**The London School of Economics and Political Science**

*Essays on Mathematical Finance: Applications of Moment Expansions and Filtering Theory*

Takeshi Yamada

A thesis submitted to the Department of Statistics of the London School of Economics for the degree of Doctor of Philosophy, London, June 2010.

UMI Number: U613433

UMI

Dissertation Publishing

ProQuest

# Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without the prior written consent of the author.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

# Acknowledgment

# Abstract

This thesis contains three essays on mathematical finance. The first discusses approximation methods for pricing swaptions based on moment expansions with multi-factor affine jump-diffusion models in Chapter 2. Two methods are examined. One is based on a Gram–Charlier expansion and the other is based on a generalized Edgeworth expansion. The density function of the forward swap is replaced with more tractable functions and their moments. Numerical simulations are conducted to confirm their accuracy. Models with a Gaussian-type or CIR-type volatility with an exponentially, normally or truncated normally distributed jump size are employed.

The second essay proposes a framework to study the spot and forward relationship in carbon allowances markets and the third deals with the same problem in a different setting. The framework is based on the no-arbitrage principle. The value of the spot price depends on two underlying variables: the forward price and the net position of the zone defined as the difference between allocated carbon allowances and emissions. In Chapter 3, the net position of the market is modelled as a Markov chain and in Chapter 4, as a linear diffusion. Two kinds of filtration used in pricing are considered: complete information where market participants observe both processes continuously and incomplete information where they observe the forward price continuously and the net position of the zone periodically. Pricing problems occur in an incomplete market, since the net position of the zone is not tradable. A locally risk-minimization approach is used to fix the martingale measure. Under the complete information setting, The analytical spot price is obtained. Under the incomplete information setting, a filtered process is used for pricing, leading to the use of filtering theory. The spot price is computed numerically. Chapter 5 concludes.

# Contents

# List of Tables

# List of Figures

8

# Chapter 1

# Introduction

The purpose of this introductory chapter is to provide a brief introduction and motivation for each article. The main part of this dissertation consists of three chapters, Chapters 2-4.

## Approximating Swaption Prices with Moment Expansions

In Chapter 2 on Approximating Swaption Prices with Moment Expansions, the validity of two approximation methods based upon moment expansions are investigated with the framework of multi-factor affine jump-diffusion models. From a practical point of view there are several requirements, such as analytical tractability and multi-factors, for modeling interest rates in order to price and hedge derivatives contracts and measure their risks. One of these is consistency. To handle a portfolio including several types of interest rate and bond derivatives contracts, pricing them, computing hedge parameters and measuring their risks with one fixed model leads to consistency. If each derivatives contract is evaluated with a different model, then this results in inconsistency, meaning that their risks are measured by relying on different assumptions. For this reason, pricing several types of derivatives contracts with one model is preferable. Some derivatives contracts, however, do not have analytical solutions for a model which does give analytical solutions for other types of derivatives contracts. A simulation approach is applicable for evaluating such derivatives contracts, albeit a time-consuming procedure. If an approximation method returns accurate prices and works efficiently, it is useful to practitioners. This is the main motivation for this research. To price interest rate and bond derivatives, several models are available and the multi-factor affine term structure model is one of the dominant

frameworks for modeling short rates. Generally speaking, it gives an analytically tractable approach for cap, floor and options on zero-coupon bonds, but not for options on swaps or on coupon-bearing bonds. In this light, we consider approximation methods for pricing swaptions with affine jump-diffusion models.

The moment expansion approaches are characterised by replacing the density function of a target with a more tractable density and adjusting their differences using cumulants. One of these is referred to as a Gram–Charlier expansion. In this expansion, a target density is approximated with the density of normal distributions in additive form and coefficients are written with the cumulants of the target random variable. Consequently, calculating swaption prices is reduced to calculating the cumulants of the underlying swap value.

