption	0.02	0.02	0.02	0.02	0.02	1.0114
Dividend option	0.03	0.06	0.14	0.41	1.24	25.88
Stock option	0.02	0.03	0.06	0.07	0.11	3.49

Table 1.3 – Computation times (in seconds) needed to price swaptions, dividend options, and stock options using a) the maximum entropy method matching N moments and b) Monte-Carlo simulation with 10^5 sample paths and weekly discretization. The swaption has a maturity of 3 months and underlying swap of 10 years, the dividend option has a maturity of 2 years, and the stock option has a maturity of 3 months. All options have ATM strike.

1.6 Extensions

1.6.1 Seasonality

It is well known that some stock markets exhibit a strongly seasonal pattern in the payment of dividends. For example, Figure 1.9 shows that the constituents of the Euro Stoxx 50 pay a large fraction of their dividends between April and June each year.[14] The easiest way to incorporate seasonality in our framework is to introduce a deterministic function of time $\delta(t)$ and redefine the cumulative dividend process as:

$$C_t = \int_0^t \delta(s)\,\mathrm{d}s + \mathrm{e}^{\beta t} p^\top H_1(X_t). \tag{1.25}$$

The function $\delta(t)$ therefore adds a deterministic shift to the instantaneous dividend rate:

$$D_t = \delta(t) + \mathrm{e}^{\beta t} p^\top (\beta \mathrm{Id} + G_1) H_1(X_t). \tag{1.26}$$

In addition to incorporating seasonality, $\delta(t)$ can also be chosen such that the observed dividend futures prices are perfectly matched. In Appendix A.1 we show how the bootstrapping method described in Chapter 3 (cf., Filipović and Willems (2018)) can be used to find such a function. We do not lose any tractability with the specification in (1.25), since the moments of $C_{T_2} - C_{T_1}$ can still easily be computed.

Alternatively, we could also introduce time dependence in the specification of X_t. Doing so in general comes at the cost of losing tractability, because we leave the class of polynomial jump-diffusions. However, it is possible to introduce a specific type of time dependence such that we *do* stay in the class of polynomial jump-diffusions. Define $\Gamma(t)$ as a vector of sine and cosine functions whose frequencies are integer multiples of 2π (so that they all have period

[14] See e.g. Marchioro (2016) for a study of dividend seasonality in other markets.

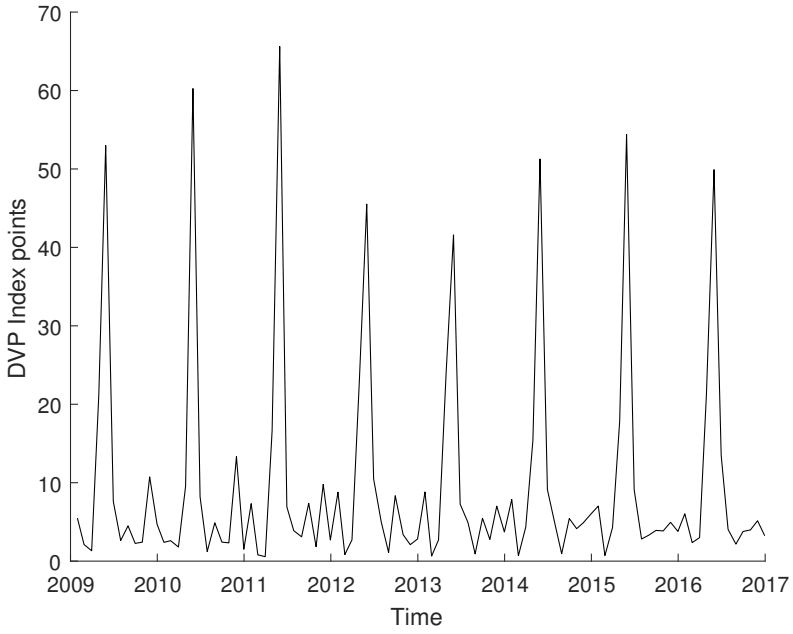


Figure 1.9 – Monthly dividend payments by Euro Stoxx 50 constituents (in index points) from January 2009 until December 2016. Source: Euro Stoxx 50 DVP index, Bloomberg.

one)

$$\Gamma(t) = \begin{pmatrix} \sin(2\pi t) \\ \cos(2\pi t) \\ \vdots \\ \sin(2\pi K t) \\ \cos(2\pi K t) \end{pmatrix} \in \mathbb{R}^{2K}, \quad K \in \mathbb{N}, \quad t \geq 0.$$

The superposition

$$z_0 + z^\top \Gamma(t), \quad (z_0, z) \in \mathbb{R}^{1+2K},$$

is a flexible function for modeling annually repeating cycles and is a standard choice for pricing commodity derivatives (see e.g. Sørensen (2002)). In fact, from Fourier analysis we know that any smooth periodic function can be expressed as a sum of sine and cosine waves. Remark now that $\Gamma(t)$ is the solution of the following linear ordinary differential equation

$$\mathrm{d}\Gamma(t) = \mathrm{blkdiag}\left(\begin{pmatrix} 0 & 2\pi \\ -2\pi & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 2\pi K \\ -2\pi K & 0 \end{pmatrix}\right)\Gamma(t)\mathrm{d}t.$$

The function $\Gamma(t)$ can therefore be seen as a (deterministic) process of the form in (1.1) and can be added to the factor process. For example, the specification for (X^D_{0t}, X^D_{1t}) in (1.23) could

be replaced by

$$\begin{cases} \mathrm{d}X_{0t}^D &= \kappa_0^D \left(X_{1t}^D - X_{0t}^D\right) \mathrm{d}t \\ \mathrm{d}X_{1t}^D &= \kappa_1^D (z_0 + z^\top \Gamma(t) - X_{1t}^D)\, \mathrm{d}t \quad + \quad \sigma^D X_{1t}^D \left(\rho\, \mathrm{d}B_{1t} + \sqrt{1-\rho^2}\, \mathrm{d}B_{2t}\right) \end{cases},$$

where the first factor mean-reverts around the second, and the second mean-reverts around a time-dependent mean. The process X_t does not belong to the class of polynomial jump-diffusions, however the augmented process $(\Gamma(t), X_t)$ does.

In the calibration exercise in Section 1.5, we did not include any seasonal behavior in the dividends because the instruments used in the estimation are not directly affected by seasonality. Indeed, all the dividend derivatives used in the calibration reference the total amount of dividends paid in a full calendar year. The timing of the dividend payments within the year does therefore not play any role. In theory, the stock price should inherit the seasonal pattern from the dividend payments, since it drops by exactly the amount of dividends paid out. In practice, however, these price drops are obscured by the volatility of the stock price since the dividend payments typically represent only a small fraction of the total stock price. Dividend seasonality only plays a role for pricing claims on dividends realized over a time period different from an integer number of calendar years.

1.6.2 Dividend Forwards

Dividend forwards, also known as dividend swaps, are the OTC equivalent of the exchange traded dividend futures. The buyer of a dividend forward receives at a future date T_2 the dividends realized over a certain time period $[T_1, T_2]$ against a fixed payment. Dividend forwards differ from dividend futures because they are not marked to market on a daily basis. The dividend forward price $D_{swap}(t, T_1, T_2)$, $t \le T_1 \le T_2$, is defined as the fixed payment that makes the forward have zero initial value

$$\begin{aligned} D_{fwd}(t, T_1, T_2) &= \frac{1}{P(t, T_2)} \frac{1}{\zeta_t} \mathbb{E}_t \left[\zeta_{T_2} (C_{T_2} - C_{T_1})\right] \\ &= D_{fut}(t, T_1, T_2) + \frac{\mathrm{Cov}_t \left[\zeta_{T_2}, C_{T_2} - C_{T_1}\right]}{P(t, T_2)\zeta_t}. \end{aligned}$$

If interest rates and dividends are independent, then we have $D_{fwd}(t, T_1, T_2) = D_{fut}(t, T_1, T_2)$. However, if there is a positive (negative) dependence between interest rates and dividends, then there is a convexity adjustment and the dividend forward price will be smaller (larger) than the dividend futures price. The following proposition derives the dividend forward price in the polynomial framework.

Proposition 1.6.1. *The dividend forward price is given by*

$$D_{fwd}(t, T_1, T_2) = \frac{\left(\mathrm{e}^{\beta T_2} w_2^\top \mathrm{e}^{G_2(T_2 - t)} - \mathrm{e}^{\beta T_1} w_1^\top \mathrm{e}^{G_2(T_1 - t)}\right) H_2(X_t)}{q^\top \mathrm{e}^{G_1(T_2 - t)} H_1(X_t)},$$

where $w_1, w_2 \in \mathbb{R}^{N_2}$ *are the unique coordinate vectors satisfying*

$$w_1^\top H_2(x) = p^\top H_1(x) q^\top \mathrm{e}^{G_1(T_2-T_1)} H_1(x), \quad w_2^\top H_2(x) = p^\top H_1(x) q^\top H_1(x).$$

1.7 Conclusion

We have introduced an integrated framework designed to jointly price the term structures of dividends and interest rates. The uncertainty in the economy is modeled with a multivariate polynomial jump-diffusion. The model is tractable because we can calculate all conditional moments of the factor process in closed form. In particular, we have derived closed form formulas for prices of bonds, dividend futures, and the dividend paying stock. Option prices are obtained by integrating the discounted payoff function with respect to a moment matched density function that maximizes the Boltzmann-Shannon entropy. We have introduced the LJD model, characterized by a martingale part that loads linearly on the factors. The LJD model allows for a flexible dependence structure between the factors. We have assumed that dividends are paid out continuously and ignored the possibility of default. These assumptions are justified when considering derivatives on a stock index, but become questionable for derivatives on a single stock. An interesting future research direction is therefore to extend our framework with discrete dividend payments and default risk.

2 Asian Option Pricing with Orthogonal Polynomials

In this chapter we derive a series expansion for the price of a continuously sampled arithmetic Asian option in the Black-Scholes setting. The expansion is based on polynomials that are orthogonal with respect to the log-normal distribution. All terms in the series are fully explicit and no numerical integration nor any special functions are involved. We provide sufficient conditions to guarantee convergence of the series. The moment indeterminacy of the log-normal distribution introduces an asymptotic bias in the series, however we show numerically that the bias can safely be ignored in practice.

2.1 Introduction

An Asian option is a derivative contract with payoff contingent on the average value of the underlying asset over a certain time period. Valuation of these contracts is not straightforward because of the path-dependent nature of the payoff. Even in the standard Black and Scholes (1973) setting the distribution of the (arithmetic) average stock price is not known. In this chapter we derive a series expansion for the Asian option price in the Black-Scholes setting using polynomials that are orthogonal with respect to the log-normal distribution. The terms in the series are fully explicit since all the moments of the average price are known. We prove that the series does not diverge by showing that the tails of the average price distribution are dominated by the tails of a log-normal distribution. As a consequence of the well known moment indeterminacy of the log-normal distribution (see e.g., Heyde (1963)), it is not theoretically guaranteed that the series converges to the true price. We show numerically that this asymptotic bias is small for standard parameterizations and the real approximation challenge lies in controlling the error coming from truncating the series after a finite number of terms.

There exists a vast literature on the problem of Asian option pricing. We give a brief overview which is by no means exhaustive. One approach is to approximate the unknown distribution of the average price with a more tractable one. Turnbull and Wakeman (1991), Levy (1992), Ritchken et al. (1993), Li and Chen (2016) use an Edgeworth expansion around a log-normal reference distribution to approximate the distribution of the arithmetic average of the ge-

ometric Brownian motion. Ju (2002) and Sun et al. (2013) use a Taylor series approach to approximate the unknown average distribution from a log-normal. Milevsky and Posner (1998) use a moment matched inverse gamma distribution as approximation. Their choice is motivated by the fact that the infinite horizon average stock price has an inverse gamma distribution. More recently, Aprahamian and Maddah (2015) use a moment matched compound gamma distribution. Although these type of approximations lead to analytic option price formulas, their main drawback is the lack of reliable error estimates. A second strand of the literature focuses on Monte-Carlo and PDE methods. Kemna and Vorst (1990) propose to use the continuously sampled geometric option price as a control variate and show that it leads to a significant variance reduction. Fu et al. (1999) argue that this is a biased control variate, but interestingly the bias approximately offsets the bias coming from discretely computing the continuous average in the simulation. Lapeyre et al. (2001) perform a numerical comparison of different Monte-Carlo schemes. Rogers and Shi (1995), Zvan et al. (1996), Vecer (2001, 2002), Marcozzi (2003) solve the pricing PDE numerically. Another approach is to derive bounds on the Asian option price, see e.g. Curran (1994), Rogers and Shi (1995), Thompson (2002), and Vanmaele et al. (2006). Finally, there are several papers that derive exact representations of the Asian option price. Yor (1992) expresses the option price as a triple integral, to be evaluated numerically. Geman and Yor (1993) derive the Laplace transform of the option price. Numerical inversion of this Laplace transform is however a delicate task, see e.g. Eydeland and Geman (1995), Fu et al. (1999), Shaw (2002). Carverhill and Clewlow (1990) relate the density of the discrete arithmetic average to an iterative convolution of densities, which is approximated numerically through the Fast Fourier Transform algorithm. Later extensions and improvements of the convolution approach include Benhamou (2002), Fusai and Meucci (2008), Černỳ and Kyriakou (2011), and Fusai et al. (2011). Dufresne (2000) derives a series representation using Laguerre orthogonal polynomials. Linetsky (2004) derives a series representation using spectral expansions involving Whittaker functions.

The approach taken in this chapter is closely related to Dufresne (2000) in the sense that both are based on orthogonal polynomial expansions. The Laguerre series expansion can be shown to diverge when directly expanding the density of the average price, which is related to the fact that the tails of the average price distribution are heavier than those of the Gamma distribution. As a workaround, Dufresne (2000) proposes to work with the reciprocal of the average, for which the Laguerre series does converge. The main downside of this approach is that the moments of the reciprocal average are not available in closed form and need to be calculated through numerical integration, which introduces a high computational cost and additional numerical errors. Asmussen et al. (2016) use a different workaround and expand an exponentially tilted transformation of the density of a sum of log-normal random variables using a Laguerre series. They show that the exponential tilting transformation guarantees the expansion to converge. However, a similar problem as in Dufresne (2000) arises: the moments of the exponentially tilted density are not available in closed form and have to be computed numerically. In contrast, our approach is fully explicit and does not involve any numerical integration, which makes it very fast.

Truncating our series after only one term is equivalent to pricing the option under the assumption that the average price is log-normally distributed. The remaining terms in the series can therefore be thought of as corrections to the log-normal distribution. This has a very similar flavour to approaches using an Edgeworth expansion around the log-normal distribution (cfr. Jarrow and Rudd (1982) and Turnbull and Wakeman (1991)). The key difference with our approach is that the Edgeworth expansion can easily diverge because it lacks a proper theoretical framework. In contrast, the series we present in this chapter is guaranteed to converge, possibly with a small asymptotic bias. A thorough study of the approximation error reveals that the asymptotic bias is positively related to the volatility of the stock price process and the option expiry. We use the integration by parts formula from Malliavin calculus to derive an upper bound on the approximation error.

The remaining of this chapter is structured as follows. Section 2.2 casts the problem and derives useful properties about the distribution of the arithmetic average. Section 2.3 describes the density expansion used to approximate the option price. In Section 2.4 we investigate the approximation error. Section 2.5 illustrates the method with numerical examples. Section 2.6 concludes. All proofs can be found in Appendix B.3.

2.2 The Distribution of the Arithmetic Average

We fix a stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{Q})$ satisfying the usual conditions and let $\mathbb{Q}$ be the risk-neutral probability measure. We consider the Black-Scholes setup where the underlying stock price S_t follows a geometric Brownian motion:

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}B_t,$$

where $r \in \mathbb{R}$ is the short-rate, $\sigma > 0$ the volatility of the asset, and B_t a standard Brownian motion. For notational convenience we assume $S_0 = 1$, which is without loss of generality. We define the average price process as

$$A_t = \frac{1}{t}\int_0^t S_u\,\mathrm{d}u, \quad t > 0.$$

The price at time 0 of an Asian option with continuous arithmetic averaging, strike $K > 0$, and expiry $T > 0$ is given by

$$\pi = \mathrm{e}^{-rT}\mathbb{E}\left[(A_T - K)^+\right].$$

The option price can not be computed explicitly since we do not know the distribution of A_T. However, we can derive useful results about its distribution.

We start by computing all the moments of A_T. Using the time-reversal property of a Brownian motion, we have the following identity in law (cfr. Dufresne (1990), Carmona et al. (1997), Donati-Martin et al. (2001), Linetsky (2004)):

Lemma 2.2.1. *The random variable TA_T has the same distribution as the solution at time T of the following SDE*

$$\mathrm{d}X_t = (rX_t + 1)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t, \quad X_0 = 0. \tag{2.1}$$

The SDE in (2.1) defines a polynomial diffusion (see e.g., Filipović and Larsson (2016)) and is a special case of the factor process (1.1) considered in the first chapter. Using the moment formula in (1.3) we can therefore compute all the moments of A_T in closed form, as shown in the following proposition.

Proposition 2.2.2. *If we denote by $H_n(x) = (1, x, \ldots, x^n)^\top$, $n \in \mathbb{N}$, then we have*

$$\mathbb{E}\left[H_n(A_T)\right] = \mathrm{e}^{G_n T} H_n(0),$$

where $G_n \in \mathbb{R}^{(n+1)\times(n+1)}$ is the following lower bidiagonal matrix

$$G_n = \begin{pmatrix} 0 & & & \\ \frac{1}{T} & r & & \\ & \ddots & \ddots & \\ & & \frac{n}{T} & (nr + \frac{1}{2}n(n-1)\sigma^2) \end{pmatrix}. \tag{2.2}$$

Given that the matrix exponential is a standard built-in function in most scientific computing packages, the above moment formula is very easy to implement. There also exist efficient numerical methods to directly compute the action of the matrix exponential, see e.g. Al-Mohy and Higham (2011) and Caliari et al. (2014). An equivalent, but more cumbersome to implement, representation of the moments can be found in Geman and Yor (1993).

The following proposition shows that A_T admits a smooth density function $g(x)$ whose tails are dominated by the tails of a log-normal density function:

Proposition 2.2.3.

1. *The random variable A_T admits an infinitely differentiable density function $g(x)$.*

2. *The density function $g(x)$ has the following asymptotic properties:*

$$g(x) = \begin{cases} \mathcal{O}\left(\exp\left\{-\frac{3}{2}\frac{\log(x)^2}{\sigma^2 T}\right\}\right) & \text{for} \quad x \to 0, \\ \mathcal{O}\left(\exp\left\{-\frac{1}{2}\frac{\log(x)^2}{\sigma^2 T}\right\}\right) & \text{for} \quad x \to \infty. \end{cases}$$

2.3 Polynomial Expansion

Armed with the moments of A_T, we now tackle the problem of pricing options on A_T. In the first chapter, we used a maximum entropy approach to find a density function satisfying the moment constraints and we used numerical integration to compute option prices. In this chapter, we employ the polynomial density expansion approach described by Filipović et al. (2013), see also Ackerer et al. (2018), Ackerer and Filipović (2019b), and Carr and Willems (2019) for similar applications. The advantage over a maximum entropy approach lies in the fact that we do not have to do any numerical integration in this case.

Define the weighted Hilbert space L^2_w as the set of measurable functions f on $\mathbb{R}$ with finite L^2_w-norm defined as

$$\|f\|^2_w = \int_0^\infty f(x)^2 w(x)\,\mathrm{d}x, \quad w(x) = \frac{1}{\sqrt{2\pi}\nu x}\exp\left\{-\frac{(\log(x)-\mu)^2}{2\nu^2}\right\}, \tag{2.3}$$

for some constants $\mu \in \mathbb{R}$, $\nu > 0$. The weight function w is the density function of a log-normal distribution. The corresponding scalar product between two functions $f, h \in L^2_w$ is defined as

$$\langle f, h\rangle_w = \int_0^\infty f(x)h(x)w(x)\,\mathrm{d}x.$$

Since the measures associated with the densities g and w are equivalent, we can define the likelihood ratio function ℓ such that

$$g(x) = \ell(x)w(x), \quad x \in (0,\infty).$$

Using Proposition 2.2.3 we now have the following result:

Proposition 2.3.1. *If $\nu^2 > \frac{1}{2}\sigma^2 T$, then $\ell \in L^2_w$, i.e.*

$$\int_0^\infty \left(\frac{g(x)}{w(x)}\right)^2 w(x)\,\mathrm{d}x < \infty.$$

Denote by $\mathrm{Pol}(\mathbb{R})$ the set of polynomials on $\mathbb{R}$ and by $\mathrm{Pol}_N(\mathbb{R})$ the subset of polynomials on $\mathbb{R}$ of degree at most $N \in \mathbb{N}$. Since the log-normal distribution has finite moments of any degree, we have $\mathrm{Pol}_N(\mathbb{R}) \subset L^2_w$ for all $N \in \mathbb{N}$. Let $b_0, b_1, \ldots, b_N$ form an orthonormal polynomial basis for $\mathrm{Pol}_N(\mathbb{R})$. Such a basis can, for example, be constructed numerically from the monomial basis using a Cholesky decomposition. Indeed, define the Hankel moment matrix $M = (M_{ij})_{0\le i,j\le N}$ as

$$M_{ij} = \langle x^i, x^j\rangle_w = \mathrm{e}^{\mu(i+j)+\frac{1}{2}(i+j)^2\nu^2}, \quad i,j = 0,\ldots,N, \tag{2.4}$$

which is positive definite by construction. If we denote by $M = LL^\top$ the unique Cholesky

decomposition of M, then

$$(b_0(x),\dots,b_N(x))^\top = L^{-1}H_N(x),$$

forms an orthonormal polynomial basis for $\mathrm{Pol}_N(\mathbb{R})$. Alternative approaches to build an orthonormal basis are, for example, the three-term recurrence relation (see Lemma B.1.1 for details) or the analytical expressions for the orthonormal polynomials derived in Theorem 1.1 of Asmussen et al. (2016).

Remark 2.3.2. *The matrix M defined in* (2.4) *can in practice be non-positive definite due to rounding errors. This problem becomes increasingly important for large N and/or large ν because the elements in M grow very fast. Similarly, the moments of A_T can also grow very large, which causes rounding errors in finite precision arithmetic. In Appendix B.1 we describe a convenient scaling technique that solves these problems in many cases.*

Define the discounted payoff function $F(x) = \mathrm{e}^{-rT}(x-K)^+$. Since $F(x) \le \mathrm{e}^{-rT}x$ for all $x \ge 0$, we immediately have that $F \in L^2_w$. Denote by $\overline{\mathrm{Pol}(\mathbb{R})}$ the closure of $\mathrm{Pol}(\mathbb{R})$ in L^2_w. We define the projected option price $\bar{\pi} = \mathbb{E}[\bar{F}(A_T)]$, where $\bar{F}$ denotes the orthogonal projection of F onto $\overline{\mathrm{Pol}(\mathbb{R})}$ in L^2_w. Elementary functional analysis gives

$$\bar{\pi} = \langle \bar{F}, \ell \rangle_w = \sum_{n\ge 0} f_n \ell_n, \tag{2.5}$$

where we define the likelihood coefficients $\ell_n = \langle \ell, b_n \rangle_w$ and payoff coefficients $f_n = \langle F, b_n \rangle_w$. Truncating the series in (2.5) after a finite number of terms finally gives the following approximation for the Asian option price:

$$\pi^{(N)} = \sum_{n=0}^{N} f_n \ell_n, \quad N \in \mathbb{N}. \tag{2.6}$$

The likelihood coefficients are available in closed form using the moments of A_T in Proposition 2.2.2:

$$(\ell_0,\dots,\ell_N)^\top = L^{-1}\mathrm{e}^{G_N T}H_N(0).$$

The payoff coefficients can also be derived explicitly, as shown in the following proposition.