Normal distributions are frequently used when approximating a density function. If the density of a target deviates from normality, which is the case when jump terms are included in a model, the accuracy of the Gram–Charlier expansion would deteriorate even if deterioration can be corrected by including higher-order cumulants. In a generalized Edgeworth expansion, any distribution can be used to approximate a target density. The idea behind using this expansion is that an approximation adopting a similar distribution to that of the underlying swap value would be expected to generate better results with a lower-order expansion. By focusing attention on swap cash-flows, we propose using the zero-coupon bond at swap maturity as a candidate for approximation since it has the largest cash-flow. By employing this, the swaption prices can be decomposed into options on the zero-coupon bond and adjustment terms. Due to the multi-factor underlying process, each term in the decomposition does not have an analytical solution, hence the options on the zero-coupon bond and adjustment terms are computed with the Gram–Charlier expansion. As a result, calculating the swaption prices is reduced to calculating the cumulants of the underlying swap value in both approaches.

While assuming the existence of cumulants, these two methods are model-free approaches. Analysis of historical data suggests the existence of jumps and shows that models containing jumps produce more accurate estimates of the interest rate curves. Affine jump-diffusion models are therefore employed to

confirm accuracy and computational burden of these approaches.

Since the moments of swap values are the product of zero-coupon bonds, as long as the models which have analytical solutions for zero-coupon bond prices are implemented, they yield the moments of swap values analytically as well. Calculating the moments of swap values is reduced to solving ordinary differential equations called Riccati equations. For example, the multi-factor model with Gaussian-type volatility, exponentially distributed jump sizes and constant jump intensity and that with CIR-type volatility, exponentially distributed jump sizes and constant jump intensity provide analytical solutions for both zero-coupon bonds and moments of swap values. We implement both analytical and numerical solvable models and compare their computational burdens with that of Monte Carlo simulation.

Five numerical examples are presented: Gaussian-type volatility and CIR-type volatility combined with an exponentially, normally or truncated normally distributed jump size. In each example, we compute the prices with Monte Carlo simulation, the Gram–Charlier and generalized Edgeworth expansions and compare their respective prices across various strike rates. Although the model with Gaussian-type volatility or CIR-type volatility and an exponentially distributed jump size leads to analytical solutions for Riccati equations, others do not have closed-form solutions. The Runge-Kutta method is employed to solve Riccati equations numerically. We compare their validity based on accuracy and computational time.

**Forward and Spot Relationship in Carbon Emissions Markets**

In Chapters 3 and 4, the spot and forward relationship in the EU Emission Trading Scheme (ETS) is investigated. We propose a framework to study it. Due to regulations and market design, the spot and forward relationship in this market is different from that of other markets such as zero-coupon bonds or stocks. The framework is based on the no-arbitrage argument. The EU ETS is one of the commodities markets and is relatively new. Free carbon emission allowances are allocated to carbon-intensive companies and these companies can emit carbon dioxide up to their allocation. There are spot and futures markets

in which companies can trade allowances to offset their excess or shortage of allowances. Trading periods in EU ETS can be split into several phases. We consider the spot and forward relationship in Phase II by extending the preceding paper by Çetin and Verschuere (2009) in which the spot and forward relationship in Phase I is investigated. Phase II prices differ from Phase I prices in the availability of banking. Banking allows companies to carry forward their unused allowances into the next year.

The pricing formula is derived by assuming an exogenous price process for the forward contract. The proportional changes in the forward price are driven by a Brownian motion and its drift term is modelled to include an unobservable variable. An insight of our setting is that the net position of the zone which is unobservable plays a crucial role in determining the spot and forward relationship. It is defined as the difference between total allocations and emissions. Depending on the market condition expressed by the net position of the zone, the reltionship varies, indicating the spot price is considered an option on both the forward price and net positon of the zone. In Chapter 3, the net position of the market is assumed to follow a continuous-time and finite-state Markov chain. In Chapter 4, it is assumed to follow a linear diffusion. We focus on differences in the information set that market participants observe. In each chapter two filtrations used in pricing are considered, complete information where market participants can observe both the forward process and net position of the zone continuously and incomplete information where market participants can observe the forward price continuously and the net position of the zone only at announcement times.

First we consider pricing under the complete information setting. This pricing considered as being in an incomplete market, since the net position of the zone is not tradable. A locally risk-minimization approach is used to derive a martingale measure referred to as the minimal martingale measure. With this approach, the martingale parts which are orthogonal to tradable instruments remain unchanged and the spot price is obtained analytically.