Proposition 2.3.3. *Let Φ be the standard normal cumulative distribution function. The payoff coefficients $f_0,\dots,f_N$ are given by*

$$(f_0,\dots,f_N)^\top = e^{-rT}L^{-1}(\tilde{f}_0,\dots,\tilde{f}_N)^\top,$$

with

$$\begin{aligned} \tilde{f}_n &= \mathrm{e}^{\mu(n+1)+\frac{1}{2}(n+1)^2\nu^2}\Phi(d_{n+1}) - K\mathrm{e}^{\mu n+\frac{1}{2}n^2\nu^2}\Phi(d_n), \quad n = 0,\dots,N, \\ d_n &= \frac{\mu+\nu^2 n - \log(K)}{\nu}, \quad n = 0,\dots,N+1. \end{aligned} \tag{2.7}$$

Equivalently, we could also derive the approximation (2.6) by projecting ℓ, instead of F, on the set of polynomials. This leads to the interpretation of (2.6) as the option price one obtains when approximating the true density $g(x)$ by

$$g^{(N)}(x) = w(x) \sum_{n=0}^{N} \ell_n b_n(x). \tag{2.8}$$

The function $g^{(N)}(x)$ integrates to one by construction

$$\int_0^\infty g^{(N)}(x)\,\mathrm{d}x = \sum_{n=0}^{N} \ell_n \langle b_n, b_0 = 1\rangle_w = \ell_0 = 1,$$

where the last equality follows from the fact that $g(x)$ integrates to one. However, it is not a true probability density function since it is not guaranteed to be non-negative.

2.4 Approximation Error

The error introduced by the the approximation in (2.6) can be decomposed as

$$\pi - \pi^{(N)} = (\pi - \bar{\pi}) + \left(\bar{\pi} - \pi^{(N)}\right).$$

The second term is guaranteed to converge to zero as $N \to \infty$. In order for the first term to vanish, we need $F \in \overline{\mathrm{Pol}(\mathbb{R})}$ and/or $\ell \in \overline{\mathrm{Pol}(\mathbb{R})}$. It is well known (see e.g., Heyde (1963)) that the log-normal distribution is not determined by its moments. As a consequence, the set of polynomials does not lie dense in L^2_w: $\overline{\mathrm{Pol}(\mathbb{R})} \subsetneq L^2_w$. Hence, the fact that $F, \ell \in L^2_w$ is not sufficient to guarantee that the first term in the error decomposition is zero. One of the goals of this chapter is to quantify the importance of the first error term. In this section we therefore investigate the error associated with projecting F and ℓ on the set of polynomials.

The L^2_w-distances of F and ℓ to their respective orthogonal projections on $\mathrm{Pol}_N(\mathbb{R})$ are given by

$$\epsilon^F_N := \left\| F - \sum_{n=0}^{N} b_n f_n \right\|^2_w = \|F\|^2_w - \sum_{n=0}^{N} f_n^2,$$
$$\epsilon^\ell_N := \left\| \ell - \sum_{n=0}^{N} b_n \ell_n \right\|^2_w = \|\ell\|^2_w - \sum_{n=0}^{N} \ell_n^2.$$

The L^2_w-norm of the payoff function F can be derived explicitly following very similar steps as in the proof of Proposition 2.3.3:

$$\|F\|^2_w = \mathrm{e}^{-2rT}\left(\mathrm{e}^{2\mu+2v^2}\Phi(d_2) - 2K\mathrm{e}^{\mu+0.5v^2}\Phi(d_1) + K^2\Phi(d_0)\right),$$

where d_0, d_1, and d_2 are defined in (2.7). Hence, we can explicitly evaluate ϵ^F_N.

The computation of ϵ_N^ℓ is more difficult since $\|\ell\|_w^2$ depends on the unknown density $g(x)$. The following lemma uses the integration by parts formula from Malliavin calculus to derive a representation of $g(x)$ in terms of an expectation, which can be evaluated by Monte-Carlo simulation:

Lemma 2.4.1. *For any $x \in \mathbb{R}$ we have*

$$g(x) = \mathbb{E}\left[\left(1_{\{A_T \geq x\}} - c(x)\right) \frac{2}{\sigma^2}\left(\frac{S_T - S_0}{TA_T^2} + \frac{\sigma^2 - r}{A_T}\right)\right], \tag{2.9}$$

where c is any deterministic finite-valued function.

Remark 2.4.2. *The purpose of the function c is to guarantee that the simulated $g(x)$ actually goes to zero for $x \to 0$. Indeed, if we set $c(x) \equiv 0$, then $g(0)$ can be different from zero due to the Monte-Carlo error, which can lead to numerical problems when evaluating $\ell(x)$. This can be avoided by, for example, using the indicator function $c(x) = 1_{x \leq \xi}$, for some $\xi > 0$.*

As a direct consequence of (2.9) we get the following expression for the L_w^2-norm of the likelihood ratio.

Corollary 2.4.3. *The L_w^2-norm of ℓ is given by*

$$\|\ell\|_w^2 = \mathbb{E}\left[\frac{(1_{\{A_T \geq \tilde{A}_T\}} - c(\tilde{A}_T))\frac{2}{\sigma^2}\left(\frac{S_T - S_0}{TA_T^2} + \frac{\sigma^2 - r}{A_T}\right)}{w(\tilde{A}_T)}\right],$$

where the random variable $\tilde{A}_T$ is independent from all other random variables and has the same distribution as A_T.

This allows us to get an estimate for ϵ_N^ℓ by simulating the random vector $(S_T, A_T, \tilde{A}_T)$. In Appendix B.2 we describe how to use the known density function of the geometric average as a powerful control variate to significantly reduce the variance in the Monte-Carlo simulation.

Using the Cauchy-Schwarz inequality we also have the following upper bound on the approximation error in terms of the projection errors ϵ_N^F and ϵ_K^ℓ.

Proposition 2.4.4. *For any $N \geq 0$ we have*

$$|\pi - \pi^{(N)}| \leq \sqrt{\epsilon_N^F \epsilon_N^\ell} \tag{2.10}$$

This upper bound will therefore be small if F and/or ℓ are well approximated by a polynomial in L_w^2. Computing the upper bound involves a Monte-Carlo simulation to compute ϵ_N^ℓ, which makes it impractical to use as a decision rule for N. This bound should be seen as a more conservative estimate of the approximation error compared to direct simulation of the option price.

2.5 Numerical Examples

In this section we compute Asian option prices using the series expansion in (2.6). The orthonormal basis is constructed using the scaling technique described in Appendix B.1. We set $v^2 = \frac{1}{2}\sigma^2 T + 10^{-4}$ so that Proposition 2.3.1 is satisfied and choose μ so that the first moment of w matches the first moment of A_T. As a consequence, we always have

$$\ell_1 = \int_0^\infty b_1(x) g(x)\,\mathrm{d}x = \langle b_0, b_1 \rangle_w = 0.$$

Remark 2.5.1. *Choosing μ and v so that the first two moments of A_T are matched is typically not possible due to the restriction $v^2 > \frac{1}{2}\sigma^2 T$ in Proposition 2.3.1. A similar problem arises in the Jacobi stochastic volatility model of Ackerer et al. (2018), where options are priced using a polynomial expansion with a normal density as weight function. Ackerer and Filipović (2019b) address this problem by using a mixture of two normal densities as weight function. Specifying w as a mixture of normal densities would not work in our setting since in this case $\ell \notin L^2_w$.*[1] *Instead, we can use a mixture of two log-normal densities:*

$$w(x) = c\,w_1(x) + (1-c)\,w_2(x),$$

where $c \in [0,1]$ is the mixture weight, and w_1 and w_2 are log-normal density functions with mean parameters $\mu_1, \mu_2 \in \mathbb{R}$, and volatility parameters $v_1, v_2 > 0$, respectively. In order for Proposition 2.3.1 to apply, it suffices to choose $v_1^2 > \frac{1}{2}\sigma^2 T$. The remaining parameters can then be chosen freely and used for higher order moment matching.

We consider a set of seven parameterizations that has been used as a set of test cases in, among others, Eydeland and Geman (1995), Fu et al. (1999), Dufresne (2000), and Linetsky (2004). The first columns of Table 2.1 contain the parameter values of the seven cases. The cases are ordered in increasing size of $\tau = \sigma^2 T$. Remark that $S_0 \neq 1$ for all cases, however we normalize the initial stock price to one and scale the strike and option price accordingly. The columns LNS10, LNS15, LNS20 (Log-Normal Series) contain the option price approximations using (2.6) for $N = 10, 15, 20$, respectively. The columns LS (Laguerre Series) and EE (Eigenvalue Expansion) correspond to the series expansions of Dufresne (2000) and Linetsky (2004), respectively. The column VEC shows the prices produced by the PDE method of Vecer (2001, 2002) using a grid with 400 space points and 200 time points.[2] The last column contains the 95% confidence intervals of a Monte-Carlo simulation using the geometric Asian option price as a control variate, cfr. Kemna and Vorst (1990). We simulate 2×10^5 price trajectories with a time step of 10^{-3} and approximate the continuous average with a discrete average.

[1] Instead of approximating the distribution of A_T, it is tempting to approximate the distribution of $\log(A_T)$ and rewrite the discounted payoff function accordingly. In this case, one can show that specifying w as a normal density function gives a series approximation that converges to the true price. The catch is that we do not know the moments of $\log(A_T)$, only those of A_T, and hence the terms in the series can not be computed explicitly.

[2] This grid choice corresponds to the one used in Vecer (2001). By significantly increasing the number of space points in the grid, the PDE method can achieve the same accuracy as Linetsky (2004). However, doing so makes the method very slow.

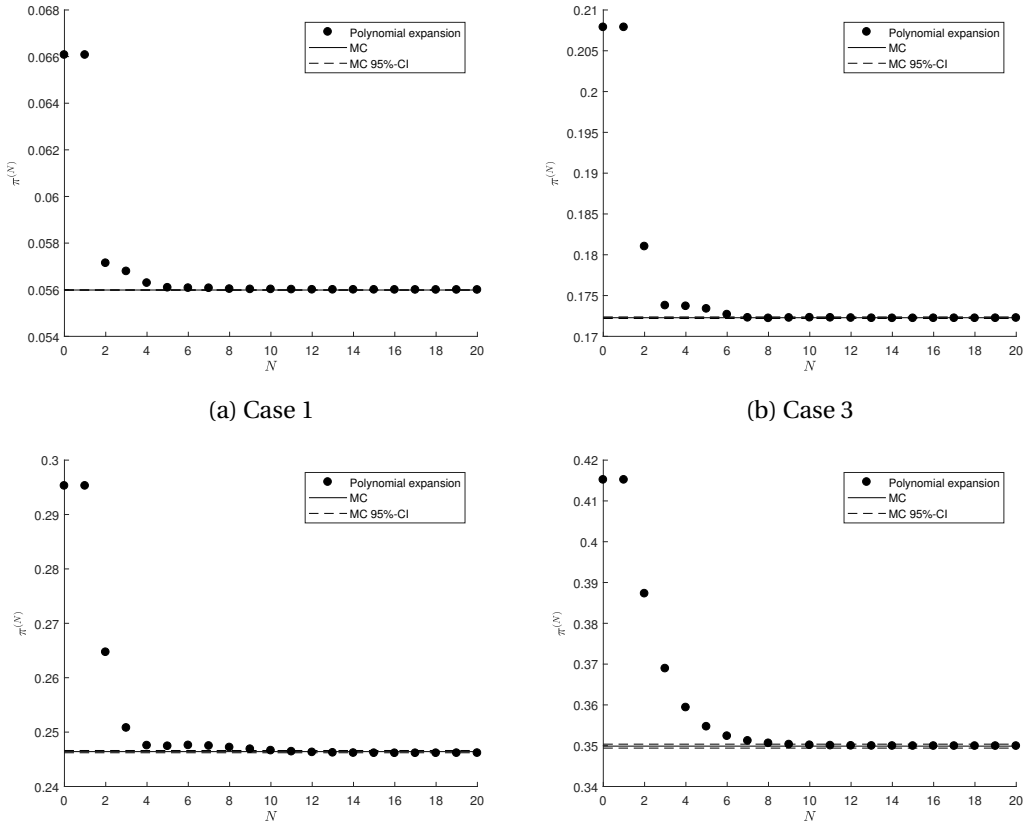
Case	r	σ	T	S_0	LNS10	LNS15	LNS20	LS	EE	VEC	MC 95% CI
1	.02	.10	1	2.0	.05601	.05600	.05599	.0197	.05599	.05595	[.05598, .05599]
2	.18	.30	1	2.0	.2185	.2184	.2184	.2184	.2184	.2184	[.2183 , .2185]
3	.0125	.25	2	2.0	.1723	.1722	.1722	.1723	.1723	.1723	[.1722 , .1724]
4	.05	.50	1	1.9	.1930	.1927	.1928	.1932	.1932	.1932	[.1929 , .1933]
5	.05	.50	1	2.0	.2466	.2461	.2461	.2464	.2464	.2464	[.2461 , .2466]
6	.05	.50	1	2.1	.3068	.3062	.3061	.3062	.3062	.3062	[.3060 , .3065]
7	.05	.50	2	2.0	.3501	.3499	.3499	.3501	.3501	.3500	[.3494 , .3504]

Table 2.1 – Price approximations for different parameterizations and different methods. The strike price is $K = 2$ for all cases. The column LNSX refers to the method presented in this chapter with the first $1 + X$ terms of the series, LS to Dufresne (2000), EE to Linetsky (2004), VEC to Vecer (2001, 2002), and MC 95% CI to the 95% confidence interval of the Monte-Carlo simulation.

For the first three cases we find virtually identical prices as Linetsky (2004), which is one of the most accurate benchmarks available in the literature. Remarkably, our method does not face any problems with the very low volatility in case 1. Many other existing method have serious difficulty with this parameterization. Indeed, the series of Dufresne (2000) does not even come close to the true price, while Linetsky (2004) requires 400 terms to obtain an accurate result. The price of Vecer (2001, 2002) is close to the true price, but still outside of the 95% Monte-Carlo confidence interval. Methods based on numerical inversion of the Laplace transform of Geman and Yor (1993) also struggle with low volatility because they involve numerical integration of highly oscillating integrands (see e.g., Fu et al. (1999)). When using exact arithmetic for case 1, our series with 20+1 terms agrees with the 400 term series of Linetsky (2004) to eight decimal places. When using double precision arithmetic, which was used for all numerical results in this section, the price agrees to four decimal places due to rounding errors. For cases 4 to 7, the LNS prices are slightly different from the EE benchmark. However, they are still very close and with the exception of case 4, they are all within the 95% confidence interval of the Monte-Carlo simulation.

Figure 2.1 plots the LNS price approximations for N ranging from 0 to 20, together with the Monte-Carlo price and the corresponding 95% confidence intervals as a benchmark.[3] We observe that the series converges very fast in all cases. In fact, truncating the series at $N = 10$ would give almost identical results. In theory, N can be chosen arbitrarily large, however in finite precision arithmetic it is inevitable that rounding errors start playing a role at some point. Remark that the prices for $N = 0$ and $N = 1$ are identical, which is a consequence of the fact that the auxiliary distribution matches the first moment of A_T. Figure 2.2 plots the simulated true density $g(x)$ and the approximating densities $g^{(0)}(x)$, $g^{(4)}(x)$, and $g^{(20)}(x)$ defined in (2.8). The true density was simulated using (B.3) in Appendix B.2, which is an extension of (2.9) using the density of the geometric average as a powerful control variate. Note that $g^{(0)}(x) = w(x)$,

[3] Figure 2.1 and 2.2 only show cases 1, 3, 5, and 7. The plots for the other cases look very similar and are available upon request.



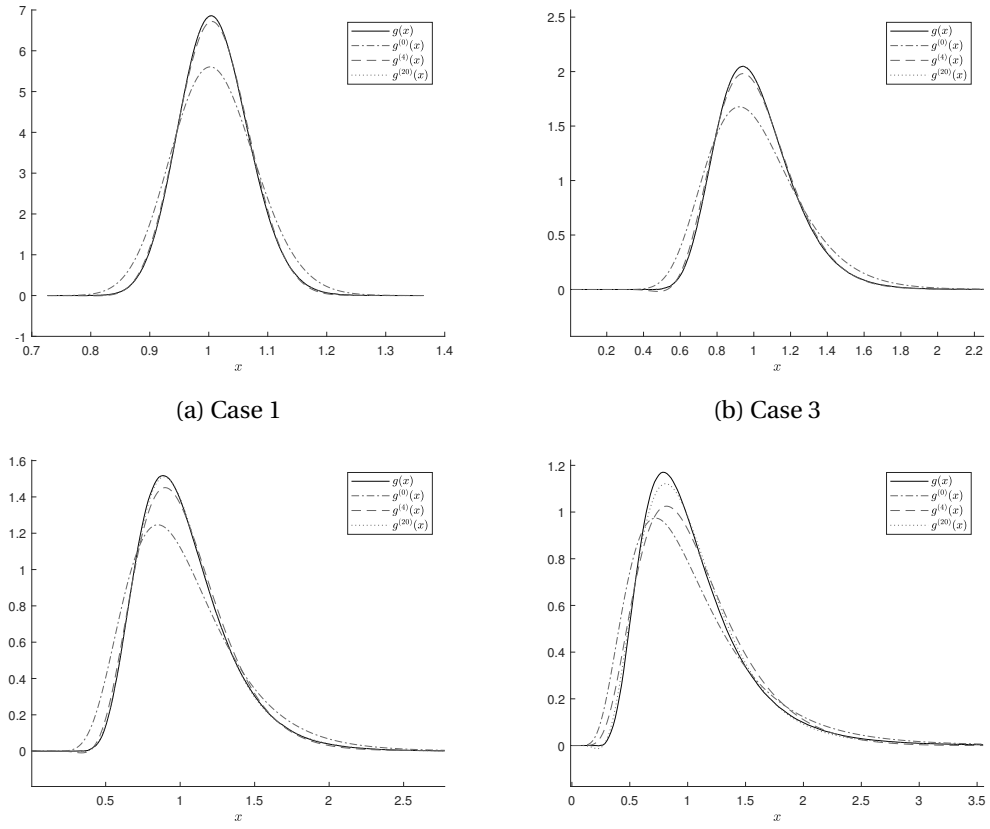
(a) Case 1

(b) Case 3

(c) Case 5

(d) Case 7

Figure 2.1 – Asian option price approximations in function of polynomial approximation order N. The cases correspond to different parameterizations shown in Table 2.1.



(a) Case 1

(b) Case 3

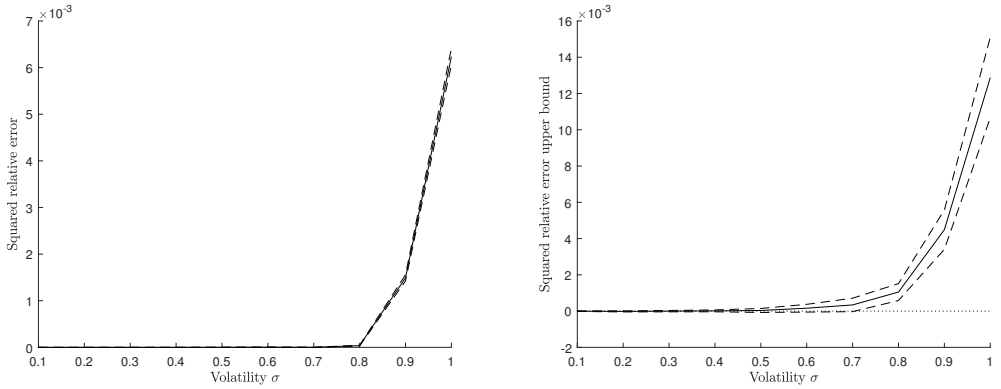
(c) Case 5

(d) Case 7

Figure 2.2 – Simulated true density function $g(x)$ and approximated density functions $g^{(n)}(x)$, $n = 0, 4, 20$. The cases correspond to different parameterizations shown in Table 2.1.

since $b_0(x) = \ell_0 = 1$. We can see that the approximating densities approach the true density as we include more terms in the expansion. In Figure 2.2a and 2.2b the approximation $g^{(20)}(x)$ is virtually indistinguishable from the true density. However, in Figure 2.2c and 2.2d there remains a noticeable difference between $g(x)$ and $g^{(20)}(x)$. This is consistent with the pricing errors we observed earlier in Table 2.1.

The above results indicate that for τ not too high, the LNS provides a very accurate approximation of the option price. This is not entirely surprising since τ determines the volatility parameter of the auxiliary log-normal density w, and hence how fast the tails of w go to zero. Loosely speaking, when τ is small, projecting the payoff or likelihood ratio function on the set of polynomials in L^2_w is almost like approximating a continuous function on a compact interval by polynomials. However, when τ becomes larger, the tails of w decay slowly and it becomes increasingly difficult to approximate a function with a polynomial in L^2_w. In other words, for larger values of τ, the moment indeterminacy of the log-normal distribution starts playing a more prominent role. A natural question is therefore whether this poses a problem for option pricing purposes. In Figure 2.3a we fix $T = 1$ and plot for a range of values for σ the



(a) Squared relative error (b) Squared relative error upper bound

Figure 2.3 – Squared relative approximation error for different values of σ. Dashed lines correspond to the 95% confidence intervals from the Monte-Carlo simulation. Parameters: $T = 1$, $r = 0.05$, $S_0 = K = 2$.

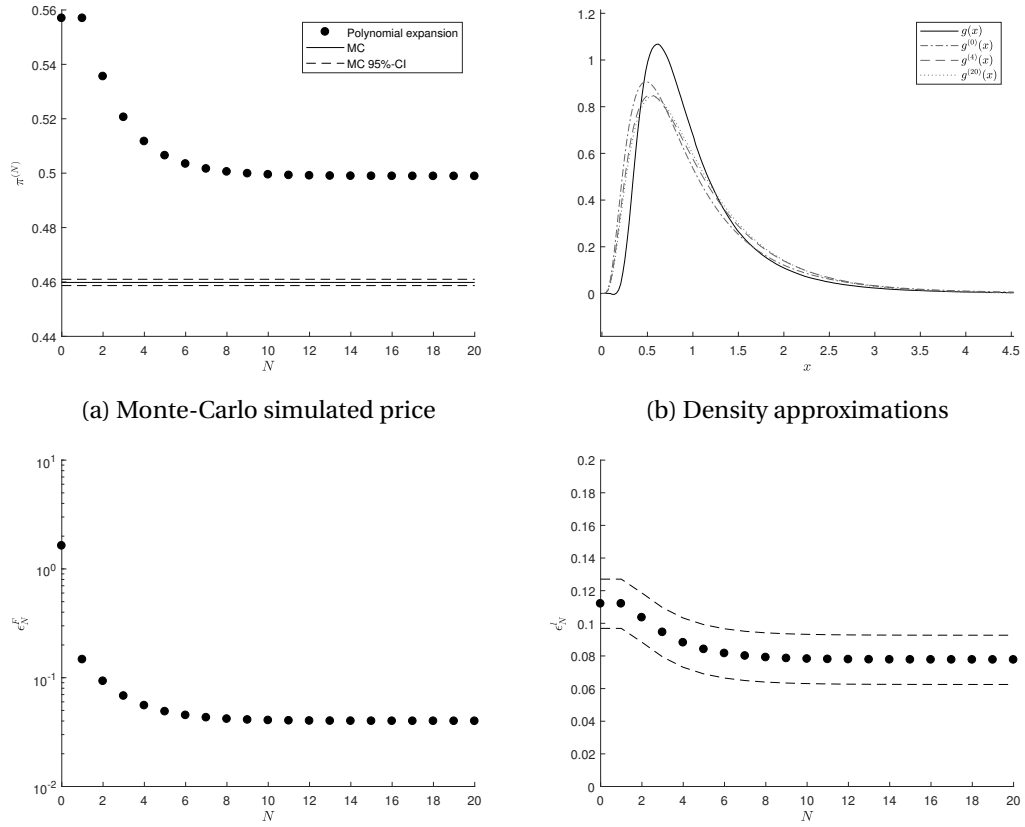
squared relative error

$$SRE = \left(\frac{\hat{\pi}^{MC} - \pi^{(20)}}{\pi^{(20)}} \right)^2,$$

where $\hat{\pi}^{MC}$ denotes the Monte-Carlo price estimate. The error starts to becomes noticeable around $\sigma = 80\%$, where $\sqrt{SRE} \approx 0.5\%$. For higher values of σ the error increases sharply. In Figure 2.3b we plot a more conservative estimate of the squared relative error using the upper bound in (2.10). This plot shows that the upper bound is only significantly different from zero for σ larger than approximately 70%. Figure 2.4 gives a more detailed insight in the extreme case of $\sigma = 100\%$. Although the LNS series converges relatively fast, it is clear from Figure 2.4a that it does not converge to the true price. The reason, as already mentioned before, is that the payoff and likelihood ratio functions are not accurately approximated by polynomials in the L^2_w-norm, as indicated by the projection errors in Figure 2.4c and 2.4d. We would obtain similar results by keeping σ fixed and varying the maturity T, since the crucial parameter for the asymptotic pricing error is $\tau = \sigma^2 T$. As a rule of thumb, we suggest to use the LNS method when $\tau \leq 0.5$.