Next, pricing under the incomplete information setting is investigated. The spot and forward relationship is derived with the help of filtering theory. Since

market participants do not always observe the net position of the zone, the pricing is done under the filtration generated by forward prices which are observed at any time and the net position of the zone which is observed periodically. From our modeling that has the drift term of the forward price as a function of the net position of the zone, the net position of the zone is observed through the fluctuations of the forward price. The distributions of underlying processes projected onto the new filtration are taken into consideration in a pricing equation. In Chapter 3 the framework that assumes the net position of the zone follows a Markov chain leads to the use of the Wonham filter. In Chapter 4 the framework leads to the use of the Kalman-Bucy filter. Since the net position of the zone is not tradable, this pricing is again that in an incomplete market. The minimal martingale measure is found and the distributions of the underlying processes under this measure are calculated. The spot price is computed with Monte Carlo simulation. The price differences between the complete and incomplete information settings are highlighted in numerical examples. Of interest are the jump of the net position of the zone at the announcement times and changes in prices.

Estimation is carried out in both chapters with historical data. Though our aim is to estimate parameters by maximizing the log-likelihood function of the data, it is not straightforward due to the existence of an unobservable variable. To avoid numerical routines, we employ the expectation-maximization (EM) algorithm for finding the parameters. In the EM algorithm, instead of maximizing the log-likelihood function including only observable variables, the complete data log-likelihood function including both observable and unobservable variables is maximized. The underlying processes are reformulated for an associated discrete-time version. In Chapter 3, since the unobservable process is modelled as a Markov chain and the observable process follows a diffusion process whose drift term is linked to the Markov chain, this system is called a hidden Markov model. In Chapter 4, since the unobservable process is modelled as a linear diffusion and the observable process follows a linear diffusion whose drift term is linked to the unobservable process, this system is called a Kalman filter.

# Chapter 2

# Approximating Swaption Prices with Moment Expansions

## 2.1 Introduction

In interest rate and bond markets, various kinds of derivatives are traded to cover possible risk exposures. Swaptions are among the most liquid derivatives traded and therefore the efficient calculation of their theoretical prices is important. The theoretical prices of these derivatives depend on the term structure model used. From a practical perspective, evaluating various kinds of derivatives with the same model leads to consistent risk management, for example, measuring the possible future loss on a portfolio or computing portfolio sensitivities to various risk factors. When using the same model to price different kinds of derivatives, there are some which do not have analytical solutions. In such cases, numerical methods are used although they are generally time-consuming. If an approximation method returns an accurate price and works fast, it is helpful to practitioners.

Affine term structure models were originally introduced by Duffie and Kan (1996) and this framework was extended to include jumps by Duffie, Pan and Singleton (2000) and Chacko and Das (2002). They are called affine jump-diffusion (hereafter AJD) models. AJD models comprise a widely-used class of asset pricing models which have both tractability and flexibility. As demonstrated by Duffie, Pan and Singleton (2000) and Chacko and Das (2002), prices of zero-coupon bonds and some options are given by solving the system of ordinary differential equations under AJD models. Some empirical studies support the assumption of jump components in the term structure of interest rates.

16

For instance, Das (2002) incorporated jump terms into the Vasicek model and showed the jump models produce more accurate estimates of the interest rate curves than pure-diffusion models using the daily Federal Funds rate. Johannes (2004) concluded that the surprise arrival of news about the macroeconomy causes jumps. Piazzesi (2005) proposed the interest rate model in which jump components represent macroeconomic announcements by the Federal Reserve.

European swaption prices can be obtained for one-factor models, since in this case the exercising boundary can be determined. Using it, swaption prices can be decomposed into a portfolio of options on zero-coupon bonds. Due to the difficulty of identifying the exercising boundary, a closed-form solution for swaptions has not been obtained for multi-factor models. Papers considering European swaption pricing in multi-factor affine term structure models include Munk (1999), Singleton and Umantsev (2002) and Collin-Dufresne and Goldstein (2002). Munk (1999) approximated swaption prices with an option on a zero-coupon bond whose maturity is equal to the stochastic duration. The paper by Singleton and Umantsev (2002) approximated the exercise region by linearization and simplified the pricing of swaptions to pricing several caplets. In their approach, exercising probabilities are computed with the Fourier inversion.