The main advantage of the method proposed in this chapter is the ease of of its implementation and the computation speed. All terms in the series are fully explicit and involve only simple linear algebra operations. Table 2.2 shows the computation times of the LNS with $N \in \{10, 25, 20\}$, as well as the computation times of the benchmark methods.[4] The LNS com-

[4] For the LS, all symbolic calculations related to the moments of the reciprocal of the average have been precomputed using Matlab's Symbolic Math Toolbox. We use 15+1 terms in the series, a higher number of terms leads to severe rounding problems in double precision arithmetic. For the EE, the integral representation (16) in Linetsky (2004) has been implemented instead of the series representation (15). The implementation of the series representation involves partial derivatives of the Whittaker function with respect to its indices. These derivatives are not available in Matlab's Symbolic Math Toolbox and numerical finite difference approximations did not give accurate results. All numerical integrations are performed using Matlab's built-in function `integral`. For case 1, the numerical integration in the EE did not finish in a reasonable amount of time. For the VEC method, we use



(a) Monte-Carlo simulated price

(b) Density approximations

(c) Payoff projection error

(d) Likelihood projection error

Figure 2.4 – Visualization of the projection bias in the extreme volatility case. Dashed lines correspond to the 95% confidence intervals from the Monte-Carlo simulation. Parameters: $\sigma = 1$, $T = 1$, $r = 0.05$, $S_0 = K = 2$.

Case	r	σ	T	S_0	LNS10	LNS15	LNS20	LS	EE	VEC	MC
1	.02	.10	1	2.0	.006	.008	.009	.930	-	.277	6.344
2	.18	.30	1	2.0	.002	.002	.003	.666	2.901	.345	5.518
3	.0125	.25	2	2.0	.002	.002	.002	.635	3.505	.374	12.138
4	.05	.50	1	1.9	.001	.002	.003	.785	3.172	.404	6.819
5	.05	.50	1	2.0	.001	.002	.002	.701	2.768	.404	5.432
6	.05	.50	1	2.1	.001	.001	.002	.687	2.719	.398	5.452
7	.05	.50	2	2.0	.002	.002	.004	.594	2.202	.438	11.699

Table 2.2 – Computation times in seconds. The column LNSX refers to the method presented in this chapter with the first $1+X$ terms of the series, LS to Dufresne (2000), EE to Linetsky (2004), VEC to Vecer (2001, 2002), and MC to the Monte-Carlo simulation.

putation times are all in the order of miliseconds. Although the EE is very accurate, it comes at the cost of long computation times (in the order of several seconds) caused by the expensive evaluations of the Whittaker function. The LS does not require calls to special functions, however the method is slowed down by the numerical integration involved in computing the moments of the reciprocal of the average. The implementation of both the EE and LS require the use of software that can handle symbolic mathematics, in contrast to the implementation of the LNS. The VEC method is the fastest among the benchmarks considered in this chapter, but still an order of magnitude slower than the LNS.

2.6 Conclusion

We have presented a series expansion for the continuously sampled arithmetic Asian option using polynomials that are orthogonal with respect to the log-normal distribution. The terms in the series are fully explicit and do not require any numerical integration or special functions, which makes the method very fast. We have shown that the series does not diverge if the volatility of the auxiliary log-normal distribution is sufficiently high. However, the series is not guaranteed to converge to the true price. We have investigated this asymptotic bias numerically and found that its magnitude is related to the size of $\tau = \sigma^2 T$.

There are several extensions to our method. First of all, we can handle discretely monitored Asian options using exactly the same setup, but replacing the moments of the continuous average with those of the discrete average. The latter are easily computed using iterated expectations. Secondly, we only look at fixed-strike Asian options in this chapter. Since the process $\left(S_t, \int_0^t S_u \,\mathrm{d}u\right)$ is jointly a polynomial diffusion, we can compute all of its mixed moments. Our method can then be extended to price floating-strike Asian options by using a bivariate log-normal as auxiliary distribution. Finally, we can define the auxiliary density w as a mixture of log-normal densities, as studied in Ackerer and Filipović (2019b). Using a

Prof. Jan Vecer's Matlab implementation, which can be downloaded at http://www.stat.columbia.edu/~vecer/asiancontinuous.m. All computations are performed on a desktop computer with an Intel Xeon 3.50GHz CPU.

mixture allows to match higher order moments, which can lead to a faster convergence of the approximating series.

3 Exact Smooth Term Structure Estimation

We present a non-parametric method to estimate the discount curve from market quotes based on the Moore–Penrose pseudoinverse. The discount curve reproduces the market quotes perfectly, has maximal smoothness, and is given in closed-form. The method is easy to implement and requires only basic linear algebra operations. We provide a full theoretical framework as well as several practical applications.

3.1 Introduction

In financial models it is often assumed that we can observe an initial term structure of zero-coupon bond prices for the continuum of maturities, also known as the discount curve. In practice, however, zero-coupon bonds are rarely traded and the discount curve has to be derived from prices of actively traded fixed-income instruments such as coupon bonds, interest rate swaps or futures. Since the discount curve is an infinite-dimensional object, we need an interpolation method to complete the information obtained from the finite number of observed market instruments. Broadly speaking we can divide term structure estimation methods in two categories: parametric methods and non-parametric methods.

Parametric methods impose a particular functional form for (parts of) the discount curve and calibrate the parameters by minimizing the pricing error. Examples of single-piece functions that are defined over the entire maturity domain include the seminal work of Nelson and Siegel (1987) and Svensson (1994). These are typically low-dimensional parametric forms and are preferred for more qualitative studies where the general shape of the curve is more important than the exact values (e.g. monetary policy in central banks). Single-piece functional forms are however too restrictive for institutions involved in trading as they prefer to have a discount curve that perfectly reproduces market quotes in order to mark to market their books within a single arbitrage-free valuation framework. Rather than specifying a single function for the entire maturity spectrum, polynomial spline methods impose a piecewise polynomial specification. The first application in term structure estimation goes back to McCulloch (1971, 1975) where quadratic and cubic splines are fitted directly to the discount function using

ordinary least squares regressions. Steeley (1991) proposed the use of B-splines to overcome ill-conditioned matrices encountered in McCulloch (1971, 1975). We refer to Hagan and West (2006) for a survey of several other spline based algorithms. A close fit to market data can be achieved by increasing the number of knot points in the spline. The choice of both the number and the positions of the knot points remains, however, completely ad hoc.

A second class of estimation methods are the non-parametric approaches. Instead of imposing a particular functional form on (a transformation of) the discount curve, these methods minimize a norm that is related to the smoothness and goodness-of-fit of the curve. Several definitions of smoothness have been considered in the literature. Delbaen and Lorimier (1992) and Frishling and Yamamura (1996) minimize the integrated squared first derivative of the forward curve, arguing that forward rates over various harizons should not vary too much. Adams and Van Deventer (1994) and Lim and Xiao (2002), among others, use the integrated squared second order derivative of the forward curve as a measure of smoothness. Both approaches lead to polynomial splines for the optimal forward curve. Manzano and Blomvall (2004), Andersen (2007), and Kwon (2002) consider combinations of these two measures and show that this results in so called hyperbolic or tension splines. All of the above papers smooth a transformation of the discount curve (typically the forward curve) and numerical routines have to be invoked to solve for the optimal curve. Exceptions are the works of Delbaen and Lorimier (1992) and Adams and Van Deventer (1994) for the special case where the set of benchmark instruments consists solely of zero-coupon bonds. In reality, however, zero-coupon bonds are hardly ever liquidly traded and the discount curve has to be estimated based on coupon bearing bonds or swaps rates.

In this chapter we introduce an easy to use non-parametric method based on the Moore–Penrose pseudoinverse. We search in an infinite-dimensional Hilbert function space for a discount curve that has minimal norm and exactly prices a benchmark set of linear fixed-income instruments, e.g. FRAs, swaps, or coupon bonds. The norm is related to the integrated squared second derivative of the discount curve. The optimal discount curve is given by a cubic spline with knot points positioned exactly at the cashflow dates of the benchmark instruments. Because we directly smooth the discount curve function, the optimal curve is given in closed form and requires only simple linear algebra calculations. The methodology in this chapter closely resembles that of Lorimier (1995), Adams and Van Deventer (1994), Tanggaard (1997), Andersen (2007), and others, however they all focus on estimating transformations of the discount curve (e.g., the forward curve). To the best of our knowledge, this work is the first to present a fully worked out treatment of non-parametrically estimating the discount curve itself. We argue that our method is a valuable and easy to use alternative to more complex numerical algorithms to find a smooth discount curve.

Our method is designed to exactly reproduce the prices of benchmark instruments. This is common practice when the benchmark instruments are liquid Libor related instruments (e.g., swaps). The prices are typically taken to be the mid-prices. When building discount curves using coupon bonds, bid-ask ranges are often wider and a discount curve is in principle

allowed to produce any price that lies within this range. We show how to optimally pick the prices within bid-ask ranges such that the smoothness of the discount curve is maximally increased. In case the benchmark instruments are coupon bonds, we show that this reduces to solving a convex quadratic programming problem with linear inequality constraints.

As highlighted initially by Vasicek and Fong (1982) and Shea (1984), fitting a polynomial spline directly to the discount curve need not lead to a positive nor a monotonically non-increasing discount curve. Barzanti and Corradi (1998) use tension splines where the tension in the spline is increased manually until problematic behaviour is avoided. Chiu et al. (2008), Laurini and Moura (2010), and Fengler and Hin (2015), among others, impose shape constraints on the B-splines used to represent the discount function. The discount curve produced by our method is not guaranteed to be positive or monotonic non-increasing, however we did not find this to be a problem in the numerical examples we have explored. We develop a finite-dimensional counterpart of our method for which positivity and monotonicity constraints can easily be incorporated by numerically solving a convex quadratic programming problem with linear inequality constraints.

The remainder of this chapter is structured as follows. Section 3.2 casts the term structure estimation problem and shortly reviews the steps taken in a traditional bootstrap. Section 3.3 presents the theory behind our proposed method. In Section 3.4 we discuss the sensitivity of the optimal discount curve with respect to the input prices. In particular this section shows how to optimally choose prices from a bid-ask range. Section 3.5 illustrates our method with market data. Section 3.6 contains a finite-dimensional equivalent of our method. Section 3.7 concludes. All proofs can be found in Appendix C.

3.2 Estimation Problem

Suppose today is time 0 and denote by $p = (p_1, \dots, p_n)^\top$ the observed prices of n fixed-income instruments. Denote by $0 \leq x_1 < \cdots < x_N$ the union of all the cashflow dates of these instruments and call $C = (c_{ij})$ the corresponding $n \times N$ cashflow matrix.[1] If instrument i does not have a cashflow at time x_j, then we simply set $c_{ij} = 0$. The information contained in these n instruments about the discount curve can be summarized by a linear system as follows:

$$Cd = p, \tag{3.1}$$

with $d = (g(x_1), \dots, g(x_N))^\top$ and where $g(x)$ denotes the price of a zero-coupon bond maturing in x years. If C were an invertible square matrix, then there would exist a unique solution to this system: $d = C^{-1}p$. In reality, however, we typically have many more cashflow dates than instruments ($n \ll N$) that we can use for the estimation. In other words, the linear system $Cd = p$ is under-determined and there exist many discount vectors d that satisfy the relation

[1] By cashflow dates we mean every date that is relevant for the pricing of the instrument.

in (3.1).[2] The first problem that arises is therefore which of the admissible discount vectors should be chosen. Second, once we have chosen a particular admissible vector d, we still face an interpolation problem to find $g(x)$ for $x \in (0, x_1)$ and $x \in (x_{i-1}, x_i)$, $i = 2, \ldots, N$.

Bootstrapping is a common practice among trading desks to construct a discount curve from a limited number of carefully selected liquid market instruments such that the resulting curve perfectly reproduces the prices of the instruments used in the estimation. There is no unique bootstrapping method and it is likely that there are at least as many methods as there are trading desks in the world. In this section we give a very brief description of some methods, for a more detailed overview of the most popular (single-curve) bootstrapping methods used in practice we refer to Hagan and West (2006, 2008). In general, one can a priori impose an explicit parametric form for the discount curve: $g(x) = g(x; z)$ for some parameter z with dimension less than or equal to the number of observed instruments n. The pricing system (3.1) then becomes a system of possibly nonlinear equations in z:

$$C(g(x_1; z), \ldots, g(x_N; z))^\top = p.$$

Assuming that the gradients $\nabla_z g(x_j, z)$, $j = 1, \ldots, N$, are linearly independent, the inverse mapping theorem asserts that this system is no longer (locally) under-determined with respect to the parameter z. If it admits a solution z^* then it is (locally around z^*) unique. The choice of a suitable parametric form $g(x; z)$ is however not straightforward. One possibility is to choose a polynomial of degree $n-1$, also known as the Lagrange polynomial. Although this function is very smooth and flexible enough to satisfy n constraints, it demonstrates strong oscillatory behavior. A standard solution to this so called 'roller coaster' effect is to describe the discount curve by splines, i.e. piece-wise low-dimensional polynomials. There are many different ways to specify the functional form of a spline and the position of the knot points, see for example McCulloch (1971, 1975), Steeley (1991) and Adams (2001). It is important to note that in all these cases the spline solution is imposed *a priori* and both the number and the position of the knot points are chosen manually. The method we present in the next section also produces a spline. However, it remains fully non-parametric in the sense that the spline is the outcome of a proper optimization problem which determines the optimal number and position of the knot points.

3.3 Pseudoinverse on Hilbert Spaces

Instead of first finding a discount vector $d = (g(x_1), \ldots, g(x_N))^\top$ satisfying (3.1) and in a second step interpolating these discount factors to a continuous discount curve, we now directly search for a discount curve in a convenient Hilbert function space that is optimal in the sense of having minimal norm. The optimal discount curve is explicitly calculated through the pseudoinverse of a continuous linear map.

[2] Assuming the system is not inconsistent, i.e. instruments that can be replicated as a linear combination of other instruments must have the same price (no-arbitrage).

We fix a finite time to maturity horizon $\bar{\tau}$ large enough to contain all cashflow dates. We define the Hilbert space of discount curves H which consists of real functions $g\colon [0,\bar{\tau}] \to \mathbb{R}$ with absolutely continuous first derivative and norm given by

$$\|g\|_H^2 = \langle g, g\rangle_H = g(0)^2 + g'(0)^2 + \int_0^{\bar{\tau}} g''(x)^2 \,\mathrm{d}x. \tag{3.2}$$

This norm serves as a measure of smoothness for the discount curve, which approximately captures the 'flatness' of the corresponding forward curve. The forward rate is the rate one can lock in today on a riskless loan over a future time period. Unless there are specific reasons to believe otherwise, the forward rate should not fluctuate too much from one period to the next. If we denote by $F(x,y)$ the simple forward rate over a future time period $[x,y]$, $0<x<y$, then we have for small $h>0$:

$$\begin{aligned} g''(x) &\approx \frac{g(x-h)-2g(x)+g(x+h)}{h^2} \\ &= \frac{g(x)}{h}\left(F(x-h,x)+\frac{1}{h}\left(\frac{1}{1+hF(x,x+h)}-1\right)\right) \\ &\approx \frac{g(x)}{h}\Big(F(x-h,x)-F(x,x+h)\Big). \end{aligned}$$

Hence, by minimizing the curvature of the discount curve, we are approximately minimizing the difference between subsequent simple forward rates.

For any $\tau \in [0,\bar{\tau}]$ we now define the linear functional $\Phi_\tau : H \to \mathbb{R}$ which evaluates the discount curve at τ:

$$\Phi_\tau(g) = g(\tau). \tag{3.3}$$

By the Riesz representation theorem there exists a unique element $\phi_\tau \in H$ such that for any $g \in H$ we have

$$\Phi_\tau(g) = \langle \phi_\tau, g\rangle_H.$$

The following lemma gives an explicit expression for this element.

Lemma 3.3.1. *The linear functional Φ_τ on H can be uniquely represented by the element $\phi_\tau \in H$ given by*

$$\phi_\tau(x) = 1 - \frac{1}{6}(x\wedge\tau)^3 + \frac{1}{2}x\tau\,(2 + x\wedge\tau)\,,$$

where we write $x\wedge\tau := \min(x,\tau)$.

Let us define the linear map $M\colon H \to \mathbb{R}^n$ by

$$Mg = C(\Phi_{x_1}(g),\ldots,\Phi_{x_N}(g))^\top, \tag{3.4}$$

where C is just as before the $n\times N$ cashflow matrix. We henceforth assume that C has full rank. This is without loss of generality. Indeed, if C did not have full rank, we would be including redundant instruments in our estimation as they can be replicated by linear combinations

of other instruments. For example, two coupon bonds with different principal but otherwise identical characteristics impose the same constraints on the discount curve (assuming their prices are consistent with the law of one price).

We now find the discount curve with minimal H-norm that matches all benchmark quotes. That is, we solve the following infinite-dimensional optimization problem:

$$\begin{aligned} \min_{g \in H} \quad & \tfrac{1}{2}\|g\|_H^2 \\ \text{s.t.} \quad & Mg = p. \end{aligned} \tag{3.5}$$

The solution of (3.5) is an explicit piecewise cubic function, as shown in the following theorem:

Theorem 3.3.2. *There exists a unique solution $g^* \in H$ to the optimization problem* (3.5) *and it is given as*

$$g^*(x) = (M^+ p)(x) = z^\top \phi(x), \tag{3.6}$$

where $M^+ \colon \mathbb{R}^n \to H$ denotes the Moore–Penrose pseudoinverse of M, $z = C^\top \left(CAC^\top\right)^{-1} p$, $\phi(x) = (\phi_{x_1}(x), \ldots, \phi_{x_N}(x))^\top$, and A is the positive definite $N \times N$-matrix with components $A_{ij} = \phi_{x_i}(x_j) = \phi_{x_j}(x_i)$.

We have therefore explicitly constructed the discount curve $x \mapsto g^*(x)$ that exactly replicates the prices p of the instruments with cashflows C and moreover it is the smoothest curve to do so among all real functions with absolutely continuous first derivative in the sense that it minimizes the norm in (3.2). The corresponding instantaneous forward curve $f^* \colon [0, \bar{\tau}] \to \mathbb{R}$ is given explicitly by:

$$f^*(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \ln(g^*(x)) = -\frac{z^\top \phi'(x)}{z^\top \phi(x)},$$

where $\phi'(x) = (\phi'_{x_1}(x), \ldots, \phi'_{x_N}(x))^\top$ and $\phi'_{x_j}(x) = x_j - \frac{1}{2}(x \wedge x_j)^2 + x_j(x \wedge x_j)$.

Remark 3.3.3. *Gourieroux and Monfort (2013) characterize dynamic term structure models in which the zero-coupon bond prices are of the form $P(t,T) = z_t^\top a(T-t)$, where z_t is a set of $N \geq 1$ linearly independent stochastic factors and $a \colon \mathbb{R}_+ \to \mathbb{R}^N$ is a deterministic function. They show that the absence of arbitrage opportunity for a self-financed portfolio of zero-coupon bonds implies that there must exist a matrix M such that $a'(\tau) = Ma(\tau)$. The function $\phi = (\phi_{x_1}, \ldots, \phi_{x_N})^\top$ does not satisfy this requirement and hence* (3.6) *is not consistent with a dynamic term structure model. Similarly, the Nelson and Siegel (1987) and Svensson (1994) specifications are also not arbitrage-free in a dynamic sense, see e.g. Filipović (2000).*

The terms $g(0)^2$ and $g'(0)^2$ are included in (3.2) to guarantee the definiteness of the norm. Since the discount curve must start at face value, we henceforth impose $g(0) = 1$ by setting $x_1 = 0$, $p_1 = 1$, $c_{11} = 1$, and $c_{1j} = c_{i1} = 0$ for all $i, j \neq 1$. Hence, minimizing $g(0)^2$ does not influence the optimal curve since it is fixed in the constraints.

The term $g'(0)^2$ leads to a minimization of the instantaneous short rate. If this is not desirable, we can easily fix the short rate to an exogenously specified value $r \in \mathbb{R}$. Indeed, note that $\psi(x) = x$ is the Riesz representation of the linear functional $\Psi(g) = g'(0)$ in H. We now add $\Psi(g) = -r$ as an additional constraint in (3.5) and find (analogously as in the proof of Theorem 3.3.2) the following unique solution:

$$g^*(x) = \tilde{z}^\top \tilde{\phi}(x), \tag{3.7}$$

with $\tilde{z} = \tilde{C}^\top \left(\tilde{C}\tilde{A}\tilde{C}^\top\right)^{-1} \tilde{p}$, $\tilde{C} = \text{blkdiag}(C, 1)$, $\tilde{p} = \begin{pmatrix} p \\ -r \end{pmatrix}$, $\tilde{\phi}(x) = \begin{pmatrix} \phi(x) \\ \psi(x) \end{pmatrix}$, and $\tilde{A}$ is the positive definite $(N+1) \times (N+1)$ matrix with components

$$\tilde{A}_{ij} = \begin{cases} A_{ij} & i \le j \le N \\ x_i & i < j = N+1 \\ 1 & i = j = N+1 \end{cases}.$$

3.4 Discount Curve Sensitivites

The optimal discount curve (3.6) depends on the benchmark quotes through the vector $z = C^\top (CAC^\top)^{-1} p$. Depending on the type of benchmark instruments used, their quotes can enter through the price vector p or through the cashflow matrix C. For example, prices of coupon bonds enter through p, while swap rates and forward rates enter through C. The results in this section are derived for the curve in (3.6), however the results for the optimal curve (3.7) with constrained short rate directly follow by replacing p, C, z, A, ϕ with $\tilde{p}, \tilde{C}, \tilde{z}, \tilde{A}, \tilde{\phi}$, respectively.

3.4.1 Portfolio Hedging

The sensitivities of the optimal discount curve $g^*(x; p, C)$ with respect to the entries of p and C are most easily expressed using directional derivatives:

Lemma 3.4.1.

1. *The directional derivative $D_p g^* \cdot v \in H$ of the optimal discount curve g^* in* (3.6) *along a vector $v \in \mathbb{R}^n$ is given by*

$$\left(D_p g^* \cdot v\right)(x) = \sum_{i=1}^{n} v_i \frac{\partial g^*}{\partial p_i}(x) = c(v)^\top \phi(x),$$

   with $c(v) = C^\top \left(CAC^\top\right)^{-1} v \in \mathbb{R}^N$.

2. *The directional derivative $D_C g^* \cdot m \in H$ of the optimal discount curve g^* in* (3.6) *along a*

matrix $m \in \mathbb{R}^{n\times N}$ is given by

$$\left(D_C g^* \cdot m\right)(x) = \sum_{i=1}^{n}\sum_{j=1}^{N} m_{ij}\frac{\partial g^*}{\partial C_{ij}}(x) = f(m)^\top \phi(x),$$

with $f(m) = \left[m^\top - C^\top(CAC^\top)^{-1}(CAm^\top + mAC^\top)\right](CAC^\top)^{-1}p \in \mathbb{R}^N$.

These sensitivities can be used in practice to hedge a portfolio of securities against changes in the discount curve. Consider for example a bond portfolio which generates fixed cashflows c_k in τ_k years, $\tau_k \in [0,\bar{\tau}]$, $k = 1,\dots,K$, and denote its current value by V_{port}. Suppose that all benchmark instruments are coupon bonds.[3] A change Δp_i in the price of the i-th benchmark instrument leads to the following change ΔV_{port} in the value of the bond portfolio:

$$\Delta V_{port} = \sum_{i=1}^{n}\frac{\partial V_{port}}{\partial p_i}\Delta p_i = \sum_{i=1}^{n}\underbrace{\sum_{k=1}^{K} c_k\left(D_p g^* \cdot e_i\right)(\tau_k)}_{=:q_i}\Delta p_i,$$

where $e_i \in \mathbb{R}^n$ denotes the i-th canonical basis vector. Hence, we can hedge the bond portfolio against changes in the prices of the benchmark coupon bonds by purchasing $-q_i$ units of the i-th benchmark coupon bond. Andersen (2007) points out that such a hedging strategy only works well in practice if the discount curve construction produces 'local perturbations'. For example if bond i has a short maturity, then $(D_p g^* \cdot e_i)(x)$ should be zero for large x in order to avoid hedging long-term cashflows with short-term instruments. In general, cubic splines are known to perform poor with this respect and the above hedging strategy might therefore give unreasonable results with the discount curve construction presented in this chapter.

Hedging against individual small movements of the benchmark prices is however not necessarily consistent with the way interest rates move over time. Indeed, there is abundant empirical evidence that interest rate movements are attributable to a small number of stochastic factors often called level, slope, and curvature (see e.g., Litterman and Scheinkman (1991)). An alternative to hedging against changes in each benchmark price is therefore to directly hedge against interest rate movements that are deemed most likely. Specifically, we first build a discount curve $g^*(x)$ and then consider functional shifts $s_j(x)$, $j = 1,\dots,J$, to, for example, the corresponding forward curve $f^*(x)$. The sensitivity of $V_{port} = V_{port}(f^*)$ to these functional shifts can be expressed through the following functional derivative:

$$\left.\frac{\mathrm{d}V_{port}(f^* + \epsilon s_j)}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = -\sum_{k=1}^{K} c_k g^*(\tau_k)\int_0^{\tau_k} s_j(x)\,\mathrm{d}x, \quad j = 1,\dots J.$$

Next, we construct a hedging portfolio such that the functional derivatives of the hedged portfolio are equal to (or as close as possible to) zero.[4] The main advantage of this approach is

[3] A similar hedging strategy can be built if the benchmark instruments have quotes that enter through C using the directional derivative with respect to C from Lemma 3.4.1.

[4] We refer to section 6.4.2-6.4.3 in Andersen and Piterbarg (2010) for more details on this hedging approach.

that the method used to construct g^* does not play a major role in determining the hedging strategies (see e.g., Hagan and West (2006)). If $J = 1$ and $s_1(x) \equiv 1$, then we have a standard duration hedge. In order to hedge more than only parallel shifts in the forward curve, a popular choice in practice for s_j are piecewise triangular functions around a pre-defined set of so called key rate horizons $0 \leq \xi_1 < \cdots < \xi_J \leq \bar{\tau}$:

$$s_1(x) = \begin{cases} \frac{\xi_2 - x}{\xi_2 - \xi_1} & x \in [\xi_1, \xi_2] \\ 0 & else \end{cases}, \quad s_J(x) = \begin{cases} \frac{x - \xi_{J-1}}{\xi_J - \xi_{J-1}} & x \in [\xi_{J-1}, \xi_J] \\ 0 & else \end{cases}$$

$$s_j(x) = \begin{cases} \frac{x - \xi_{j-1}}{\xi_j - \xi_{j-1}} & x \in [\xi_{j-1}, \xi_j] \\ \frac{\xi_{j+1} - x}{\xi_{j+1} - \xi_j} & x \in [\xi_j, \xi_{j+1}] \\ 0 & else \end{cases}, \quad j = 2, \ldots, J-1.$$

3.4.2 Optimal Market Quotes

So far we have assumed that market quotes are observed without any error. In practice, however, we do not observe a single price but rather a bid-ask range. Any price in this range can be used to estimate the discount curve and this flexibility can be used to increase the smoothness of the discount curve.

The following lemma provides an explicit expression for the norm of the optimal discount curve we have derived before (i.e., for the case with equality constraints):

Lemma 3.4.2. *The squared norm of the optimal discount curve g^* in* (3.6) *is given by*

$$\|g^*\|_H^2 = p^\top (CAC^\top)^{-1} p.$$

We first assume that the benchmark instruments are coupon bonds for which we observe bid prices $p_b \in \mathbb{R}^n$ and ask prices $p_a \in \mathbb{R}^n$. We are now interested in solving the following optimization problem:

$$\begin{array}{ll} \min\limits_{p \in \mathbb{R}^n} & p^\top (CAC^\top)^{-1} p \\ \text{s.t.} & p_b \leq p \leq p_a \end{array}. \tag{3.8}$$

Remark that $(CAC^\top)^{-1} \in \mathbb{R}^{n \times n}$ is a positive definite matrix. Indeed, C is assumed to have full rank and A is positive definite as a consequence of the definiteness of the inner product. Hence, (3.8) is a convex quadratic programming problem where the unique solution p^* can easily be found using standard techniques. The optimal discount curve is then given by $g^* = M^+ p^*$.

Assume now that the benchmark instruments have quotes that enter through C. This is for example the case for swaps and FRAs. Denote the bid and ask quotes by α_b and α_a, respectively.

The optimization problem now becomes:

$$\begin{array}{ll} \min\limits_{\alpha\in\mathbb{R}^n} & p^\top (CAC^\top)^{-1} p \\ \text{s.t.} & \alpha_b \le \alpha \le \alpha_a \end{array} . \tag{3.9}$$

This problem is more difficult to solve than (3.8) because it is not necessarily a convex programming problem. However, we are able to compute the gradient explicitly:

Lemma 3.4.3. *The partial derivative with respect to α_i of the squared norm of the optimal discount curve in* (3.6) *is given by:*

$$\frac{\partial \|g^*\|_H^2}{\partial \alpha_i} = -2p^\top (CAC^\top)^{-1} \frac{\partial C}{\partial \alpha_i} AC^\top (CAC^\top)^{-1} p,$$

where $\frac{\partial C}{\partial \alpha_i}$ denotes the componentwise derivative of C with respect to the quote α_i.

We can therefore use a wide range of gradient-based constrained optimization algorithms. A similar idea was used by Kwon (2002), however his approach requires a numerical evaluation of the gradient at every iteration step. In contrast, we have the gradient in closed form which can be beneficial for both the computation time and accuracy of the numerical procedure.

3.5 Numerical Examples

In this section we discuss three practical illustrations of the pseudoinverse method using different types of benchmark instruments.

3.5.1 Coupon Bonds

In this example we estimate the discount curve from prices of coupon bonds. Specifically, we consider data from 4th of September 1996 on nine UK government bonds with semi-annual coupons and times to maturity varying approximately from 2 months to 12 years, see Table 3.1 for details. The vector p and the first ten columns (out of 104+1) of the matrix C are shown below:

$$p = \begin{pmatrix} 1 \\ 103.82 \\ 106.04 \\ 118.44 \\ 106.28 \\ 101.15 \\ 111.06 \\ 106.24 \\ 98.49 \\ 110.87 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 105 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 4.875 & 0 & 0 & 0 & 0 & \dots \\ 0 & 6.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6.125 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.5 & 0 & 0 & \dots \\ 0 & 0 & 0 & 3.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.875 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 4.25 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.875 & 0 & \dots \\ 0 & 0 & 4.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}.$$

The first row and column of C and p correspond to the restriction $g(0) = 1$. Figure 3.1a shows

	Coupon (%)	Next coupon	Maturity date	Dirty price (p_i)
Bond 1	10	15/11/96	15/11/96	103.82
Bond 2	9.75	19/01/97	19/01/98	106.04
Bond 3	12.25	26/09/96	26/03/99	118.44
Bond 4	9	03/03/97	03/03/00	106.28
Bond 5	7	06/11/96	06/11/01	101.15
Bond 6	9.75	27/02/97	27/08/02	111.06
Bond 7	8.5	07/12/96	07/12/05	106.24
Bond 8	7.75	08/03/97	08/09/06	98.49
Bond 9	9	13/10/96	13/10/08	110.87

Table 3.1 – Market data on UK gilts, 04/09/1996. Source: James and Webber (2000).

the continuously compounded yield curve and the instantaneous forward curve obtained from the pseudoinverse method presented in Section 3.3. The yield curve looks very smooth, but the forward curve exhibits oscillatory behavior that might be undesirable.

Next, we assume that the observed coupon bond prices are mid prices and we assume a relative bid-ask spread of 0.50% for every bond price. We use the approach outlined in Section 3.4.2 to find a discount curve which produces coupon bond prices within the bid-ask range. Figure 3.1b shows the resulting yield curve and instantaneous forward curve. We observe a significant increase in smoothness for the forward curve. Remark also the decrease of the short rate from approximately 5.5% to 4% as a consequence of the $g'(0)^2$ term in the norm definition (3.2). As explained at the end of Section 3.3 this can be avoided by fixing the short rate to an exogenous constant, for example $r = 5.5\%$. The result of this estimation is shown in Figure 3.1c and Figure 3.1d for the original and optimal prices, respectively.

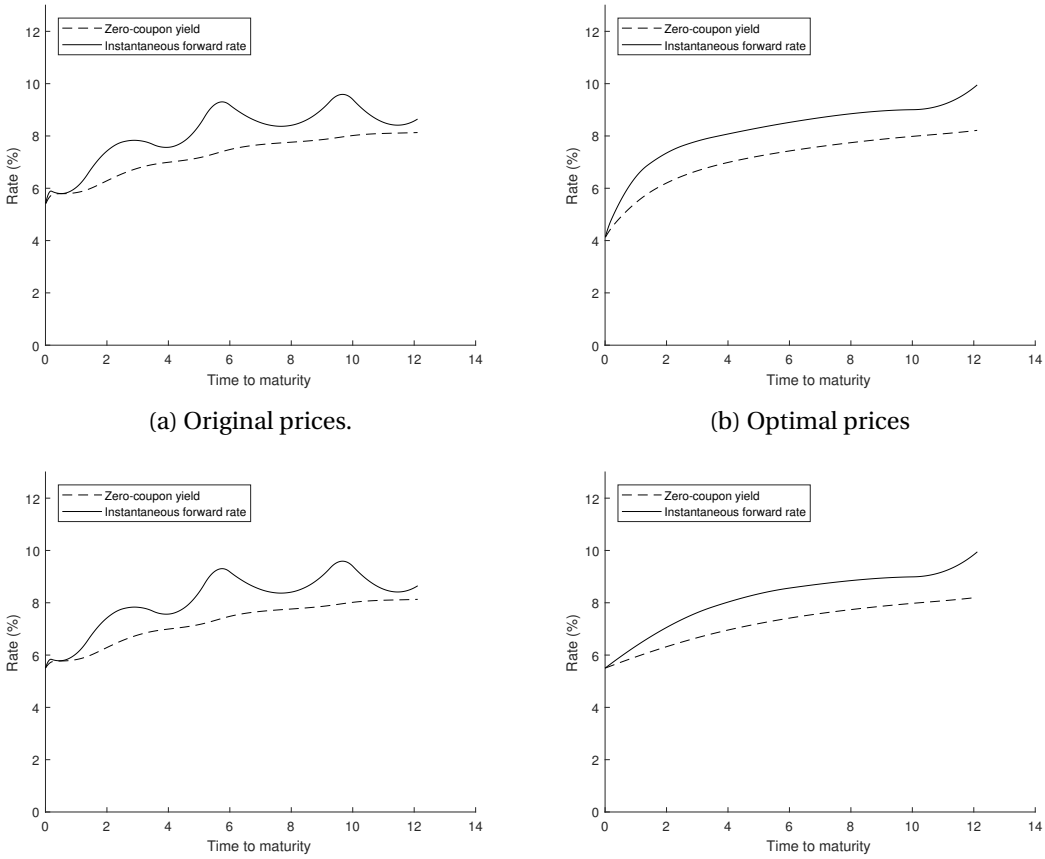
3.5.2 Libor Single Curve

In this example we use the same curve for both discounting cashflows as well as projecting forward rates. We consider data from the US money and swap markets as of 1st of October 2012, as shown in Table 3.2. More specifically we look at three USD Libor rates (overnight, 1M and 3M) with maturity dates $S = \{S_1, S_2, S_3\}$, five futures contracts on the 3M Libor and nine par swap rates with annual paying fixed leg. The futures contracts are quoted as:

$$100(1 - F_{futures}(T_{i-1}, T_i)), \quad i = 1, \ldots, 7,$$

with $F_{futures}(T_{i-1}, T_i)$ denoting the futures rate and $T = \{T_0, T_1, \ldots, T_7\}$ the corresponding reset/settlement dates. We ignore any convexity adjustments and take the futures rate as the simple forward rate to keep the estimation procedure model-independent.[5] Finally we denote

[5] The convexity adjustments are only a fraction of basis points because of the short maturities, so they do not make much qualitative difference.



(a) Original prices.

(b) Optimal prices

(c) Original prices, fixed $r = 5.50\%$.

(d) Optimal prices, fixed $r = 5.50\%$.

Figure 3.1 – Zero-coupon yield and instantaneous forward rate from the discount curve estimated using prices of UK government bonds.

Maturity date	Market quote (%)	Instrument
$S_1 = 04/10/2012$	0.1501	o/n Libor
$S_2 = 05/11/2012$	0.2135	1M Libor
$S_3 = 03/01/2013$	0.3553	3M Libor
$T_1 = 20/03/2013$	99.685	Futures
$T_2 = 19/06/2013$	99.675	Futures
$T_3 = 18/09/2013$	99.655	Futures
$T_4 = 18/12/2013$	99.645	Futures
$T_5 = 19/03/2014$	99.620	Futures
$U_2 = 03/10/2014$	0.361	Swap
$U_3 = 05/10/2015$	0.431	Swap
$U_4 = 03/10/2016$	0.564	Swap
$U_5 = 03/10/2017$	0.754	Swap
$U_7 = 03/10/2019$	1.174	Swap
$U_{10} = 03/10/2022$	1.683	Swap
$U_{15} = 04/10/2027$	2.192	Swap
$U_{20} = 04/10/2032$	2.405	Swap
$U_{30} = 03/10/2042$	2.579	Swap

Table 3.2 – Market quotes for Libor rates, futures prices and swap rates from the USD market as of 1st of October 2012. All contracts are spot ($T+2$) starting. Source: Bloomberg.

by $U = \{U_1, \ldots, U_{30}\}$ the cashflow dates of the swaps.

Traditional bootstrap

We first perform a traditional bootstrap where we interpolate all the missing simple spot and swap rates linearly. Remark first the overlapping cashflow dates of the different instruments:

$$S_1 < S_2 < T_0 < S_3 < T_1 < T_2 < T_3 < U_1 < T_4 < T_5 < U_2 < \cdots < U_{30}. \tag{3.10}$$

The prices of the discount bonds $g(S_1)$, $g(S_2)$, and $g(S_3)$ are readily obtained from the given Libor rates. At the reset date of the first futures contract, we obtain the simple spot rate $L(T_0)$ by linear interpolation of the last two Libor rates:

$$L(T_0) = wL(S_2) + (1-w)L(S_3), \quad w = \frac{\delta(T_0, S_3)}{\delta(S_2, S_3)},$$

where $\delta(x, y)$ denotes the day count fraction between dates x and y. The discount factor $g(T_0)$ is now recovered from the interpolated simple spot rate $L(T_0)$. Treating the futures rates as simple forward rates, we have all the information needed to compute $g(T_i)$, $i = 1, \ldots, 5$, iteratively from:

$$P(T_i) = \frac{g(T_{i-1})}{1 + \delta(T_{i-1}, T_i)F(T_{i-1}, T_i)}.$$

For the swaps we exploit again the overlapping cashflow dates to obtain $g(U_1)$ by linearly interpolating between $L(T_3)$ and $L(T_4)$. The swap rate $R_{swap}(U_1)$ of the swap with just one cashflow at U_1 can be calculated from the discount bond price $g(U_1)$. The remaining discount factors are now obtained by iterated use of the formula:

$$g(U_i) = \frac{1 - R_{swap}(U_i)\sum_{j=1}^{i-1}\delta(U_{j-1},U_j)g(U_j)}{1 + R_{swap}(U_i)\delta(U_{i-1},U_i)}, \quad i = 2,\ldots,30,$$

where all the missing swap rates are obtained by linear interpolation.

Finally, we obtain the discount curve for the continuum of maturities by linearly interpolating the zero-coupon bond yields between cashflow dates. Figure 3.2a and Figure 3.2b show the zero-coupon bond yields and the instantaneous forward rates[6]. Although the zero-coupon yield curve looks fairly smooth, the same cannot be said of the instantaneous forward curve. This well known 'sawtooth'-behaviour of the forward curve is a consequence of the linear interpolation. Other interpolation techniques may lead to improved smoothness of the forward curve, however choosing the correct technique remains somewhat arbitrary and can lead to a significant increase in complexity.

Pseudoinverse

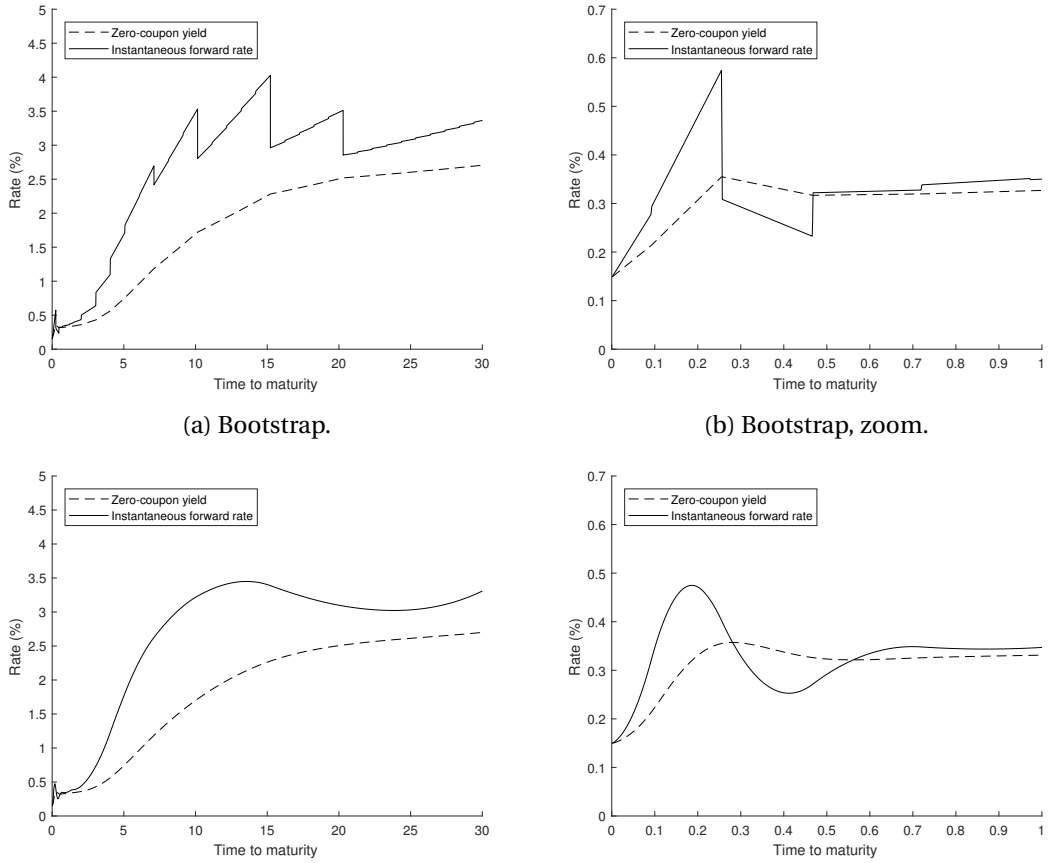
The pseudoinverse method that we introduced in Section 3.3 does not require any ad hoc interpolation. We only have to construct the cashflow matrix and the smoothness maximization criterion uses the remaining degrees of freedom in an optimal way. The cashflow matrix C in this example has dimension 18×40, one row for every observed instrument and an additional row to impose $g(0) = 1$. The columns represent all the dates relevant in the valuation of the instruments.

The Libor rates $L(S_i)$, $i = 1,2,3$, can be represented as instruments that have price 1 today and cashflow $1 + \delta(0,S_i)L(S_i)$ at time S_i. The simple forward rates $F(T_{i-1},T_i)$, $i = 1,\ldots,5$, can be seen as instruments with price 0 today, cashflow -1 at time T_{i-1} and another cashflow of $1 + \delta(T_{i-1},T_i)F(T_{i-1},T_i)$ at time T_i. For the swaps with maturity U_i, $i = 2,3,4,5,7,10,15,20,30$, we recall the definition of the par swap rate $R_{swap}(U_i)$:

$$1 - g(U_i) = R_{swap}(U_i)\sum_{j=1}^{i}\delta(U_{j-1},U_j)g(0,U_j)$$

where we set $U_0 := 0$. We see that this is equivalent to an instrument with price 1 today, cashflow $\delta(U_{j-1},U_j)R_{swap}(U_i)$ at time U_j, $j = 1,\ldots,i-1$, and a final cashflow $1 + \delta(U_{i-1},U_i)R_{swap}(U_i)$ at time U_i. The vector p and the first 13 columns of the matrix C therefore take the following form:

[6] We have approximated the instantaneous forward rates using first order finite differences on a fine grid.



(a) Bootstrap.

(b) Bootstrap, zoom.

(c) Pseudoinverse.

(d) Pseudoinverse, zoom.

Figure 3.2 – Zero-coupon yield and instantaneous forward rate from the discount curve estimated using USD Libor swaps and futures.

$$
p = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad C = \begin{array}{c} \begin{array}{ccccccccccccccc} 0 & S_1 & S_2 & T_0 & S_3 & T_1 & T_2 & T_3 & U_1 & T_4 & T_5 & U_2 & U_3 & \cdots & \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & c_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & c_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & c_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & -1 & 0 & c_{45} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & -1 & c_{56} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & c_{67} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & c_{79} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & c_{8,10} & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{98} & 0 & 0 & c_{9,11} & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{10,8} & 0 & 0 & c_{10,11} & c_{10,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{11,8} & 0 & 0 & c_{11,11} & c_{11,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{12,8} & 0 & 0 & c_{12,11} & c_{12,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{13,8} & 0 & 0 & c_{13,11} & c_{13,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{14,8} & 0 & 0 & c_{14,11} & c_{14,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{15,8} & 0 & 0 & c_{15,11} & c_{15,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{16,8} & 0 & 0 & c_{16,11} & c_{16,12} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{17,8} & 0 & 0 & c_{17,11} & c_{17,12} & \cdots \end{pmatrix} \end{array} \begin{array}{c} \\ g(0)=1 \\ Libor \\ Libor \\ Libor \\ Futures \\ Futures \\ Futures \\ Futures \\ Futures \\ Swap \\ Swap \\ Swap \\ Swap \\ Swap \\ Swap \\ Swap \\ Swap \\ Swap \end{array}.
$$

In Figure 3.2c and Figure 3.2d we have plotted zero-coupon yields and instantaneous forward rates. We observe that the pseudoinverse method produces a substantially smoother forward curve than the one in Figure 3.2a. Note that fixing the short rate to an exogenous constant would not change much here since we have included the overnight Libor rate as a benchmark instrument.

As an example of a hedging strategy, we consider a payer swap maturing in 13 years. We want to hedge this swap with the nine receiver swaps used in our estimation through the hedging approach outlined in Section 3.4.1 with triangular functional shifts to the forward curve. Following Andersen and Piterbarg (2010) we choose to use key rate horizons ξ_j, $j = 1, \dots, J$, spaced three months apart over the interval $[0, \bar{\tau}]$. Note that $J > 9$, which means we cannot build a perfect hedge but only an approximate one (in a least squares sense). The hedging quantities are shown in Figure 3.3. The hedging strategy consists roughly in combining the swaps with maturity in 10 and 15 years. This makes sense intuitively since these are the two hedging instruments whose cashflows resemble the one of the 13 year swap the most.