Collin-Dufresne and Goldstein (2002) proposed an approximation method for pricing European swaptions based on an Edgeworth expansion and Tanaka, Yamada and Watanabe (2010) applied a Gram-Charlier expansion and derived an alternative expression for swaption prices for multi-factor affine diffusion models. Both the Edgeworth and Gram-Charlier expansions use the density function of the normal distribution to approximate the density of swap values and require the calculation of higher-order moments of swap values, which are available in an exponential-affine form under the affine term structure models. This kind of approximation method, which replaces the density of the target variable with a more tractable density, can be used for pricing options in different settings such as pricing a European stock option. For example, Jarrow and Rudd (1982) proposed a generalized Edgeworth expansion to incorporate the effect of the skewness and kurtosis of an underlying stock into pricing an option on it. In their approach, a density function of an underlying variable is approximated with an arbitrary distribution. Turnbull and Wakeman (1991) applied the gen-

17

eralized Edgeworth expansion to evaluate a European average option. All these methods are referred to as moment expansions in the sense that the density of the target is approximated with a more tractable density and its higher-order moments.

The objective of this article is to investigate the validity of two methods based on the moment expansions for pricing European swaptions in multi-factor AJD models. The first method is to apply the Gram-Charlier expansion and it is, therefore, an extension of Tanaka, Yamada and Watanabe (2010) from affine diffusions to AJDs. The second one is to apply the generalized Edgeworth expansion. In the latter approach, European swaption values are decomposed into European options on a zero-coupon bond and adjustment terms. This decomposition indicates a swaption can be partially hedged with options on zero-coupon bonds.

The outline of the chapter is as follows. After this introductory section, we present the mathematical structure and no-arbitrage pricing used for evaluating swaptions. Two approximation methods, the Gram–Charlier and generalized Edgeworth expansion methods, are explained and the associated approximations for swaption prices are derived in Section 2.2. In Section 2.3, we review the AJD models. Section 2.4 tests the approximations numerically under Gaussian-type or CIR-type volatility with exponential, normal or truncated normal jump size settings. Section 2.5 concludes.

## 2.2 Approximating Swaption Prices with Moment Expansions

### 2.2.1 The Valuation of Interest Rate Derivatives

We will examine the mathematical structure of interest rate and bond markets. In this and the next subsection, we introduce some concepts of interest rates and derivative pricing used in the chapter. The approach is similar to that in Bingham and Kiesel (2004) and Musiela and Rutkowski (2005). For more details see them and references therein. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \leq T^*}$, where $T^*$ is a fixed terminal time horizon. Absence

of arbitrage is guaranteed by the existence of an equivalent martingale measure under which the process of relative price with respect to a numéraire of the basic security has to be a martingale.

**Definition 2.1** *(Risk-free short rate and money market account) Let $r_t$ denote the risk-free short rate at time $t$ over the infinitesimal time interval $[t, t + dt]$ and $r_t$ is assumed to be an adapted process. We may then introduce an adapted process, $D_t$, referred to as a money-market account which satisfies the following ordinary differential equation (ODE),*

$$dD_t = r_t D_t \, dt, \quad D_0 = 1.$$

*Consequently, it is given by*

$$D_t = \exp\left(\int_0^t r_s \, ds\right), \quad \forall \, t \in [0, T^*].$$

Fundamental instruments in interest rate and bond markets are zero-coupon bonds.

**Definition 2.2** *(Zero-coupon bonds) Zero-coupon bonds are the financial contracts which guarantee the holder the payment of one unit of cash at prescribed maturity $T \leq T^*$.*

We denote by $P(t, T)$ the time-$t$ price of a zero-coupon bond maturing at $T$. By definition, $P(T, T) = 1$. We assume that zero-coupon bonds with various maturities and a money-market account are tradable.

**Definition 2.3** *(No-arbitrage zero-coupon bond price) Zero-coupon bonds with several maturities, $P(t, T)$, $t \leq T \leq T^*$ are called an arbitrage-free bond family if the following conditions hold:*

- $P(T, T) = 1, \forall T$.

- *There exists a probability measure $\mathbb{Q}$ such that $\forall \, t \in [0, T]$ the relative price of a zero-coupon bond,*

$$\frac{P(t, T)}{D_t},$$

*is a martingale under $\mathbb{Q}$.*

19

To ensure no-arbitrage between zero-coupon bonds of different maturities, we assume there exists the risk-neutral measure $\mathbb{Q}$ under which the relative price of a zero-coupon bond is a martingale. The price of an $\mathcal{F}_T$-measurable contingent claim is evaluated as an expectation under $\mathbb{Q}$.