3.5.3 Libor Multi Curve

After the credit crisis of 2008 it became clear that using one and the same curve for both discounting and projecting cashflows was no longer realistic. Today's market standard is to use overnight indexed swaps (OIS) to extract a risk-free curve for the purpose of discounting cashflows and to use separate curves to project forward rates with different tenors. We show in this example how the pseudoinverse method can easily be used to extract all these different curves from market data.

Table 3.3 shows quotes of four different swap instruments from the eurozone market as of

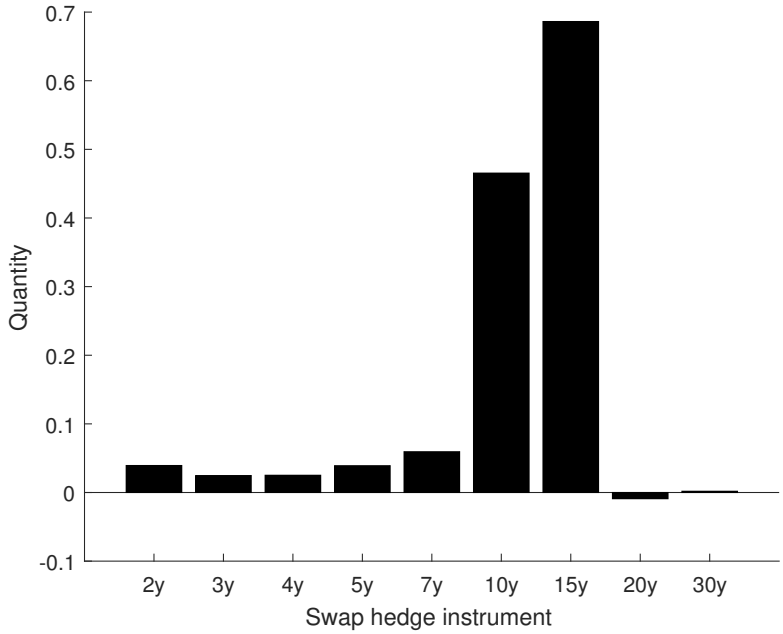


Figure 3.3 – Hedging a 13 year swap using the nine swaps used in the estimation as hedging instruments.

4th of November 2013. The first is an OIS that pays fixed and receives floating tied to Eonia. Eonia swaps with maturity longer than one year have annual payment frequency on both legs while those with maturity less than one year only have a cashflow at maturity. The second swap instrument pays fixed and receives floating tied to 6M Euribor. The 6M Euribor swap has annual payments on its fixed leg and semi-annual on the floating leg. The remaining two swap instruments are basis swaps that swap cashflows tied to floating rates. The first one swaps 3M Euribor against 6M Euribor and the second one 1M Euribor against 6M Euribor. These swaps are quoted in terms of the spread that has to be added to the shorter leg such that the two legs have identical present value.

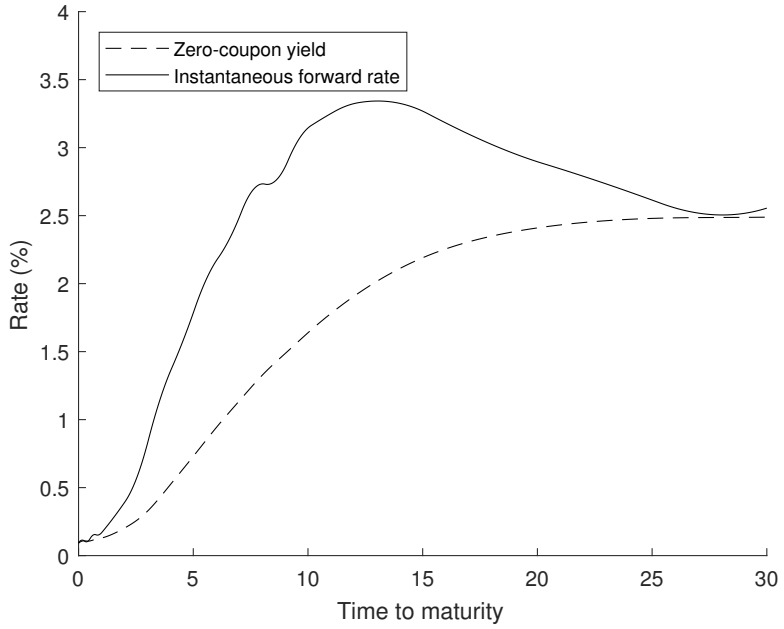
The curve g_{OIS} used for discounting is extracted from the OIS quotes. This can be done with the pseudoinverse method in exactly the same way as in the single curve example. The corresponding yield and instantaneous forward curve are plotted in Figure 3.4a. The 6M Euribor swap and the two basis swaps have payment frequencies that are multiples of one month. In the following we therefore consider a time grid $T = \{t_1, \dots, t_N\}$ where t_i and t_{i-1}, $i = 1, \dots, N$, are one month apart and $N = 30 \times 12$.

We start with the instruments that swap the 6M Euribor against a fixed rate. These are quoted in terms of the swap rate K that equates the value of the fixed and the floating leg:

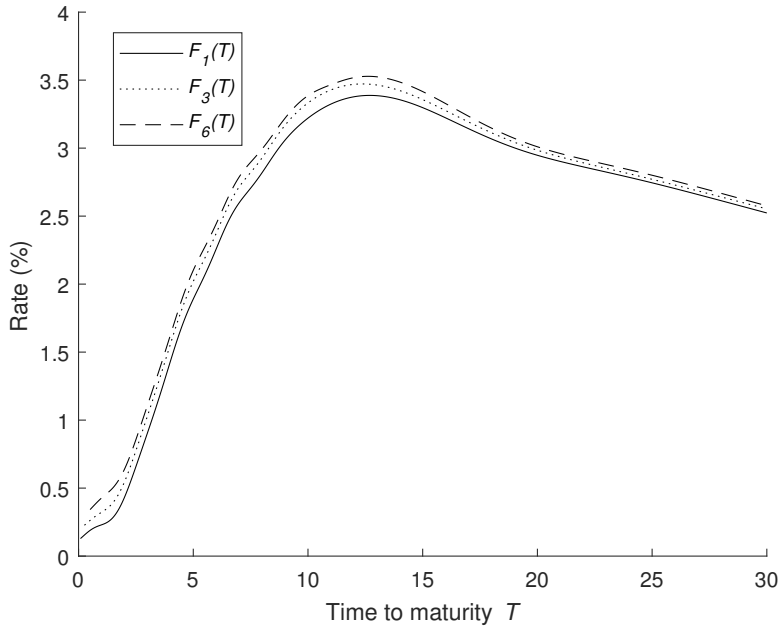
$$K \sum_{i=1}^{n} \delta(t_{12(i-1)}, t_{12i}) g_{OIS}(t_{12i}) = \sum_{i=1}^{2n} \delta(t_{6(i-1)}, t_{6i}) g_{OIS}(t_{6i}) F_6(t_{6i}),$$

Tenor	Eonia-Fix (%)	6M-Fix (%)	3M-6M (bps)	1M-3M (bps)	Cash (%)
o/n	0.092				
1m	0.102				0.129
3m	0.109				0.227
6m	0.108				0.342
9m	0.121				
1y	0.130	0.386	10.85	8.90	
2y	0.205	0.482	12.15	10.90	
3y	0.334	0.656	12.85	12.60	
4y	0.533	0.870	13.00	13.70	
5y	0.742	1.097	12.95	14.20	
6y	0.952	1.306	12.70	14.30	
7y	1.145	1.503	12.35	14.20	
8y	1.328	1.677	11.90	14.00	
9y	1.479	1.833	11.40	13.80	
10y	1.625	1.973	10.90	13.60	
11y	1.757	2.095			
12y	1.872	2.199			
15y	2.124	2.418	8.85	11.80	
20y	2.317	2.570	7.40	10.10	
25y	2.385	2.618	6.50	8.90	
30y	2.406	2.625	5.90	8.10	

Table 3.3 – Mid swap rates for Eonia swaps, 6M Euribor swaps, 3M-6M basis swaps, 1M-6M basis swaps and Euribor cash rates as of 04/11/2013. Source: Bloomberg.



(a) Zero-coupon yield and instantaneous forward curve corresponding to the OIS discount curve g_{OIS}.



(b) Forward curves with tenors one, three, and six months.

Figure 3.4 – Multicurve term structure estimation from Euribor and Eonia swaps with pseudoinverse method.

where $t_0 = 0$, t_{12n} is the maturity of the swap, and $F_k(t_i)$ denotes the k-month simple forward rate with reference period $[t_{i-k}, t_i]$. With our estimate for the OIS discount curve we are able to value the fixed leg of this swap. Notice that the right-hand side is a linear function of the unknown 6M forward rates with known coefficients. In other words, using our methodology from before we are able to extract the smoothest possible 6M forward curve that exactly fits the quoted swap rates. The pricing system therefore becomes $Cf = p$, where $f = (F_6(t_6), \ldots, F_6(t_N))^\top$, the price vector p takes the form

$$p = \left(F_6(t_6),\, K\delta(t_0, t_{12}) g_{OIS}(t_{12}),\, \ldots,\, K \sum_{i=1}^{30} \delta(t_{12(i-1)}, t_{12i}) g_{OIS}(t_{12i}) \right)^\top$$

and the first three rows of the 'cashflow' matrix C for become:

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \delta(t_0,t_6) g_{OIS}(t_6) & \delta(t_6,t_{12}) g_{OIS}(t_{12}) & 0 & 0 & 0 & \cdots \\ \delta(t_0,t_6) g_{OIS}(t_6) & \delta(t_6,t_{12}) g_{OIS}(t_{12}) & \delta(t_{12},t_{18}) g_{OIS}(t_{18}) & \delta(t_{18},t_{24}) g_{OIS}(t_{24}) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Next to the quotes of the swaps, we also included here the given 6M Euribor spot rate $F_6(t_6)$ by adding the row $(1, 0, \ldots, 0)$ to the matrix C and the rate to the vector p.

In a next step we use the 6M forward curve to extract the 3M forward curve from the 3M-6M tenor basis swaps. Basis swaps are quoted in terms of the spread S that has to be added to the leg with the highest frequency in order to equate the value of both legs:

$$\begin{aligned} &\sum_{i=1}^{2n} \delta(t_{6(i-1)}, t_{6i}) g_{OIS}(t_{6i}) F_6(t_{6i}) \\ =&\sum_{i=1}^{4n} \delta(t_{3(i-1)}, t_{3i}) g_{OIS}(t_{3i}) (F_3(t_{3i}) + S). \end{aligned}$$

Rearranging terms we get:

$$\begin{aligned} &\sum_{i=1}^{2n} \delta(t_{6(i-1)}, t_{6i}) g_{OIS}(t_{6i}) F_6(t_{6i}) - S \sum_{i=1}^{4n} \delta(t_{3(i-1)}, t_{3i}) g_{OIS}(t_{3i}) \\ =&\sum_{i=1}^{4n} \delta(t_{3(i-1)}, t_{3i}) g_{OIS}(t_{3i}) F_3(t_{3i}). \end{aligned}$$

The left-hand side can be evaluated with the discount and 6M forward curve we have extracted earlier and the right-hand side is a linear function of the unknown 3M forward rates. We can therefore use the pseudoinverse method to extract the 3M forward curve. Once again we can also easily include the spot 3M Euribor rate $F_3(t_3)$. In a very similar fashion we also obtain the 1M forward curve from the 1M-3M basis swap quotes, where we also include the spot 1M Euribor rate $F_1(t_1)$. All three curves are plotted in Figure 3.4b.

Remark 3.5.1. *Although Eonia swaps are quoted up to 60 years of maturity, it is sometimes*

argued that the Eonia swaps with maturity longer than 1 year are not sufficiently liquid to be used in the construction of the OIS discount curve. The remaining part of the curve is instead often calculated from OIS-3M basis swaps, which are more actively traded. This does however create a circular dependency between the OIS discount curve, the 3M forward and 6M forward curves. The easy extension of the pseudoinverse method to the multi-curve world was mainly due to the fact that we can estimate all the curves one by one. If we have to estimate a part of the OIS discount curve from OIS-3M swaps, then we need to solve for all three curves at once. This increases the complexity of the problem because in the swap valuations the forward rates are multiplied with the discount rates, i.e. we face non-linear constraints in the optimization problem. One possible workaround is to jointly estimate the OIS discount curve $g_1(x) := g_{OIS}(x)$ *and the 'discounted' forward curves* $g_2(x) := g_{OIS}(x)F_3(x)$, $g_3(x) := g_{OIS}(x)F_6(x)$. *By smoothing the discount curve and discounted forward curves, we are again solving an optimization problem with linear constraints:*

$$\begin{aligned} \min_{g_1,g_2,g_3 \in H} \quad & \|g_1\|_H^2 + \|g_2\|_H^2 + \|g_3\|_H^2 \\ s.t. \quad & M_1 g_1 + M_2 g_2 + M_3 g_3 = p, \end{aligned}$$

for some appropriately defined linear maps M_1, M_2, *and* M_3. *We leave the implementation of this extension for further research.*

3.6 Pseudoinverse on the Euclidean Space

For readers unfamiliar with infinite-dimensional Hilbert spaces, we introduce in this section a finite-dimensional analogue of the method introduced in Section 3.3. Instead of looking for a discount curve in a Hilbert function space, we now search for a vector of discount factors in the Euclidian space which has maximal smoothness in some sense. Suppose we are interested in the discount factors at dates $0 = u_1 < \cdots < u_K$:

$$d = (g(u_1), \ldots, g(u_K))^\top,$$

for some $K \geq 1$ and (for simplicity) $u_{i+1} - u_i \equiv \delta > 0$. Suppose furthermore that d contains all the discount factors that are required to value the n instruments we observe, i.e. $\{x_1, \ldots, x_N\} \subseteq \{u_1, \ldots, u_K\}$. Redefine $C \in \mathbb{R}^{n \times K}$ as the cashflow matrix of the n benchmark instruments on the dates $\{u_1, \ldots, u_K\}$.

We cast the smoothness criterion (3.2) in a discrete form using a left Riemann sum for the

integral and forward finite differences for the derivatives (other choices are possible of course):

$$\begin{aligned}
&g(0)^2 + g'(0)^2 + \int_0^{u_{K-1}} g''(x)^2 \,\mathrm{d}x \\
&\approx g(u_0)^2 + \frac{1}{\delta}(g(u_1) - g(u_0))^2 + \sum_{i=0}^{K-2} \left(\frac{g(u_{i+2}) - 2g(u_{i+1}) + g(u_i)}{\delta^2} \right)^2 \delta \\
&= \|Ad\|_K^2,
\end{aligned}$$

where $\|\cdot\|_K^2$ denotes the Euclidian norm on $\mathbb{R}^K$ and

$$A = \mathrm{diag}(1, \delta^{-1/2}, \delta^{-3/2}, \ldots, \delta^{-3/2}) \begin{pmatrix} 1 & 0 & \ldots & \ldots & \ldots & 0 \\ -1 & 1 & 0 & & & \vdots \\ 1 & -2 & 1 & \ddots & & \vdots \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{K \times K}.$$

The finite dimensional optimization problem now becomes:

$$\begin{aligned} \min_{d \in \mathbb{R}^K} \quad & \tfrac{1}{2}\|Ad\|_K^2 \\ \text{s.t.} \quad & Cd = p \end{aligned} \,. \tag{3.11}$$

The above is a convex quadratic programming problem with linear inequality constraints, for which we obtain the following solution:

Theorem 3.6.1. *There exists a unique solution $d^* \in \mathbb{R}^K$ to the optimization problem* (3.11) *and it is given as*

$$d^* = A^{-1} M^+ p,$$

where $M = CA^{-1}$ and $M^+ = M^\top (MM^\top)^{-1}$ is the Moore–Penrose pseudoinverse of the matrix M.

Remark that in the finite-dimensional case it becomes very easy to impose positivity and monotonicity constraints on the discount factors:

$$\begin{aligned} \min_{d \in \mathbb{R}^K} \quad & \tfrac{1}{2}\|Ad\|_K^2 \\ \text{s.t.} \quad & Cd = p \\ & d_1 > \cdots > d_K > 0 \end{aligned} \,.$$

We therefore obtain a convex quadratic programming problem with linear inequality constraints. Such a problem has a unique solution that can easily be found with established algorithms implemented in many numerical software packages.

3.7 Conclusion

We have introduced a novel method based on the Moore–Penrose pseudoinverse to extract a discount curve that exactly reproduces the prices of the benchmark instruments and that has maximal smoothness in the sense that it has minimal integrated squared second-order derivatives. The optimal discount curve is a piecewise-cubic function and is obtained as the unique solution of an infinite-dimensional optimization problem. Bid-ask spreads can be incorporated to further increase the smoothness of the discount curve. The pseudoinverse method is very easy to implement, making it an interesting method of first resort before considering more complex alternatives.

A Appendix to Chapter 1

A.1 Bootstrapping an Additive Seasonality Function

In this section we explain how to bootstrap a smooth curve $T \mapsto f_t(T)$ of (unobserved) futures prices corresponding to the instantaneous dividend rate D_T. The curve should perfectly reproduce observed dividend futures prices and in addition incorporate a seasonality effect. Once we have this function, we define the function $\delta(T)$ as

$$\delta(T) = f_t(T) - p^\top(\beta \mathrm{Id} + G_1)\mathbb{E}_t[H_1(X_T)], \quad T \geq t,$$

so that the specification in (1.25) perfectly reproduces observed futures contracts and incorporates seasonality.

Suppose for notational simplicity that today is time 0 and we observe the futures prices F_i of the dividends realized over one calendar year $[i-1, i]$, $i = 1, \ldots, I$. Divide the calendar year in $J \geq 1$ buckets and assign a seasonal weight $w_j \geq 0$ to each bucket, with $w_1 + \cdots + w_J = 1$. These seasonal weights can for example be estimated from a time series of dividend payments. We search for the twice continuously differentiable curve f_0 that has maximal smoothness subject to the pricing and seasonality constraints:

$$\begin{aligned} \min_{f_0 \in C^2(\mathbb{R})} \quad & f_0(0)^2 + f_0'(0)^2 + \int_0^I f_0''(u)^2 \, \mathrm{d}u \\ \text{s.t.} \quad & \int_{i-1+\frac{j-1}{J}}^{i-1+\frac{j}{J}} f_0(u) \, \mathrm{d}u = w_j F_i, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J. \end{aligned}$$

This can be cast in an appropriate Hilbert space as a convex variational optimization problem with linear constraints. In particular, it has a unique solution that can be solved in closed form using techniques similar to the ones presented in Chapter 3 (cf., Filipović and Willems (2018)). By discretizing the optimization problem, a non-negativity constraint on f can be added as well.

A.2 Proofs

This section contains all the proofs of the propositions in the paper.

Proof of Proposition 1.2.2

Using the moment formula (1.3) we have for $t \leq T$

$$\mathbb{E}_t[C_T] = \mathrm{e}^{\beta T} p^\top \mathrm{e}^{G_1(T-t)} H_1(X_t).$$

Differentiating with respect to T gives

$$\frac{\mathrm{d}\mathbb{E}_t[C_T]}{\mathrm{d}T} = \beta \mathrm{e}^{\beta T} p^\top \mathrm{e}^{G_1(T-t)} H_1(X_t) + \mathrm{e}^{\beta T} p^\top G_1 \mathrm{e}^{G_1(T-t)} H_1(X_t).$$

The result now follows from

$$D_t = \left.\frac{\mathrm{d}\mathbb{E}_t[C_T]}{\mathrm{d}T}\right|_{T=t}.$$

Proof of Proposition 1.2.3

Plugging in the specifications for ζ_t and D_t in (1.11) gives:

$$S_t^* = \frac{1}{\zeta_t} \int_t^\infty \mathrm{e}^{-(\gamma-\beta)s} \mathbb{E}_t \left[p^\top (\beta \mathrm{Id} + G_1) H_1(x) H_1(X_s)\, q^\top H_1(X_s) \right] \mathrm{d}s.$$

Since X_t is a polynomial process, we can find a closed form expression for the expectation inside the integral:

$$\mathbb{E}_t \left[p^\top (\beta \mathrm{Id} + G_1) H_1(X_s)\, q^\top H_1(X_s) \right] = v^\top \mathrm{e}^{G_2(s-t)}\, H_2(X_t).$$

The fundamental stock price therefore becomes:

$$\begin{aligned}
S_t^* &= \frac{\mathrm{e}^{\beta t} v^\top}{q^\top H_1(X_t)} \int_t^\infty \mathrm{e}^{-(\gamma-\beta)(s-t)} \mathrm{e}^{G_2(s-t)}\, \mathrm{d}s\, H_2(X_t) \\
&= \frac{\mathrm{e}^{\beta t} v^\top}{q^\top H_1(X_t)} \left(G_2 - (\gamma-\beta)\,\mathrm{Id}\right)^{-1} \exp\left\{\left(G_2 - (\gamma-\beta)\,\mathrm{Id}\right)(s-t)\right\} \Big|_{s=t}^{s=\infty} H_2(X_t) \\
&= \frac{\mathrm{e}^{\beta t} v^\top}{q^\top H_1(X_t)} \left((\gamma-\beta)\,\mathrm{Id} - G_2\right)^{-1} H_2(X_t) \\
&< \infty,
\end{aligned}$$

where we have used the fact that the eigenvalues of the matrix $G_2 - (\gamma-\beta)\,\mathrm{Id}$ have negative real parts.

Proof of Proposition 1.2.4

The market is arbitrage free if and only if the deflated gains process

$$G_t = \zeta_t S_t + \int_0^t \zeta_s D_s \, \mathrm{d}s \tag{A.1}$$

is a non-negative local martingale.

If S_t is of the form in (1.13), then we have

$$G_t = \mathbb{E}_t \left[\int_0^\infty \zeta_s D_s \, \mathrm{d}s \right] + L_t,$$

which is clearly a non-negative local martingale and therefore the market is arbitrage free.

Conversely, suppose that the market is arbitrage free and hence (A.1) holds. As a direct consequence, the process

$$\zeta_t S_t - \zeta_t S_t^* = G_t - \mathbb{E}_t \left[\int_0^\infty \zeta_s D_s \right]$$

must be a local martingale. To show nonnegativity, note that a local martingale bounded from below is a supermartingale, so that we have for all $T \geq t$

$$\begin{aligned} \zeta_t S_t - \zeta_t S_t^* &\geq \mathbb{E}_t \left[G_T - \int_0^\infty \zeta_s D_s \right] \\ &= \mathbb{E}_t \left[\zeta_T S_T - \int_T^\infty \zeta_s D_s \, \mathrm{d}s \right] \\ &\geq \mathbb{E}_t \left[- \int_T^\infty \zeta_s D_s \, \mathrm{d}s \right] \xrightarrow{T \to \infty} 0, \end{aligned}$$

where we have used the limited liability of the stock in the last inequality.

Proof of Proposition 1.2.6

Similarly as in the proof of Proposition 1.2.3 we get

$$\begin{aligned} \int_t^\infty (s-t) \, \mathbb{E}_t[\zeta_s D_s] \, \mathrm{d}s &= \upsilon^\top \int_t^\infty (s-t) \, \mathrm{e}^{(\beta-\gamma)s} \mathrm{e}^{G_2(s-t)} \, \mathrm{d}s H_2(X_t) \\ &= \mathrm{e}^{(\beta-\gamma)t} \upsilon^\top \int_t^\infty (s-t) \, \mathrm{e}^{[G_2-(\gamma-\beta)\mathrm{Id}](s-t)} \, \mathrm{d}s H_2(X_t). \end{aligned}$$

Applying integration by parts gives

$$\begin{aligned}\int_t^\infty (s-t)\,\mathbb{E}_t[\zeta_s D_s]\,\mathrm{d}s &= \mathrm{e}^{(\beta-\gamma)t}\,v^\top[(\gamma-\beta)\mathrm{Id}-G_2]^{-1}\int_t^\infty \mathrm{e}^{[G_2-(\gamma-\beta)\mathrm{Id}](s-t)}\,\mathrm{d}s H_2(X_t)\\ &= \mathrm{e}^{(\beta-\gamma)t}\,v^\top[(\gamma-\beta)\mathrm{Id}-G_2]^{-2}H_2(X_t)\\ &= \mathrm{e}^{(\beta-\gamma)t}\,w^\top[(\gamma-\beta)\mathrm{Id}-G_2]^{-1}H_2(X_t).\end{aligned}$$

The result now follows from (1.12) and (1.15).

Proof of Proposition 1.4.1

We start by showing that there exists a unique strong solution X_t to (1.21) with values in $(0,\infty)^d$. Due to the global Lipschitz continuity of the coefficients, the SDE in (1.21) has a unique strong solution in $\mathbb{R}^d$ for every $X_0 \in \mathbb{R}^d$, see Theorem III.2.32 in Jacod and Shiryaev (2003). It remains to show that X_t is $(0,\infty)^d$-valued for all $t \geq 0$ if $X_0 \in (0,\infty)^d$. First, we prove the statement for the diffusive case.

Lemma A.2.1. *Consider the SDE*

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \mathrm{diag}(X_t)\Sigma\,\mathrm{d}W_t, \tag{A.2}$$

for some d-dimensional Brownian motion W_t and κ, θ, Σ as assumed in Proposition 1.4.1. If $X_0 \in (0,\infty)^d$, then $X_t \in (0,\infty)^d$ for all $t \geq 0$.

Proof. Replace X_t in the drift of (A.2) by X_t^+ componentwise and consider the SDE

$$\mathrm{d}Y_t = \kappa(\theta - Y_t^+)\,\mathrm{d}t + \mathrm{diag}(Y_t)\Sigma\,\mathrm{d}W_t, \tag{A.3}$$

with $Y_0 = X_0 \in (0,\infty)^d$. The function $y \mapsto y^+$ componentwise is still Lipschitz continuous, so that there exists a unique solution Y_t to (A.3). Now consider the SDE

$$\mathrm{d}Z_t = -\mathrm{diag}(\mathrm{diag}(\kappa))Z_t^+\,\mathrm{d}t + \mathrm{diag}(Z_t)\Sigma\,\mathrm{d}W_t, \tag{A.4}$$

with $Z_0 = X_0 \in (0,\infty)^d$. Its unique solution is the $(0,\infty)^d$-valued process given by

$$Z_t = Z_0 \exp\left\{\left(-\mathrm{diag}(\kappa) - \frac{1}{2}\mathrm{diag}(\Sigma\Sigma^\top)\right)t + \Sigma W_t\right\}.$$

By assumption, we have that the drift function of (A.3) is always greater than or equal to the drift function of (A.4):

$$\kappa\theta - \kappa x^+ \geq -\mathrm{diag}(\mathrm{diag}(\kappa))x^+, \quad \forall x \in \mathbb{R}^d.$$

By the comparison theorem from (Geiß and Manthey, 1994, Theorem 1.2) we have almost surely

$$Y_t \geq Z_t, \quad t \geq 0.$$

Hence, $Y_t \in (0,\infty)^d$ and therefore Y_t also solves the SDE (A.2). By uniqueness we conclude that $X_t = Y_t$ for all t, which proves the claim. □

Define τ_i as the ith jump time of N_t and $\tau_0 = 0$. We argue by induction and assume that $X_{\tau_i} > 0$ for some $i = 0, 1, \ldots$. Since the process X_t is right-continuous, we have the following diffusive dynamics for the process $X_t^{(\tau_i)} = X_{t+\tau_i}$ on the interval $[0, \tau_{i+1} - \tau_i)$

$$\mathrm{d}X_t^{(\tau_i)} = \left(\kappa\theta + \left(-\kappa - \xi\,\mathrm{diag}\left(\int_{\mathscr{S}} z\,\mathrm{d}F(\mathrm{d}z)\right)\right) X_t^{(\tau_i)}\right)\mathrm{d}t + \mathrm{diag}(X_t^{(\tau_i)})\Sigma\,\mathrm{d}B_t^{(\tau_i)},$$

with $X_0^{(\tau_i)} = X_{\tau_i}$ and $B_t^{(\tau_i)} = B_{\tau_i+t} - B_{\tau_i}$. The stopping time τ_i is a.s. finite and therefore the process $B_t^{(\tau_i)}$ defines a d-dimensional Brownian motion with respect to its natural filtration, see Theorem 6.16 in Karatzas and Shreve (1991). By Lemma A.2.1 we have $X_t^{(\tau_i)} \in (0,\infty)^d$ for all $t \in [0, \tau_{i+1} - \tau_i)$. As a consequence, we have $X_t \in (0,\infty)^d$ for all $t \in [\tau_i, \tau_{i+1})$. The jump size $X_{\tau_{i+1}} - X_{\tau_{i+1}-}$ at time τ_{i+1} satisfies

$$X_{\tau_{i+1}} - X_{\tau_{i+1}-} = \mathrm{diag}(X_{\tau_{i+1}-})Z_{i+1} > -X_{\tau_{i+1}-},$$

where the Z_{i+1} are i.i.d. random variables with distribution $F(\mathrm{d}z)$. Rearranging terms gives $X_{\tau_{i+1}} \in (0,\infty)^d$. By induction we conclude that $X_t \in (0,\infty)^d$ for $t \in [0, \tau_i)$, $i \in \mathbb{N}$. The claim now follows because $\tau_i \to \infty$ for $i \to \infty$ a.s.

Next, we prove that X_t is a polynomial jump-diffusion. The action of the generator of X_t on a C^2 function $f\colon \mathbb{R}^d \to \mathbb{R}$ is given by

$$\begin{aligned}\mathscr{G}f(x) =& \frac{1}{2}\mathrm{tr}\left(\mathrm{diag}(x)\Sigma\Sigma^\top\mathrm{diag}(x)\nabla^2 f(x)\right) + \nabla f(x)^\top\kappa(\theta - x) \\ &+ \xi\left(\int_{\mathscr{S}} f(x + \mathrm{diag}(x)z)\,F(\mathrm{d}z) - f(x) - \nabla f(x)^\top\mathrm{diag}(x)\int_{\mathscr{S}} z\,F(\mathrm{d}z)\right),\end{aligned} \tag{A.5}$$

where $\mathscr{S}$ denotes the support of F and we assume that f is such that the integrals are finite. Now suppose that $f \in \mathrm{Pol}_n(\mathbb{R}^d)$ and assume without loss of generality that f is a monomial with $f(x) = x^\alpha = x_1^{\alpha_1}\cdots x_d^{\alpha_d}$, $|\alpha| = n$. We now apply the generator to this function. It follows immediately that the first two terms in (A.5) are again a polynomial of degree n or less. Indeed, the gradient (hessian) in the second (first) term lowers the degree by one (two), while the remaining factors augment the degree by at most one (two). The third term in (A.5) becomes (we slightly abuse the notation α to represent both a multi-index and a vector):

$$\begin{aligned}&\xi\left(x^\alpha\int_{\mathscr{S}}\prod_{j=1}^d(1+z_j)^{\alpha_j}\,F(\mathrm{d}z) - x^\alpha - x^\alpha\alpha^\top\int_{\mathscr{S}} z\,F(\mathrm{d}z)\right) \\ =&\xi x^\alpha\int_{\mathscr{S}}\left(\mathrm{e}^{\alpha^\top\log(1+z)} - 1 - \alpha^\top z\right)F(\mathrm{d}z),\end{aligned} \tag{A.6}$$

where the logarithm is applied componentwise. Hence, we conclude that $\mathscr{G}$ maps polynomials to polynomials of the same degree or less.

Proof of Proposition 1.4.2

This proof is similar to the one of Theorem 5 in Filipović et al. (2017). From (1.5) we have that $D_t \geq 0$ if and only if

$$\beta \geq \sup_{x\in(0,\infty)^d} -\frac{p^\top G_1 H_1(x)}{p^\top H_1(x)}, \tag{A.7}$$

provided it is finite. Using (1.2) we have

$$-\frac{p^\top G_1 H_1(x)}{p^\top H_1(x)} = \frac{-\tilde{p}^\top \kappa\theta + \sum_{j=1}^d \tilde{p}^\top \kappa_j x_j}{p_0 + \sum_{j=1}^k p_j x_j}. \tag{A.8}$$

Using the assumption $\kappa_{ij} \leq 0$ for $i \neq j$ (cfr., Proposition 1.4.1), we have for all $j > k$ that

$$\tilde{p}^\top \kappa_j = \sum_{i=1}^d p_i \kappa_{ij} = \sum_{i=1}^k p_i \kappa_{ij} \leq 0. \tag{A.9}$$

Combining (A.8) with (A.9) gives

$$\sup_{x\in(0,\infty)^d} \frac{-\tilde{p}^\top \kappa\theta + \sum_{j=1}^d \tilde{p}^\top \kappa_j x_j}{p_0 + \sum_{j=1}^k p_j x_j} = \sup_{x\in(0,\infty)^k} \frac{-\tilde{p}^\top \kappa\theta + \sum_{j=1}^k \tilde{p}^\top \kappa_j x_j}{p_0 + \sum_{j=1}^k p_j x_j}. \tag{A.10}$$

If $p_0 > 0$, then the fraction on the right-hand side of (A.10) can be seen as a convex combination of

$$\left\{-\frac{\tilde{p}^\top \kappa\theta}{p_0}, \frac{\tilde{p}^\top \kappa_1}{p_1}, \ldots, \frac{\tilde{p}^\top \kappa_k}{p_k}\right\},$$

with coefficients $p_0, p_1 x_1, \ldots, p_k x_k$. As a consequence, we have in this case

$$\sup_{x\in(0,\infty)^d} -\frac{p^\top G_1 H_1(x)}{p^\top H_1(x)} = \max\left\{-\frac{\tilde{p}^\top \kappa\theta}{p_0}, \frac{\tilde{p}^\top \kappa_1}{p_1}, \ldots, \frac{\tilde{p}^\top \kappa_k}{p_k}\right\}.$$

If $p_0 = 0$, then using the assumption $\kappa\theta \geq 0$ (cfr., Proposition 1.4.1) we get

$$\begin{aligned}
\sup_{x\in(0,\infty)^d} -\frac{p^\top G_1 H_1(x)}{p^\top H_1(x)} &= \sup_{x\in(0,\infty)^k} \frac{-\tilde{p}^\top \kappa\theta + \sum_{j=1}^k \tilde{p}^\top \kappa_j x_j}{\sum_{j=1}^k p_j x_j} \\
&= \sup_{x\in(0,\infty)^k} \frac{\sum_{j=1}^k \tilde{p}^\top \kappa_j x_j}{\sum_{j=1}^k p_j x_j} \\
&= \max\left\{\frac{\tilde{p}^\top \kappa_1}{p_1}, \ldots, \frac{\tilde{p}^\top \kappa_k}{p_k}\right\}.
\end{aligned}$$

Proof of Proposition 1.4.3

Suppose first that κ is lower triangular. In order to get a specific idea what the matrix G_2 looks like, we start by fixing a monomial basis for $\mathrm{Pol}_2(\mathbb{R}^d)$ using the graded lexicographic ordering

of monomials:

$$H_2(x) = (1, x_1, \dots, x_d, x_1^2, x_1x_2, \dots, x_1x_d, x_2^2, x_2x_3, \dots, x_d^2)^\top, \quad x \in \mathbb{R}^d. \tag{A.11}$$

It follows by inspection of (A.5) and (A.6) that, thanks to the triangular structure of κ, the matrix G_2 is lower triangular with respect to this basis. Indeed, the first and third term in (A.5) only contribute to the diagonal elements of G_2, while the second term contributes to the lower triangular part (including the diagonal). The eigenvalues of G_2 are therefore given by its diagonal elements.

Each element in the monomial basis can be expressed as as $f(x) = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, for some $\alpha \in \mathbb{N}^d$ with $\sum_{i=1}^d \alpha_i \le 2$. In order to find the diagonal elements of G_2, we need to find the coefficient of the polynomial $\mathcal{G}f(x)$ associated with the basis element $f(x)$. It follows from (A.5) and (A.6) that this coefficient is given by

$$\begin{aligned} &-\sum_{i=1}^d \kappa_{ii}\alpha_i + \frac{1}{2}\sum_{i<j}(\Sigma\Sigma^\top)_{ij}\alpha_i\alpha_j + \sum_{i=1}^d (\Sigma\Sigma^\top)_{ii}\alpha_i(\alpha_i - 1) \\ &+ \xi\int_{\mathscr{S}} \left(\mathrm{e}^{\alpha^\top \log(1+z)} - 1 - \alpha^\top z\right) F(\mathrm{d}z). \end{aligned}$$

The restriction $\sum_{i=1}^d \alpha_i \le 2$ allows to summarize all diagonal elements, and hence the eigenvalues, of G_2 as follows

$$\begin{gathered} 0, -\kappa_{11}, \dots, -\kappa_{dd}, \\ -\kappa_{ii} - \kappa_{jj} + (\Sigma\Sigma^\top)_{ij} + \xi\int_{\mathscr{S}} z_i z_j F(\mathrm{d}z), \quad 1 \le i, j \le d. \end{gathered}$$

Note that a change of basis will lead to a different matrix G_2, however its eigenvalues are unaffected. The choice of the basis in (A.11) is therefore without loss of generality.

If κ is upper triangular, we consider a different ordering for the monomial basis:

$$H_2(x) = (1, x_d, \dots, x_1, x_d^2, x_dx_{d-1}, \dots, x_dx_1, x_{d-1}^2, x_{d-1}x_{d-2}, \dots, x_1^2)^\top, \quad x \in \mathbb{R}^d.$$

The result now follows from the same arguments as in the lower triangular case.

Proof of Proposition 1.6.1

Using the law of iterated expectations and the moment formula (1.3) we get:

$$\begin{aligned} \mathbb{E}_t[\zeta_{T_2}(C_{T_2} - C_{T_1})] &= \mathrm{e}^{-\gamma T_2}\left(\mathrm{e}^{\beta T_2}\mathbb{E}_t[q^\top H_1(X_{T_2})p^\top H_1(X_{T_2})] - \mathrm{e}^{\beta T_1}\mathbb{E}_t[p^\top H_1(X_{T_1})\mathbb{E}_{T_1}[q^\top H_1(X_{T_2})]]\right) \\ &= \mathrm{e}^{-\gamma T_2}\left(\mathrm{e}^{\beta T_2} w_2^\top \mathrm{e}^{G_2(T_2-t)}H_2(X_t) - \mathrm{e}^{\beta T_1}\mathbb{E}_t[p^\top H_1(X_{T_1})q^\top \mathrm{e}^{G_1(T_2-T_1)}H_1(X_{T_1})]\right) \\ &= \mathrm{e}^{-\gamma T_2}\left(\mathrm{e}^{\beta T_2} w_2^\top \mathrm{e}^{G_2(T_2-t)}H_2(X_t) - \mathrm{e}^{\beta T_1} w_1^\top \mathrm{e}^{G_1(T_1-t)}H_2(X_t)\right). \end{aligned}$$

Note that the vectors w_1 and w_2 are unique since the basis elements are linearly independent by definition. Finally, using the bond price formula (1.9) we get

$$\begin{aligned} D_{fwd}(t,T_1,T_2) &= \frac{1}{\zeta_t P(t,T_2)} \mathbb{E}_t[\zeta_{T_2}(C_{T_2} - C_{T_1})] \\ &= \frac{\mathrm{e}^{\beta T_2} w_2^\top \mathrm{e}^{G_2(T_2-t)} H_2(X_t) - \mathrm{e}^{\beta T_1} w_1^\top \mathrm{e}^{G_1(T_1-t)} H_2(X_t)}{q^\top \mathrm{e}^{G_1(T_2-t)} H_1(X_t)}. \end{aligned}$$

B Appendix to Chapter 2

B.1 Scaling with Auxiliary Moments

In this appendix we show how to avoid rounding errors by scaling the problem using the moments of the auxiliary density w.

Using $(L^{-1})^\top L^{-1} = M^{-1}$ we get from (2.6):

$$\begin{aligned}\pi^{(N)} &= (f_0,\dots,f_N)(\ell_0,\dots,\ell_N)^\top \\ &= \mathrm{e}^{-rT}(\tilde{f}_0,\dots,\tilde{f}_N)M^{-1}\mathrm{e}^{G_N T}H_N(0).\end{aligned}$$

Define $S \in \mathbb{R}^{(N+1)\times(N+1)}$ as the diagonal matrix with the moments of w on its diagonal:

$$S = \mathrm{diag}(s_0,\dots,s_N), \quad s_i = \mathrm{e}^{i\mu+\frac{1}{2}i^2\nu^2}.$$

We can now write

$$\begin{aligned}\pi^{(N)} &= \mathrm{e}^{-rT}(\overline{f}_0,\dots,\overline{f}_N)S^{-1}SM^{-1}SS^{-1}\mathrm{e}^{G_N T}SS^{-1}H_N(0) \\ &= \mathrm{e}^{-rT}(\overline{f}_0,\dots,\overline{f}_N)\overline{M}^{-1}\mathrm{e}^{\overline{G}_N T}H_N(0),\end{aligned} \tag{B.1}$$

where we have defined the matrices $((\overline{G}_N)_{ij})_{0\le i,j\le N}$, $(\overline{M}_{ij})_{0\le i,j\le N}$ and the vector $(\overline{f}_0,\dots,\overline{f}_N)^\top$ as

$$\overline{M}_{ij} = \mathrm{e}^{ijv^2}, \quad \overline{f}_i = \mathrm{e}^{\mu+\frac{1}{2}(2i+1)v^2}\Phi(d_{i+1}) - K\Phi(d_i), \quad (\overline{G}_N)_{ij} = \begin{cases} ir + \frac{1}{2}i(i-1)\sigma^2 & j = i \\ \frac{i}{T}\mathrm{e}^{-\mu+\frac{1}{2}(1-2i)v^2} & j = i-1 \\ 0 & else \end{cases},$$

for $i,j = 0,\dots,N$. The components of $\overline{M}$ and $(\overline{f}_0,\dots,\overline{f}_N)^\top$ grow *much* slower for increasing N as their counterparts M and $(\tilde{f}_0,\dots,\tilde{f}_N)^\top$, respectively. The vector $\mathrm{e}^{\overline{G}_N T}H_N(0)$ corresponds to the moments of A_T divided by the moments corresponding to w. Since both moments grow approximately at the same rate, this vector will have components around one. This scaling is

important since for large N the raw moments of A_T are enormous. This was causing trouble for example in Dufresne (2000). We therefore circumvent the numerical inaccuracies coming from the explosive moment behavior by directly computing the relative moments.

The likelihood and payoff coefficients can be computed by performing a Cholesky decomposition on $\overline{M}$ instead of M:

$$(f_0,\dots,f_N)^\top = \mathrm{e}^{-rT}\overline{L}^{-1}(\overline{f}_0,\dots,\overline{f}_N)^\top,$$
$$(\ell_0,\dots,\ell_N)^\top = \overline{L}^{-1}\mathrm{e}^{\overline{G}_N T}H_N(0),$$

where $\overline{M} = \overline{L}\,\overline{L}^\top$ is the Cholesky decomposition of $\overline{M}$.

Remark that to compute the option price in (B.1), we do not necessarily need to do a Cholesky decomposition. Indeed, we only need to invert the matrix $\overline{M}$. Doing a Cholesky decomposition is one way to solve a linear system, but there are other possibilities. In particular, remark that $\overline{M}$ is a Vandermonde matrix and its inverse can be computed analytically (see e.g. Turner (1966)). There also exist specific numerical methods to solve linear Vandermonde systems, see e.g. Björck and Pereyra (1970). However, we have not found any significant differences between the Cholesky method and alternative matrix inversion techniques for the examples considered in this chapter.

For very large values of ν, even the matrix $\overline{M}$ might become ill conditioned. In this case it is advisable to construct the orthonormal basis using the three-term recurrence relation:

Lemma B.1.1. *The polynomials $b_n \in \mathrm{Pol}_n(\mathbb{R})$, $n = 0,1,\dots$, defined recursively by*

$$b_0(x) = 1, \quad b_1(x) = \frac{1}{\beta_1}(x-\alpha_0),$$
$$b_n(x) = \frac{1}{\beta_n}\big((x-\alpha_{n-1})b_{n-1}(x) - \beta_{n-1}b_{n-2}(x)\big), \quad n = 2,3,\dots,$$

with

$$\alpha_n = \mathrm{e}^{\mu+\nu^2(n-\frac{1}{2})}\left(\mathrm{e}^{\nu^2(n+1)} + \mathrm{e}^{\nu^2 n} - 1\right), \quad n = 0,1,\dots$$
$$\beta_n = \mathrm{e}^{\mu+\frac{1}{2}\nu^2(3n-2)}\sqrt{\mathrm{e}^{\nu^2 n}-1}, \quad n = 1,2,\dots,$$

satisfy

$$\int_{\mathbb{R}} b_i(x)b_j(x)w(x)\,\mathrm{d}x = \begin{cases} 1 & i = j \\ 0 & else \end{cases}, \quad i,j = 0,1,\dots.$$

Proof. Straightforward application of the moment-generating function of the normal distribution and Theorem 1.29 in Gautschi (2004). □

The above recursion suffers from rounding errors in double precision arithmetic for small ν, but is very accurate for large ν.

B.2 Control Variate for Simulating $g(x)$

In order to increase the efficiency of the Monte-Carlo simulation of $g(x)$, we describe in this appendix how to use the density of the geometric average as a control variate. This idea is inspired by Kemna and Vorst (1990), who report a very substantial variance reduction when using the geometric Asian option price as a control variate in the simulation of the arithmetic Asian option price. Denote by $Q_T = \exp\left(\frac{1}{T}\int_0^T \log(S_s)\,\mathrm{d}s\right)$ the geometrical price average. It is not difficult to see that $\log(Q_T)$ is normally distributed with mean $\frac{1}{2}(r-\frac{1}{2}\sigma^2)T$ and variance $\frac{\sigma^2}{3}T$. Hence, Q_T is log-normally distributed and we know its density function, which we denote by $q(x)$, explicitly. Similarly as in Lemma 2.4.1, we can also express $q(x)$ as an expectation:

Lemma B.2.1. *For any $x \in \mathbb{R}$*

$$q(x) = \mathbb{E}\left[\left(1_{\{Q_T \geq x\}} - c(x)\right)\left(\frac{2B_T}{\sigma T Q_T} + \frac{1}{Q_T}\right)\right], \tag{B.2}$$

where c is any deterministic finite-valued function.

We now propose the following estimator for the density of the arithmetic average:

$$\begin{aligned} g(x) =& \mathbb{E}\left[(1_{\{A_T \geq x\}} - c_1(x))\frac{2}{\sigma^2}\left(\frac{S_T - S_0}{TA_T^2} + \frac{\sigma^2 - r}{A_T}\right)\right. \\ &\left. + \left(q(x) - \left(1_{\{Q_T \geq x\}} - c_2(x)\right)\left(\frac{2B_T}{\sigma T Q_T} + \frac{1}{Q_T}\right)\right)\right], \end{aligned} \tag{B.3}$$

for some deterministic finite-valued functions c_1 and c_2. Given the typically high correlation between the geometric and arithmetic average, the above estimator has a significantly smaller variance than the estimator in (2.9). In numerical examples the functions c_1 and c_2 are chosen as follows:

$$c_1(x) = 1_{x \leq m_1^A}, \quad c_2(x) = 1_{x \leq m_1^Q},$$

where m_1^A and m_1^Q denote the first moments of A_T and Q_T, respectively.

Finally, we use (B.3) to express $\|\ell\|_w^2$ as an expectation that can be evaluated using Monte-Carlo

simulation:

$$\|\ell\|_w^2 = \mathbb{E}\left[\frac{(1_{\{A_T\geq\tilde{A}_T\}} - c_1(\tilde{A}_T))\frac{2}{\sigma^2}\left(\frac{S_T-S_0}{TA_T^2}+\frac{\sigma^2-r}{A_T}\right)}{w(\tilde{A}_T)}\right.$$
$$\left. + \frac{q(\tilde{A}_T) - \left(1_{\{Q_T\geq\tilde{A}_T\}} - c_2(\tilde{A}_T)\right)\left(\frac{2B_T}{\sigma TQ_T}+\frac{1}{Q_T}\right)}{w(\tilde{A}_T)}\right], \tag{B.4}$$

where the random variable $\tilde{A}_T$ is independent from all other random variables and has the same distribution as A_T. In numerical examples we find a variance reduction of roughly a factor ten.