**Proposition 2.1** *(No-arbitrage price of a contingent claim) The no-arbitrage price of an $\mathcal{F}_T$-measurable contingent claim, H, is given by*

$$\pi_t := D_t \mathbb{E}_{\mathbb{Q}} \left[ D_T^{-1} H \mid \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{Q}} \left[ e^{-\int_t^T r_s \, ds} H \mid \mathcal{F}_t \right], \quad \forall t < T.$$

*where $\mathbb{E}_{\mathbb{Q}}[\cdot]$ is the expectation operator under measure $\mathbb{Q}$.*

With the assumption of no-arbitrage among zero-coupon bonds, the zero-coupon bond price can be written (noting $P(T, T) = 1$) as follows:

$$P(t, T) = D_t \mathbb{E}_{\mathbb{Q}} \left[ D_T^{-1} \mid \mathcal{F}_t \right], \quad \forall t < T.$$

This leads to the expression which zero-coupon bonds satisfy:

$$P(t, T) = \mathbb{E}_{\mathbb{Q}} \left[ e^{-\int_t^T r_s \, ds} \mid \mathcal{F}_t \right], \quad \forall t < T.$$

For evaluating a contingent claim, any strictly positive process can be used as a numéraire. Under the risk-neutral measure, the money-market account is used as the numéraire. Changing the pricing measure is permitted and is called the change of measure technique. In particular, the approach in which the zero-coupon bond price $P(t, T)$ is used as the numéraire is called the $T$-forward measure approach since under the measure, denoted by $\mathbb{P}_T$, the forward prices $P(t, U)/P(t, T)$, for any $U < T$, are martingales.

**Proposition 2.2** *(No-arbitrage price of a contingent claim under $\mathbb{P}_T$) Under the $T$-forward measure, an $\mathcal{F}_T$-measurable contingent claim H is evaluated as*

$$\pi_t := P(t, T) \mathbb{E}_{\mathbb{P}_T} \left[ H \mid \mathcal{F}_t \right], \quad \forall t < T.$$

*The Radon-Nikodým derivative is given by*

$$\frac{d\mathbb{P}_T}{d\mathbb{Q}} = \frac{P(t, T)}{D_t P(0, T)}.$$

The forward measure approach is convenient for computing the expectation in the case where $H$ is correlated with the money-market account. Under $\mathbb{P}_T$ we only need to know the distribution of the underlying process under the measure, while under $\mathbb{Q}$, the correlation between the money-market account and the payoff must be taken into consideration.

## 2.2.2 The Valuation of Swaptions

A swaption is an option which gives the option holder the right to enter into a swap contract in the future. The most common form of an interest-rate swap consists of exchanges of interest payments between two different institutions. For instance, one side pays a fixed interest rate on a notional amount semi-annually and the other side pays a floating rate on the same notional amount at the same time. The reference rate on the floating side is generally the LIBOR rate.

**Definition 2.4** *(LIBOR rate) LIBOR rate, denoted by $L(t,T)$, is the constant rate at which an investment has to be made to produce one unit of cash at $T$ starting at $t$, and is given by*

$$L(t,T) = \frac{1 - P(t,T)}{(T-t)P(t,T)}.$$

Consider the swap contract which has a unit notional amount, starting at $T_0$ and exchanging cash-flow at dates $T_1, T_2, \ldots, T_N$, which are set at regularly spaced time intervals, with $\delta = T_i - T_{i-1}$ for all $i$. At time $T_i$, $i = 1, 2, \ldots, N$, the fixed side pays the amount $\delta\kappa$, where $\kappa$ is a fixed rate, and the floating side pays the amount $\delta L(T_{i-1}, T_i)$, where $L(T_{i-1}, T_i)$ is the LIBOR rate reset at the previous instant $T_{i-1}$. The discounted total payoff at time $t < T_0$ of the fixed side is expressed as

$$\sum_{i=1}^{N} \delta\kappa P(t, T_i),$$

whereas the discounted total payoff of the floating side is expressed as

$$\sum_{i=1}^{N} \delta L(T_{i-1}, T_i) P(t, T_i) = \sum_{i=1}^{N} \frac{1 - P(T_{i-1}, T_i)}{P(T_{i-1}, T_i)} P(t, T_i)$$