B.3 Proofs

This appendix contains all the proofs.

Proof of Lemma 2.2.1

Using the time-reversal property of a Brownian motion, we have the following identity in law for fixed $t > 0$:

$$\begin{aligned} tA_t &= \int_0^t \mathrm{e}^{(r-\frac{1}{2}\sigma^2)u+\sigma B_u}\,\mathrm{d}u \\ &\overset{\text{law}}{=} \int_0^t \mathrm{e}^{(r-\frac{1}{2}\sigma^2)(t-u)+\sigma(B_t-B_u)}\,\mathrm{d}u \\ &= S_t\int_0^t S_u^{-1}\,\mathrm{d}u. \end{aligned}$$

Applying Itô's lemma to $X_t := S_t\int_0^t S_u^{-1}\,\mathrm{d}u$ gives

$$\begin{aligned} \mathrm{d}X_t &= S_tS_t^{-1}\,\mathrm{d}t + \int_0^t S_u^{-1}\,\mathrm{d}u\,(rS_t\,\mathrm{d}t+\sigma S_t\,\mathrm{d}B_t) \\ &= (rX_t+1)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t. \end{aligned}$$

Proof of Proposition 2.2.2

Applying the infinitesimal generator $\mathscr{G}$ corresponding to the diffusion in (2.1) to a monomial x^n gives:

$$\mathscr{G}x^n = x^n(nr+\frac{1}{2}n(n-1)\sigma^2) + nx^{n-1}.$$

Hence, we have $\mathcal{G}H_n(x) = \tilde{G}_n H_n(x)$ componentwise, where $\tilde{G}_n$ is defined as

$$\tilde{G}_n = \begin{pmatrix} 0 & & & \\ 1 & r & & \\ & \ddots & \ddots & \\ & & n & (nr + \frac{1}{2}n(n-1)\sigma^2) \end{pmatrix}.$$

Using the identity in distribution of Lemma 2.2.1, we get

$$\begin{aligned}
\mathbb{E}[H_n(A_T)] &= \operatorname{diag}(H_n(T^{-1}))\mathbb{E}[H_n(X_T)] \\
&= \operatorname{diag}(H_n(T^{-1}))\left(H_n(X_0) + \int_0^T \mathbb{E}[\mathcal{G}H_n(X_u)]\,\mathrm{d}u\right) \\
&= \operatorname{diag}(H_n(T^{-1}))H_n(0) + \operatorname{diag}(H_n(T^{-1}))\tilde{G}_n \int_0^T \mathbb{E}[H_n(X_u)]\,\mathrm{d}u \\
&= H_n(0) + \operatorname{diag}(H_n(T^{-1}))\tilde{G}_n \operatorname{diag}(H_n(T)) \int_0^T \mathbb{E}[H_n(A_u)]\,\mathrm{d}u \\
&= H_n(0) + G_n \int_0^T \mathbb{E}[H_n(A_u)]\,\mathrm{d}u,
\end{aligned}$$

where G_n was defined in (2.2). The result now follows from solving the above linear first order ODE.

Proof of Proposition 2.2.3

1. We will show that the solution at time $T > 0$ of the SDE in (2.1) admits a smooth density function. The claim then follows by the identity in law.

   Define the volatility and drift functions $a(x) = \sigma x$ and $b(x) = rx + 1$. Hörmander's condition (see for example Section 2.3.2 in Nualart (2006)) becomes in this case:

   $$a(X_0) \neq 0 \quad \text{or} \quad a^{(n)}(X_0)b(X_0) \neq 0 \text{ for some } n \geq 1.$$

   Hörmander's condition is satisfied since for $n = 1$ we have $a'(0)b(0) = \sigma \neq 0$. Since $a(x)$ and $b(x)$ are infinitely differentiable functions with bounded partial derivatives of all orders, we conclude by Theorem 2.3.3 in Nualart (2006) that X_T, and therefore A_T, admits a smooth density function.

2. We start from the following two observations:

   $$A_T \leq \sup_{0\leq u\leq T} S_u \quad \text{and} \quad P\left(\sup_{0\leq u\leq T} \sigma B_u \geq x\right) = 2P\left(Z \geq \frac{x}{\sigma\sqrt{T}}\right),$$

where $Z \sim N(0,1)$. Using the fact that the exponential is an increasing function, we get

$$\begin{aligned} P(A_T \geq x) &\leq P\left(\sup_{0\leq u\leq T} S_u \geq x\right) \\ &= P\left(\sup_{0\leq u\leq T} (r-\frac{1}{2}\sigma^2)u + \sigma B_u \geq \log(x)\right) \\ &\leq \begin{cases} 2P\left(Z \geq \frac{\log(x)-(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) & \text{if} \quad r \geq \frac{1}{2}\sigma^2 \\ 2P\left(Z \geq \frac{\log(x)}{\sigma\sqrt{T}}\right) & \text{if} \quad r \leq \frac{1}{2}\sigma^2 \end{cases}. \end{aligned}$$

Applying the rule of l'Hôpital gives

$$\begin{aligned} \lim_{x\to\infty} g(x)\left(\mathrm{e}^{-\frac{\log(x)^2}{2\sigma^2 T}}\right)^{-1} &= \lim_{x\to\infty} P(A_T \geq x)\left(\int_x^\infty \mathrm{e}^{-\frac{\log(y)^2}{2\sigma^2 T}}\,\mathrm{d}y\right)^{-1} \\ &\leq 2\frac{1}{\sqrt{2\pi T}\sigma}\lim_{x\to\infty}\int_x^\infty \mathrm{e}^{-\frac{(\log(y)-(r-\frac{1}{2}\sigma^2)+T)^2}{2\sigma^2 T}}\,\mathrm{d}y\left(\int_x^\infty \mathrm{e}^{-\frac{\log(y)^2}{2\sigma^2 T}}\,\mathrm{d}y\right)^{-1} \\ &= \sqrt{\frac{2}{\pi T}}\frac{1}{\sigma}. \end{aligned}$$

Hence we have show that $g(x) = \mathcal{O}\left(\exp\left\{-\frac{1}{2}\frac{\log(x)^2}{\sigma^2 T}\right\}\right)$ for $x \to \infty$.

Since the exponential is a convex function, we have that the arithmetic average is always bounded below by the geometric average:

$$A_T \geq Q_T = \exp\left(\frac{1}{T}\int_0^T \log(S_s)\,\mathrm{d}s\right).$$

It is not difficult to see that $\log(Q_T)$ is normally distributed with mean $\frac{1}{2}(r-\frac{1}{2}\sigma^2)T$ and variance $\frac{\sigma^2}{3}T$. By similar arguments as before we therefore have

$$g(x) = \mathcal{O}\left(\exp\left\{-\frac{3}{2}\frac{\log(x)^2}{\sigma^2 T}\right\}\right) \quad \text{for} \quad x \to 0.$$

Proof of Proposition 2.3.1

We can write the squared norm of ℓ as

$$\begin{aligned} \|\ell\|_w^2 &= \int_0^\infty \left(\frac{g(x)}{w(x)}\right)^2 w(x)\,\mathrm{d}x \\ &= \int_0^a \frac{g(x)^2}{w(x)}\,\mathrm{d}x + \int_a^b \frac{g(x)^2}{w(x)}\,\mathrm{d}x + \int_b^\infty \frac{g(x)^2}{w(x)}\,\mathrm{d}x, \end{aligned} \tag{B.5}$$

for some $0 < a < b < \infty$. The second term is finite since the function $\frac{g^2}{w}$ is continuous over the compact interval $[a, b]$. From Proposition 2.2.3 we have

$$g(x)^2 = \mathcal{O}\left(\exp\left\{-\frac{\log(x)^2}{\sigma^2 T}\right\}\right), \quad \text{for} \quad x \to \infty \quad \text{and} \quad x \to 0.$$

For the log-normal density we have

$$w(x) = \mathcal{O}\left(\exp\left\{-\frac{\log(x)^2}{2\nu}\right\}\right), \quad \text{for} \quad x \to \infty \quad \text{and} \quad x \to 0.$$

Since $2\nu > \sigma^2 T$ by assumption, we are guaranteed that the first and last term in (B.5) are finite for a sufficiently small (resp. large) choice of a (resp. b).

Proof of Proposition 2.3.3

The payoff coefficients can be written as

$$(f_0, \ldots, f_N)^\top = e^{-rT} C (\tilde{f}_0, \ldots, \tilde{f}_N)^\top,$$

with

$$\begin{aligned}\tilde{f}_n &= \frac{1}{\sqrt{2\pi}\nu} \int_0^\infty (e^x - K)^+ \mathrm{e}^{nx} \mathrm{e}^{-\frac{(x-\mu)^2}{2\nu^2}} \,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\nu} \left(\int_{\log(K)}^\infty \mathrm{e}^{(n+1)x} \mathrm{e}^{-\frac{(x-\mu)^2}{2\nu^2}} \,\mathrm{d}x - K \int_{\log(K)}^\infty \mathrm{e}^{nx} \mathrm{e}^{-\frac{(x-\mu)^2}{2\nu^2}} \,\mathrm{d}x \right).\end{aligned}$$

Completing the square in the exponent gives

$$\begin{aligned}\frac{1}{\sqrt{2\pi}\nu} \int_{\log(K)}^\infty \mathrm{e}^{nx} \mathrm{e}^{-\frac{(x-\mu)^2}{2\nu^2}} \,\mathrm{d}x &= \frac{1}{\sqrt{2\pi}\nu} \mathrm{e}^{\mu n + \frac{1}{2} n^2 \nu^2} \int_{\log(K)}^\infty \mathrm{e}^{-\frac{(x-(\mu+\nu^2 n))^2}{2\nu^2}} \,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mu n + \frac{1}{2} n^2 \nu^2} \int_{\frac{\log(K)-(\mu+\nu^2 n)}{\nu}}^\infty \mathrm{e}^{-\frac{1}{2} y^2} \,\mathrm{d}y \\ &= \mathrm{e}^{\mu n + \frac{1}{2} n^2 \nu^2} \Phi(d_n),\end{aligned}$$

where d_n is defined in (2.7). We finally get

$$\tilde{f}_n = \mathrm{e}^{\mu(n+1) + \frac{1}{2}(n+1)^2 \nu^2} \Phi(d_{n+1}) - K \mathrm{e}^{\mu n + \frac{1}{2} n^2 \nu^2} \Phi(d_n).$$

Proof of Lemma 2.4.1

This proof is based on Malliavin calculus techniques, we refer to Nualart (2006) for an overview of standard results in this area. A similar approach is taken by Fournié et al. (1999) to compute the Greeks of an Asian option by Monte-Carlo simulation.

Denote by $D : \mathbb{D}^{1,2} \to L^2(\Omega \times [0,T])$, $F \mapsto \{D_t F, t \in [0,T]\}$, the Malliavin derivative operator. By Theorem 2.2.1 in (Nualart, 2006) we have $S_t, A_t \in \mathbb{D}^{1,2}$ for $t \in (0,T]$ and

$$D_u S_t = \sigma S_t 1_{\{u \le t\}}, \quad D_u A_t = \frac{\sigma}{t} \int_u^t S_s \, \mathrm{d}s.$$

Denote by

$$\delta : Dom(\delta) \to L^2(\Omega), \{X_t, t \in [0,T]\} \mapsto \delta(X)$$

the Skorohod integral and by $Dom(\delta) \subseteq L^2(\Omega \times [0,T])$ the corresponding domain. The Skorohod integral is defined as the adjoint operator of the Malliavin derivative and can be shown to extend the Itô integral to non-adapted processes. In particular, we have immediately that $\{S_t, t \in [0,T]\} \in Dom(\delta)$ and

$$\delta(S) = \int_0^T S_s \, \mathrm{d}B_s. \tag{B.6}$$

For $\phi \in C_c^\infty$ we have $\phi(A_T) \in \mathbb{D}^{1,2}$ and

$$\int_0^T (D_u \phi(A_T)) S_u \, \mathrm{d}u = \phi'(A_T) \int_0^T (D_u A_T) S_u \, \mathrm{d}u.$$

Using the duality relationship between the Skorohod integral and the Malliavin derivative we get

$$\begin{aligned} \mathbb{E}[\phi'(A_T)] &= \mathbb{E}\left[\int_0^T (D_u \phi(A_T)) \frac{S_u}{\int_0^T (D_u A_T) S_u \, \mathrm{d}u} \, \mathrm{d}u \right] \\ &= \mathbb{E}\left[\phi(A_T) \delta\left(\frac{S}{\int_0^T (D_u A_T) S_u \, \mathrm{d}u} \right) \right]. \end{aligned} \tag{B.7}$$

By Lemma 1 in Bally (2003) (see also Proposition 2.1.1 in Nualart (2006) for a similar approach) we obtain the following representation of the density function of A_T:[1]

$$\begin{aligned} g(x) &= \mathbb{E}\left[1_{\{A_T \ge x\}} \delta\left(\frac{S}{\int_0^T (D_u A_T) S_u \, \mathrm{d}u} \right) \right] \\ &= \frac{T}{\sigma} \mathbb{E}\left[1_{\{A_T \ge x\}} \delta\left(\frac{S}{\int_0^T S_u \int_u^T S_s \, \mathrm{d}s \, \mathrm{d}u} \right) \right]. \end{aligned} \tag{B.8}$$

Interchanging the order of integration gives

$$\begin{aligned} \int_0^T S_u \int_u^T S_s \, \mathrm{d}s \, \mathrm{d}u &= \left(\int_0^T S_u \, \mathrm{d}u \right)^2 - \int_0^T \int_0^u S_u S_s \, \mathrm{d}s \, \mathrm{d}u \\ &= \left(\int_0^T S_u \, \mathrm{d}u \right)^2 - \int_0^T S_s \int_s^T S_u \, \mathrm{d}u \, \mathrm{d}s, \end{aligned}$$

[1] Informally speaking one applies a regularization argument in order to use (B.7) for the (shifted) Heaviside function $\phi(y) = 1_{\{y \ge x\}}$.

which gives $\int_0^T S_u \int_u^T S_s \,\mathrm{d}s\,\mathrm{d}u = \frac{T^2}{2} A_T^2$. Plugging this into (B.8) gives

$$g(x) = \frac{2}{T\sigma}\mathbb{E}\left[1_{\{A_T \geq x\}} \delta\left(\frac{S}{A_T^2}\right)\right].$$

We use Proposition 1.3.3 in Nualart (2006) to factor out the random variable A_T^{-2} from the Skorohod integral:

$$\begin{aligned}\delta\left(\frac{S}{A_T^2}\right) &= A_T^{-2}\delta(S) - \int_0^T D_t\left(A_T^{-2}\right) S_t \,\mathrm{d}t \\ &= A_T^{-2}\frac{1}{\sigma}\left(S_T - S_0 - r\int_0^T S_s \,\mathrm{d}s\right) - \int_0^T D_t\left(A_T^{-2}\right) S_t \,\mathrm{d}t,\end{aligned}$$

where we used (B.6) in the last equation. Using the chain rule for the Malliavin derivative we get

$$\begin{aligned}\delta\left(\frac{S}{A_T^2}\right) &= A_T^{-2}\frac{1}{\sigma}\left(S_T - S_0 - r\int_0^T S_s \,\mathrm{d}s\right) + 2A_T^{-3}\int_0^T D_t A_T S_t \,\mathrm{d}t \\ &= A_T^{-2}\frac{1}{\sigma}\left(S_T - S_0 - rTA_T\right) + 2A_T^{-3}\frac{1}{T}\int_0^T S_t \int_t^T \sigma S_u \,\mathrm{d}u\,\mathrm{d}t \\ &= A_T^{-2}\frac{1}{\sigma}\left(S_T - S_0 - rTA_T\right) + A_T^{-1}\sigma T \\ &= A_T^{-2}\frac{1}{\sigma}\left(S_T - S_0\right) + \frac{T}{\sigma}A_T^{-1}(\sigma^2 - r).\end{aligned}$$

Putting everything back together we finally get:

$$g(x) = \frac{2}{\sigma^2}\mathbb{E}\left[1_{\{A_T \geq x\}}\left(\frac{S_T - S_0}{TA_T^2} + \frac{\sigma^2 - r}{A_T}\right)\right].$$

Since the Skorohod integral has zero expectation we also have

$$g(x) = \frac{2}{\sigma^2}\mathbb{E}\left[\left(1_{\{A_T \geq x\}} - c(x)\right)\left(\frac{S_T - S_0}{TA_T^2} + \frac{\sigma^2 - r}{A_T}\right)\right],$$

for any deterministic finite-valued function c.

Proof of Corollary 2.4.3

The result follows immediately from (2.9) and

$$\|\ell\|_w^2 = \int_0^\infty \ell^2(x)\, w(x) \,\mathrm{d}x = \int_0^\infty \frac{g(x)}{w(x)} g(x) \,\mathrm{d}x.$$

Proof of Proposition 2.4.4

Using the Cauchy-Schwarz inequality and the orthonormality of the polynomials $b_0, \dots, b_N$ we get

$$\begin{aligned}
|\pi - \pi^{(N)}| &= \left| \langle F, l \rangle_w - \sum_{n=0}^{N} f_n \ell_n \right| \\
&= \left\langle F - \sum_{n=0}^{N} b_n f_n \,,\, \ell - \sum_{n=0}^{N} b_n \ell_n \right\rangle_w \\
&\leq \left\| F - \sum_{n=0}^{N} b_n f_n \right\|_w \left\| \ell - \sum_{n=0}^{N} b_n \ell_n \right\|_w \\
&= \left(\|F\|_w^2 - \sum_{n=0}^{N} f_n^2 \right)^{\frac{1}{2}} \left(\|\ell\|_w^2 - \sum_{n=0}^{N} \ell_n^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Proof of Lemma B.2.1

Applying the Malliavin derivative to Q_T gives

$$D_u Q_T = Q_T D_u \left(\frac{1}{T} \int_0^T \log(S_s)\, \mathrm{d}s \right) = Q_T \frac{\sigma}{T} (T-u) 1_{u \leq T}.$$

Similarly as in the proof of Lemma 2.4.1 we can write

$$\begin{aligned}
q(x) &= \mathbb{E}\left[1_{\{Q_T \geq x\}} \delta\left(\frac{1}{\int_0^T D_u Q_T \,\mathrm{d}u} \right) \right] \\
&= \mathbb{E}\left[1_{\{Q_T \geq x\}} \delta\left(\frac{2}{Q_T \sigma T} \right) \right]. \qquad \text{(B.9)}
\end{aligned}$$

Using Proposition 1.3.3 in Nualart (2006) to factor out the random variable from the Skorohod integral gives

$$\begin{aligned}
\delta\left(\frac{2}{Q_T \sigma T} \right) &= \frac{2B_T}{\sigma T Q_T} - \frac{2}{\sigma T} \int_0^T D_u (Q_T^{-1})\, \mathrm{d}u \\
&= \frac{2B_T}{\sigma T Q_T} - \frac{2}{\sigma T} Q_T^{-2} \int_0^T Q_T \frac{\sigma}{T} (T-u)\, \mathrm{d}u \\
&= \frac{2B_T}{\sigma T Q_T} + \frac{1}{Q_T}.
\end{aligned}$$

Plugging this back into (B.9) finally gives

$$q(x) = \mathbb{E}\left[1_{\{Q_T \geq x\}} \left(\frac{2B_T}{\sigma T Q_T} + \frac{1}{Q_T} \right) \right].$$

Since the Skorohod integral has zero expectation we also have

$$q(x) = \mathbb{E}\left[\left(1_{\{Q_T \geq x\}} - c(x)\right)\left(\frac{2B_T}{\sigma T Q_T} + \frac{1}{Q_T}\right)\right],$$

for any deterministic finite-valued function c.

C Appendix to Chapter 3

Proof of Lemma 3.3.1

Integration by parts gives

$$\begin{aligned} g(\tau) &= g(0) + \int_0^\tau g'(x)\,\mathrm{d}x \\ &= g(0) + \tau g'(0) - \int_0^\tau (x-\tau) g''(x)\,\mathrm{d}x. \end{aligned}$$

From the definition of the scalar product $\langle \cdot, \cdot \rangle_H$ we get the following conditions for the function ϕ_τ:

$$\begin{cases} \phi_\tau(0) = 1 \\ \phi'_\tau(0) = \tau \\ \phi''_\tau(x) = (\tau - x) 1_{[0,\tau]}(x), \quad x \in [0, \bar{\tau}]. \end{cases}$$

Integrating two time we arrive at:

$$\begin{aligned} \phi'_\tau(x) &= \tau - \tfrac{1}{2}(x \wedge \tau)^2 + \tau(x \wedge \tau), \quad x \in [0, \bar{\tau}], \\ \phi_\tau(x) &= 1 - \tfrac{1}{6}(x \wedge \tau)^3 + \tfrac{\tau}{2}(x \wedge \tau)^2 - \tfrac{\tau^2}{2}(x \wedge \tau) + x(1 + \tfrac{\tau}{2})\tau, \quad x \in [0, \bar{\tau}]. \end{aligned}$$

Proof of Theorem 3.3.2

The transpose (adjoint operator) $M^\top : \mathbb{R}^n \to H$ of the linear map $M \colon H \to \mathbb{R}^n$ is defined by:

$$\langle Mg, z \rangle_{\mathbb{R}^n} = \langle g, M^\top z \rangle_H, \quad \forall g \in H, \forall z \in \mathbb{R}^n.$$

Using the definition of M and the Riesz representation of the linear functional Φ we easily get:

$$M^\top z = \sum_{j=1}^N \phi_{x_j} C_j^\top z, \quad z \in \mathbb{R}^n,$$

where C_j represents the j-th column of the matrix C.

The Lagrangian $\mathscr{L}\colon H \times \mathbb{R}^n \to \mathbb{R}$ for problem (3.5) is defined as

$$\begin{aligned}\mathscr{L}(g,\lambda) &= \frac{1}{2}\|g\|_H^2 + \lambda^\top \left(Mg - p\right) \\ &= \frac{1}{2}\|g\|_H^2 + \langle \lambda, Mg\rangle_{\mathbb{R}^n} - \langle \lambda, p\rangle_{\mathbb{R}^n} \\ &= \frac{1}{2}\|g\|_H^2 + \langle M^\top \lambda, g\rangle_H - \langle \lambda, p\rangle_{\mathbb{R}^n}.\end{aligned}$$

The optimizers g^* and λ^* satisfy the following first-order conditions with respect to the Fréchet derivative in H and $\mathbb{R}^n$

$$g^* + M^\top \lambda^* = 0 \tag{C.1}$$

$$Mg^* - p = 0. \tag{C.2}$$

From (C.1) we get $g^* = -M^\top \lambda^*$. Plugging this into (C.2) gives $-MM^\top \lambda^* = p$. Observe now that $MM^\top : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map that can be represented by the matrix $CAC^\top$, where A is the positive definite $N \times N$ matrix with components

$$A_{ij} = \langle \phi_{x_i}, \phi_{x_j}\rangle_H = \phi_{x_i}(x_j) = \phi_{x_j}(x_i).$$

We now obtain the following unique solution for the optimal Lagrange multiplier and optimal discount curve:

$$\lambda^* = -\left(CAC^\top\right)^{-1} p, \quad g^* = M^\top \left(CAC^\top\right)^{-1} p.$$

Note that the matrix $CAC^\top$ is invertible because A is positive definite and C has full rank.