$$= \sum_{i=1}^{N} P(t, T_{i-1}) - P(t, T_i)$$

$$= P(t, T_0) - P(t, T_N).$$

**Definition 2.5** *(Cash-flows of swap) The swap contract which receives a fixed amount and pays a floating amount is termed a receiver swap, and the one which receives a floating amount and pays a fixed amount is a payer swap. The value $SV(t)$ of the swap contract at time $t$ is therefore given by*

$$SV(t) = \begin{cases} -P(t, T_0) + \delta\kappa \sum_{i=1}^{N} P(t, T_i) + P(t, T_N), & \text{for the receiver's swap,} \\ P(t, T_0) - \delta\kappa \sum_{i=1}^{N} P(t, T_i) - P(t, T_N), & \text{for the payer's swap,} \end{cases}$$

$$=: \sum_{i=0}^{N} a_i P(t, T_i). \tag{2.1}$$

The swap contract which starts at some future time is called a forward start swap. Note that the swap rate which makes the forward start swap value zero is called an at-the-money forward (ATMF), and is defined as

$$ATMF = \frac{P(t, T_N) - P(t, T_0)}{\delta \sum_{i=1}^{N} P(t, T_i)}. \tag{2.2}$$

We consider a swaption with option expiry at $T_0$ which coincides with the first reset date on the floating rate.

**Definition 2.6** *(Swaption pricing) Swaptions are exercised only when the underlying swap value has positive value. Assuming absence of no-arbitrage, the swaption value, $SOV(t)$, at time $t$ is evaluated under the risk-neutral measure $\mathbb{Q}$. By setting $H = SV(t)$ from Definition 2.5, and applying Proposition 2.1, we get:*

$$SOV(t) = \mathbb{E}_{\mathbb{Q}} \left[ e^{-\int_t^{T_0} r_s ds} \max(SV(T_0), 0) | \mathcal{F}_t \right].$$

*By applying the change of measure technique, the swaption value is converted to the expected value of exercising, from Proposition 2.2, under the $T_0$-forward*

*measure, $\mathbb{P}_{T_0}$, as follows:*

$$SOV(t) = P(t, T_0)\mathbb{E}_{\mathbb{P}_{T_0}}\left[1_{\{SV(T_0)>0\}}SV(T_0) \mid \mathcal{F}_t\right] \qquad (2.3)$$

$$= P(t, T_0)\int_0^\infty xf(x)dx, \qquad (2.4)$$

*where $f$ is the density function of the swap value SV at the option expiry date $T_0$ under the $T_0$-forward measure, conditioned on $\mathcal{F}_t$.*

Although the swap value has a multiple cash-flow, an analytical solution is available under one-factor models since the exercising boundary is identified numerically (see, for example, Jamshidian (1989)). Under multi-factor models, there is no analytical solution since the exercising boundary cannot be identified. We tackle this problem by adopting a moment expansion approach under the multi-factor AJD setting. Two expansion methods, a Gram–Charlier and generalized Edgeworth expansions, are proposed to approximate $f$ and swaption prices are derived accordingly based on these methods.

### 2.2.3   The Gram–Charlier Expansion

As shown below, the Gram–Charlier expansion is obtained by using the inverse Fourier transform of the characteristic function and reordered as an orthogonalized series in additive form. We define the Chebyshev–Hermite polynomial as

$$H_n(x) := \frac{(-1)^n}{\phi(x)}D^n\phi(x), \quad H_0(x) = 1,$$

$$\phi(x) := \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),$$

where $D = \frac{d}{dx}$. By definition, the Chebyshev–Hermite polynomial is the following series:

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x,$$

$$H_4(x) = x^4 - 6x^2 + 3, \quad H_5(x) = x^5 - 10x^3 + 15x,$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15, \quad H_7(x) = x^7 - 21x^5 + 105x^3 - 105x.$$

The Chebyshev–Hermite polynomials have the orthogonal property with respect to the Gaussian measure.

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)\phi(x)dx = \delta_{mn}n!$$

where $\delta_{mn}$ is a Kronecker's symbol that returns 1 provided that $m = n$, 0 otherwise. As shown in equation (2.5) below, by using the properties of the Chebyshev–Hermite polynomials, the Gram–Charlier expansion is an orthogonal decomposition with $\{H_n\phi\}_n$ of a density function that has coefficients $q_n$, each of which depends on a given set of cumulants.