The map

$$M^+ \colon \mathbb{R}^n \to H,\ z \mapsto M^\top \left(MM^\top\right)^{-1} z,$$

is known as the *Moore–Penrose pseudoinverse* of the linear map M. We can therefore write the optimal discount curve as

$$g^* = M^+ p.$$

Proof of Lemma 3.4.1

The optimal discount curve g^* can be written as

$$g^*(x) = \sum_{j=1}^N z_j \phi_{x_j}(x) = p^\top \left(CAC^\top\right)^{-1} C\phi(x), \tag{C.3}$$

with $\phi(x) := (\phi_{x_1}(x), \ldots, \phi_{x_N}(x))^\top$. Differentiating (C.3) with respect to the components of p immediately gives the first statement of the theorem:

$$(D_p g^* \cdot v)(x) = v^\top \left(CAC^\top\right)^{-1} C\phi(x).$$

Differentiating (C.3) with respect to the cashflow C_{ij} gives

$$\begin{aligned}\frac{\partial g^*}{\partial C_{ij}}(x) &= -p^\top \left(CAC^\top\right)^{-1} \frac{\partial CAC^\top}{\partial C_{ij}} \left(CAC^\top\right)^{-1} C\phi(x) + p^\top \left(CAC^\top\right)^{-1} I_{ij}\phi(x) \\ &= p^\top \left(CAC^\top\right)^{-1} \left(I_{ij} - (CAI_{ji} + I_{ij}AC^\top)\left(CAC^\top\right)^{-1} C\right)\phi(x),\end{aligned}$$

where $I_{ij} \in \mathbb{R}^{n\times N}$ denotes a matrix with the (i,j)-th entry equal to one and all of the other entries equal to zero. The second statement of the theorem now easily follows from the distributive property of matrix multiplication and

$$\sum_{\substack{1\le i\le n\\ 1\le j\le N}} m_{ij} I_{ij} = m.$$

Proof of Lemma 3.4.2

Using the notation of the proof of Theorem 3.3.2, the optimal discount curve g^* can be written as

$$g^* = M^+ p = M^\top (MM^\top)^{-1} p.$$

Using the fact that $M^\top$ is the dual operator of M, we get:

$$\begin{aligned}\|g^*\|^2 &= \langle g^*, g^*\rangle_H \\ &= \left\langle M^\top (MM^\top)^{-1} p, M^\top (MM^\top)^{-1} p\right\rangle_H \\ &= \left\langle (MM^\top)^{-1} p, MM^\top (MM^\top)^{-1} p\right\rangle_{\mathbb{R}^n} \\ &= \left\langle (CAC)^{-1} p, p\right\rangle_{\mathbb{R}^n} \\ &= p^\top (CAC^\top)^{-1} p.\end{aligned}$$

Proof of Theorem 3.6.1

The Lagrangian is defined as

$$\mathscr{L}(\lambda, d) = \frac{1}{2}\|Ad\|_K^2 + \lambda^\top (Cd - p).$$

The first-order optimiality conditions for the optimal λ^* and d^* are

$$A^\top A d^* + C^\top \lambda^* = 0, \quad Cd^* - p = 0.$$

Straightforward calculations give the following unique solution:

$$d^* = (A^\top A)^{-1} C^\top \left(C(A^\top A)^{-1} C^\top\right)^{-1} p.$$

Using the fact that A is invertible and defining $M := CA^{-1}$, we finally obtain

$$d^* = A^{-1}M^+ p,$$

where $M^+ = M^\top (MM^\top)^{-1}$ is the Moore–Penrose pseudoinverse of the matrix M.

Bibliography

Ackerer, D. and D. Filipović (2019a). Linear credit risk models. *Finance and Stochastics*, Forthcoming.

Ackerer, D. and D. Filipović (2019b). Option pricing with orthogonal polynomial expansions. *Mathematical Finance*, Forthcoming.

Ackerer, D., D. Filipović, and S. Pulido (2018). The Jacobi stochastic volatility model. *Finance and Stochastics 22*(3), 667–700.

Adams, K. (2001). Smooth interpolation of zero curves. *Algo Research Quarterly 4*(1/2), 11–22.

Adams, K. J. and D. R. Van Deventer (1994). Fitting yield curves and forward rate curves with maximum smoothness. *Journal of Fixed Income 4*(1), 52–62.

Agmon, N., Y. Alhassid, and R. D. Levine (1979). An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics 30*(2), 250–258.

Al-Mohy, A. H. and N. J. Higham (2011). Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM Journal on Scientific Computing 33*(2), 488–511.

Andersen, L. (2007). Discount curve construction with tension splines. *Review of Derivatives Research 10*(3), 227–267.

Andersen, L. and V. Piterbarg (2010). *Interest Rate Modeling–Volume I: Foundations and Vanilla Models*. Atlantic Financial Press.

Aprahamian, H. and B. Maddah (2015). Pricing Asian options via compound gamma and orthogonal polynomials. *Applied Mathematics and Computation 264*, 21–43.

Asmussen, S., P.-O. Goffard, and P. J. Laub (2016). *Risk and Stochastics - Festschrift for Ragnar Norberg*, Chapter Orthonormal polynomial expansions and lognormal sum densities. World Scientific.

Avellaneda, M. (1998). Minimum-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance 1*(04), 447–472.

Bally, V. (2003). *An elementary introduction to Malliavin calculus*. Ph. D. thesis, INRIA.

Barone-Adesi, G., H. Rasmussen, and C. Ravanelli (2005). An option pricing formula for the GARCH diffusion model. *Computational Statistics & Data Analysis 49*(2), 287–310.

Barzanti, L. and C. Corradi (1998). A note on interest rate term structure estimation using tension splines. *Insurance: Mathematics and Economics 22*(2), 139–143.

Bekaert, G. and S. R. Grenadier (1999). Stock and bond pricing in an affine economy. Technical report, National Bureau of Economic Research.

Benhamou, E. (2002). Fast Fourier transform for discrete Asian options. *Journal of Computational Finance 6*(1), 49–68.

Bernhart, G. and J.-F. Mai (2015). Consistent modeling of discrete cash dividends. *Journal of Derivatives 22*(3), 9–19.

Björck, k. and V. Pereyra (1970). Solution of Vandermonde systems of equations. *Mathematics of Computation 24*(112), 893–903.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics 3*(1-2), 167–179.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Bos, M., A. Shepeleva, and A. Gairat (2003). Dealing with discrete dividends. *Risk 16*(9), 109–112.

Bos, M. and S. Vandermark (2002). Finessing fixed dividends. *Risk 15*(1), 157–158.

Brennan, M. J. (1998). Stripping the S&P 500 index. *Financial Analysts Journal 54*(1), 12–22.

Brennan, M. J. and E. S. Schwartz (1979). A continuous time approach to the pricing of bonds. *Journal of Banking & Finance 3*(2), 133–155.

Buchen, P. W. and M. Kelly (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis 31*(1), 143–159.

Buehler, H. (2010). Volatility and dividends–Volatility modelling with cash dividends and simple credit risk. *Working Paper*.

Buehler, H. (2015). Volatility and dividends II–Consistent cash dividends. *Working Paper*.

Buehler, H., A. S. Dhouibi, and D. Sluys (2010). Stochastic proportional dividends. *Working Paper*.

Caliari, M., P. Kandolf, A. Ostermann, and S. Rainer (2014). Comparison of software for computing the action of the matrix exponential. *BIT Numerical Mathematics 54*(1), 113–128.

Carmona, P., F. Petit, and M. Yor (1997). On the distribution and asymptotic results for exponential functionals of Lévy processes. *Exponential functionals and principal values related to Brownian motion*, 73–121.

Carr, P. and S. Willems (2019). A lognormal type stochastic volatility model with quadratic drift. *Working Paper*.

Carverhill, A. and L. Clewlow (1990). Valuing average rate (Asian) options. *Risk 3*(4), 25–29.

Černỳ, A. and I. Kyriakou (2011). An improved convolution algorithm for discretely sampled Asian options. *Quantitative Finance 11*(3), 381–389.

Chance, D. M., R. Kumar, and D. R. Rich (2002). European option pricing with discrete stochastic dividends. *Journal of Derivatives 9*(3), 39–45.

Chiu, N.-C., S.-C. Fang, J. E. Lavery, J.-Y. Lin, and Y. Wang (2008). Approximating term structure of interest rates using cubic L_1 splines. *European Journal of Operational Research 184*(3), 990–1004.

Collin-Dufresne, P. and R. S. Goldstein (2002a). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *Journal of Finance 57*(4), 1685–1730.

Collin-Dufresne, P. and R. S. Goldstein (2002b). Pricing swaptions within an affine framework. *Journal of Derivatives 10*(1), 9–26.

Corrado, C. J. and T. Su (1996a). Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research 19*(2), 175–192.

Corrado, C. J. and T. Su (1996b). S&P 500 index option tests of Jarrow and Rudd's approximate option valuation formula. *Journal of Futures Markets 16*(6), 611–629.

Cox, A. M. and D. G. Hobson (2005). Local martingales, bubbles and option prices. *Finance and Stochastics 9*(4), 477–492.

Cuchiero, C., M. Keller-Ressel, and J. Teichmann (2012). Polynomial processes and their applications to mathematical finance. *Finance and Stochastics 16*(4), 711–740.

Curran, M. (1994). Valuing Asian and portfolio options by conditioning on the geometric mean price. *Management science 40*(12), 1705–1711.

d'Addona, S. and A. H. Kind (2006). International stock–bond correlations in a simple affine asset pricing model. *Journal of Banking and Finance 30*(10), 2747–2765.

Dechow, P. M., R. G. Sloan, and M. T. Soliman (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies 9*(2-3), 197–228.

Delbaen, F. and S. Lorimier (1992). Estimation of the yield curve and the forward rate curve starting from a finite number of observations. *Insurance: Mathematics and Economics 11*(4), 259–269.

Donati-Martin, C., R. Ghomrasni, and M. Yor (2001). On certain Markov processes attached to exponential functionals of Brownian motion; applications to Asian options. *Revista Matematica Iberoamericana 17*(1), 179.

Duffie, D., D. Filipović, and W. Schachermayer (2003). Affine processes and applications in finance. *Ann. Appl. Probab. 13*(3), 984–1053.

Dufresne, D. (1990). The distribution of a perpetuity, with applications to risk theory and pension funding. *Scandinavian Actuarial Journal 1990*(1), 39–79.

Dufresne, D. (2000). Laguerre series for Asian and other options. *Mathematical Finance 10*(4), 407–428.

Eydeland, A. and H. Geman (1995). Domino effect: Inverting the Laplace transform. *Risk 8*(4), 65–67.

Fengler, M. R. and L.-Y. Hin (2015). A simple and general approach to fitting the discount curve under no-arbitrage constraints. *Finance Research Letters 15*, 78–84.

Filipović, D. (2000). Exponential-polynomial families and the term structure of interest rates. *Bernoulli 6*(6), 1081–1107.

Filipović, D. and M. Larsson (2016). Polynomial diffusions and applications in finance. *Finance and Stochastics 20*(4), 931–972.

Filipović, D. and M. Larsson (2017). Polynomial jump-diffusion models. *Swiss Finance Institute Research Paper* (17-60).

Filipović, D., M. Larsson, and A. B. Trolle (2017). Linear-rational term structure models. *Journal of Finance 72*, 655–704.

Filipović, D., E. Mayerhofer, and P. Schneider (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics 176*(2), 93 – 111.

Filipović, D. and S. Willems (2018). Exact smooth term structure estimation. *SIAM Journal on Financial Mathematics 9*(3), 907–929.

Filipović, D. and S. Willems (2019). A term structure model for dividends and interest rates. *Swiss Finance Institute Research Paper* (17-52).

Fournié, E., J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi (1999). Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics 3*(4), 391–412.

Frishling, V. and J. Yamamura (1996). Fitting a smooth forward rate curve to coupon instruments. *The Journal of Fixed Income 6*(2), 97–103.

Fu, M. C., D. B. Madan, and T. Wang (1999). Pricing continuous Asian options: a comparison of Monte Carlo and Laplace transform inversion methods. *Journal of Computational Finance 2*(2), 49–74.

Fusai, G., D. Marazzina, and M. Marena (2011). Pricing discretely monitored asian options by maturity randomization. *SIAM Journal on Financial Mathematics 2*(1), 383–403.

Fusai, G. and A. Meucci (2008). Pricing discretely monitored Asian options under Lévy processes. *Journal of Banking & Finance 32*(10), 2076–2088.

Fusai, G. and A. Tagliani (2002). An accurate valuation of Asian options using moments. *International Journal of Theoretical and Applied Finance 5*(02), 147–169.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press.

Geiß, C. and R. Manthey (1994). Comparison theorems for stochastic differential equations in finite and infinite dimensions. *Stochastic Processes and their Applications 53*(1), 23–35.

Geman, H. and M. Yor (1993). Bessel processes, Asian options, and perpetuities. *Mathematical Finance 3*(4), 349–375.

Geske, R. (1978). The pricing of options with stochastic dividend yield. *Journal of Finance 33*(2), 617–625.

Gourieroux, C. and A. Monfort (2013). Linear-price term structure models. *Journal of Empirical Finance 24*, 24–41.

Guennoun, H. and P. Henry-Labordère (2017). Equity modeling with stochastic dividends. *Working Paper*.

Hagan, P. S. and G. West (2006). Interpolation methods for curve construction. *Applied Mathematical Finance 13*(2), 89–129.

Hagan, P. S. and G. West (2008). Methods for constructing a yield curve. *Wilmott Magazine, May*, 70–81.

Heyde, C. (1963). On a property of the lognormal distribution. *Journal of the Royal Statistical Society, Series B 25*, 392–393.

Holly, A., A. Monfort, and M. Rockinger (2011). Fourth order pseudo maximum likelihood methods. *Journal of Econometrics 162*(2), 278–293.

Jackwerth, J. C. and M. Rubinstein (1996). Recovering probability distributions from option prices. *Journal of Finance 51*(5), 1611–1631.

Jacod, J. and A. Shiryaev (2003). *Limit Theorems for Stochastic Processes*, Volume 2. Springer-Verlag.

James, J. and N. Webber (2000). *Interest Rate Modelling*. Wiley-Blackwell Publishing Ltd.

Jarrow, R. and A. Rudd (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics 10*(3), 347–369.

Jarrow, R. A., P. Protter, and K. Shimbo (2007). Asset price bubbles in complete markets. *Advances in Mathematical Finance*, 97–121.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review 106*(4), 620.

Jondeau, E. and M. Rockinger (2001). Gram–Charlier densities. *Journal of Economic Dynamics and Control 25*(10), 1457–1483.

Ju, N. (2002). Pricing Asian and basket options via Taylor expansion. *Journal of Computational Finance 5*(3), 79–103.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus* (2nd ed.). Springer-Verlag.

Kemna, A. G. and A. Vorst (1990). A pricing method for options based on average asset values. *Journal of Banking & Finance 14*(1), 113–129.

Kim, I.-M. (1995). An alternative approach to dividend adjustments in option pricing models. *Journal of Financial Engineering 4*, 351–373.

Korn, R. and L. G. Rogers (2005). Stocks paying discrete dividends: modeling and option pricing. *Journal of Derivatives 13*(2), 44–48.

Kragt, J., F. De Jong, and J. Driessen (2018). The dividend term structure. *Journal of Financial and Quantitative Analysis*, Forthcoming.

Kwon, O. K. (2002). A general framework for the construction and the smoothing of forward rate curves. Technical report, University of Technology Sydney.

Lapeyre, B., E. Temam, et al. (2001). Competitive Monte Carlo methods for the pricing of Asian options. *Journal of Computational Finance 5*(1), 39–58.

Lasserre, J.-B., T. Prieto-Rumeau, and M. Zervos (2006). Pricing a class of exotic options via moments and SDP relaxations. *Mathematical Finance 16*(3), 469–494.

Laurini, M. P. and M. Moura (2010). Constrained smoothing B-splines for the term structure of interest rates. *Insurance: Mathematics and Economics 46*(2), 339–350.

Lemke, W. and T. Werner (2009). The term structure of equity premia in an affine arbitrage free model of bond and stock market dynamics. Technical report, ECB Working Paper.

Lettau, M. and J. A. Wachter (2007). Why is long-horizon equity less risky? A duration-based explanation of the value premium. *Journal of Finance 62*(1), 55–92.

Lettau, M. and J. A. Wachter (2011). The term structures of equity and interest rates. *Journal of Financial Economics 101*(1), 90–113.

Levy, E. (1992). Pricing European average rate currency options. *Journal of International Money and Finance 11*(5), 474–491.

Li, W. and S. Chen (2016). Pricing and hedging of arithmetic Asian options via the Edgeworth series expansion approach. *The Journal of Finance and Data Science 2*(1), 1–25.

Lim, K. G. and Q. Xiao (2002). Computing maximum smoothness forward rate curves. *Statistics and Computing 12*(3), 275–279.

Linetsky, V. (2004). Spectral expansions for Asian (average price) options. *Operations Research 52*(6), 856–867.

Lioui, A. (2006). Black-Scholes-Merton revisited under stochastic dividend yields. *Journal of Futures Markets 26*(7), 703–732.

Litterman, R. B. and J. Scheinkman (1991). Common factors affecting bond returns. *The Journal of Fixed Income 1*(1), 54–61.

Lorimier, S. (1995). *Interest rate term structure estimation based on the optimal degree of smoothness of the forward rate curve*. Ph. D. thesis, University of Antwerp.

Mamaysky, H. et al. (2002). On the joint pricing of stocks and bonds: Theory and evidence. Technical report, Yale School of Management.

Manzano, J. and J. Blomvall (2004). Positive forward rates in the maximum smoothenss framework. *Quantitative finance 4*(2), 221–232.

Marchioro, M. (2016). Seasonality of dividend point indexes. *Statpro Quantitative Research Series*.

Marcozzi, M. D. (2003). On the valuation of Asian options by variational methods. *SIAM Journal on Scientific Computing 24*(4), 1124–1140.

McCulloch, J. H. (1971). Measuring the term structure of interest rates. *The Journal of Business 44*(1), 19–31.

McCulloch, J. H. (1975). The tax-adjusted yield curve. *The Journal of Finance 30*(3), 811–830.

Mead, L. R. and N. Papanicolaou (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics 25*(8), 2404–2417.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics 4*(1), 141–183.

Milevsky, M. A. and S. E. Posner (1998). Asian options, the sum of lognormals, and the reciprocal gamma distribution. *Journal of Financial and Quantitative Analysis 33*(3), 409–422.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 473–489.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*(1), 7–38.

Nualart, D. (2006). *The Malliavin calculus and related topics*. Springer.

Overhaus, M., A. Bermúdez, H. Buehler, A. Ferraris, C. Jordinson, and A. Lamnouar (2007). *Equity Hybrid Derivatives*. John Wiley & Sons.

Pilipović, D. (1997). *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill.

Ritchken, P., L. Sankarasubramanian, and A. M. Vijh (1993). The valuation of path dependent contracts on the average. *Management Science 39*(10), 1202–1213.

Rockinger, M. and E. Jondeau (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics 106*(1), 119–142.

Rogers, L. C. G. and Z. Shi (1995). The value of an Asian option. *Journal of Applied Probability 32*(4), 1077–1088.

Rompolis, L. S. (2010). Retrieving risk neutral densities from european option prices based on the principle of maximum entropy. *Journal of Empirical Finance 17*(5), 918–937.

Shaw, W. (2002). Pricing Asian options by contour integration, including asymptotic methods for low volatility. *Working paper*.

Shea, G. S. (1984). Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations. *Journal of Financial and Quantitative Analysis 19*(03), 253–269.

Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets 22*(5), 393–426.

Steeley, J. M. (1991). Estimating the Gilt-edged term structure: Basis splines and confidence intervals. *Journal of Business Finance & Accounting 18*(4), 513–529.

Sun, J., L. Chen, and S. Li (2013). A quasi-analytical pricing model for arithmetic Asian options. *Journal of Futures Markets 33*(12), 1143–1166.

Suzuki, M. (2014). Measuring the fundamental value of a stock index through dividend future prices. *Working Paper*.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, National Bureau of Economic Research.

Tanggaard, C. (1997). Nonparametric smoothing of yield curves. *Review of Quantitative Finance and Accounting 9*(3), 251–267.

Thompson, G. (2002). Fast narrow bounds on the value of Asian options. Technical report, Judge Institute of Management Studies.

Tunaru, R. S. (2018). Dividend derivatives. *Quantitative Finance 18*(1), 63–81.

Turnbull, S. M. and L. M. Wakeman (1991). A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis 26*(3), 377–389.

Turner, L. R. (1966). Inverse of the Vandermonde matrix with applications. Technical report, NASA.

Vanmaele, M., G. Deelstra, J. Liinev, J. Dhaene, and M. J. Goovaerts (2006). Bounds for the price of discrete arithmetic Asian options. *Journal of Computational and Applied Mathematics 185*(1), 51–90.

Vasicek, O. A. and H. G. Fong (1982). Term structure modeling using exponential splines. *The Journal of Finance 37*(2), 339–348.

Vecer, J. (2001). A new PDE approach for pricing arithmetic average Asian options. *Journal of Computational Finance 4*(4), 105–113.

Vecer, J. (2002). Unified pricing of Asian options. *Risk 15*(6), 113–116.

Vellekoop, M. H. and J. W. Nieuwenhuis (2006). Efficient pricing of derivatives on assets with discrete dividends. *Applied Mathematical Finance 13*(3), 265–284.

Weber, M. (2018). Cash flow duration and the term structure of equity returns. *Journal of Financial Economics 128*(3), 486–503.

Willems, S. (2019a). Asian option pricing with orthogonal polynomials. *Quantitative Finance 19*(4), 605–618.

Willems, S. (2019b). Linear stochastic dividend model. *arXiv preprint arXiv:1908.05850*.

Yan, W. (2014). Estimating a unified framework of co-pricing stocks and bonds. *Working Paper*.

Yor, M. (1992). On some exponential functionals of Brownian motion. *Advances in Applied Probability 24*(3), 509–531.

Zvan, R., P. A. Forsyth, and K. R. Vetzal (1996). Robust numerical methods for PDE models of Asian options. Technical report, University of Waterloo, Faculty of Mathematics.

Sander Willems

Contact Information

Email: willems.sander@gmail.com
Tel.: (+41) 78 818 73 38
Web: `www.sanderwillems.com`

Swiss Finance Institute at EPFL
Quartier UNIL-Dorigny, Extranef 220
CH-1015 Lausanne

Education

- **École Polytechnique Fédérale de Lausanne**, Swiss Finance Institute
  Ph.D., Mathematical Finance 2014 – 2019
  Supervisor: Prof. Damir Filipović
  Thesis title: Pricing interest rate, dividend, and equity risk.
- **New York University**, Tandon School of Engineering
  Visiting Scholar Fall 2018
  Host: Prof. Peter Carr
- **Université Libre de Bruxelles**, Solvay Business School
  Advanced Master, Quantitative Finance 2013 – 2014
  Graduated with high honors
- **Ghent University**
  Bachelor of Science, Mathematics 2008 – 2011
  Master of Science, Mathematics 2011 – 2013
  Graduated with great distinction

Academic Experience

- **Adjunct Professor**, New York University, USA Fall 2018
  Quantitative Methods in Finance, Master of Financial Engineering
  Student course evaluation: 5/5
- **Teaching Assistant**, EPFL, Switzerland 2015 – 2018
  Fixed Income Analysis, Master of Financial Engineering
  Interest Rate Models, Coursera MOOC

Professional Experience

- **Quantitative Analyst Intern**, AXA, Belgium Summer 2014
  Summer internship in AXA Belgium's Investments & ALM department.
  Implemented equity option pricing models with stochastic interest rates and stochastic volatility.
- **Risk Reporting Intern**, GIMV, Belgium Summer 2011
  Summer internship in one of the largest Belgian private equity firms.

Programming

Daily usage: Matlab, LaTeX
Basic knowledge: `C#`, `C++`, Mathematica, Python, VBA

Awards and Fellowships

- *Best Teaching Assistant Award*, Master of Financial Engineering, EPFL 2018
- *Best Discussant Doctoral Award*, Swiss Finance Institute Research Days 2018
- *Outstanding Presentation Award*, 8th General AMaMeF conference 2017
- *Swiss Finance Institute Graduate Fellowship* 2014
- *Valedictorian of AM in Quantitative Finance*, Solvay Business School 2014
- *Merit Based Tuition Fellowship*, Solvay Business School 2013

Research Interests

Mathematical finance, derivative pricing, term-structure models, affine and polynomial processes, hybrid derivatives

Other Information

- Nationality: Belgian
- Languages: Dutch (native), English (fluent), French (working proficiency)
- Leisure: hiking, skiing (certified instructor)

**Ce document a été imprimé au Centre d'impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.**