**Theorem 2.1** *(Tanaka, Yamada and Watanabe (2010)) Assume that a random variable $Y$ has density function $f$ and cumulants $c_k$ ($k \geq 1$), all of which are finite and known. Then the following hold:*

*(i) $f$ can be expanded with the Gram–Charlier expansion as*

$$f(x) = \sum_{k=0}^{\infty} \frac{q_k}{\sqrt{c_2}} H_k\left(\frac{x - c_1}{\sqrt{c_2}}\right)\phi\left(\frac{x - c_1}{\sqrt{c_2}}\right), \qquad (2.5)$$

*where*

$$q_k = \frac{1}{k!}\mathbb{E}\left[H_k\left(\frac{Y - c_1}{\sqrt{c_2}}\right)\right]$$

$$= \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k = 1, 2, \\ \sum_{m=1}^{[k/3]} \sum_{\substack{k_1+\cdots+k_m=k \\ k_i \geq 3}} \frac{c_{k_1}\cdots c_{k_m}}{m!k_1!\cdots k_m!}\left(\frac{1}{\sqrt{c_2}}\right)m^k, & \text{if } k \geq 3. \end{cases}$$

*(ii) for any $a \in \mathbb{R}$,*

$$\mathbb{E}[1_{\{Y>a\}}] = \Phi\left(\frac{c_1 - a}{\sqrt{c_2}}\right) + \sum_{k=3}^{\infty}(-1)^{k-1}q_k H_{k-1}\left(\frac{c_1 - a}{\sqrt{c_2}}\right)\phi\left(\frac{c_1 - a}{\sqrt{c_2}}\right),$$

$$\mathbb{E}[1_{\{Y>a\}}Y] = c_1\Phi\left(\frac{c_1 - a}{\sqrt{c_2}}\right) + \sqrt{c_2}\phi\left(\frac{c_1 - a}{\sqrt{c_2}}\right)$$
$$+ \sum_{k=3}^{\infty}(-1)^k q_k\left(-aH_{k-1}\left(\frac{c_1 - a}{\sqrt{c_2}}\right) + \sqrt{c_2}H_{k-2}\left(\frac{c_1 - a}{\sqrt{c_2}}\right)\right)\phi\left(\frac{c_1 - a}{\sqrt{c_2}}\right),$$

*where $\Phi(d)$ is the standard cumulative normal distribution evaluated at d.*

**Proof** (The proof of $(i)$) The characteristic function $\phi_Y$ of a random variable $Y$ is defined by the Fourier transform of $f$ as

$$\phi_Y(t) = \int_{-\infty}^{\infty} e^{itx}f(x)dx = e^{itc_1}\int_{-\infty}^{\infty} e^{i\sqrt{c_2}tx}\sqrt{c_2}f(c_1 + \sqrt{c_2}x)dx. \qquad (2.6)$$

On the other hand, by the definition of the cumulants, this can be expressed as

$$\phi_Y(t) = \exp\left(\sum_{k=1}^{\infty} \frac{c_k}{k!}(it)^k\right)$$

$$= e^{itc_1} \int_{-\infty}^{\infty} e^{i\sqrt{c_2}tx} \exp\left(\sum_{k=3}^{\infty} \frac{(-1)^k c_k}{k!}\left(\frac{D}{\sqrt{c_2}}\right)^k\right)\phi(x)dx. \qquad (2.7)$$

This is because, for any sequence $\{a_n\}$, it follows that

$$\exp\left(-\frac{c_2}{2}t^2 + \sum_{n=1}^{\infty} a_n(-i\sqrt{c_2}t)^n\right) = \int_{-\infty}^{\infty} e^{i\sqrt{c_2}tx} \exp\left(\sum_{n=1}^{\infty} a_n D^n\right)\phi(x)dx.$$

We further expand the integral of equation (2.7) by using the Taylor expansion and then reorder the terms as follows:

$$\exp\left(\sum_{k=3}^{\infty} \frac{(-1)^k c_k}{k!}\left(\frac{D}{\sqrt{c_2}}\right)^k\right)\phi(x)$$

$$= \left(1 + \sum_{m=1}^{\infty} \frac{1}{m!}\left(\sum_{k=3}^{\infty} \frac{(-1)^k c_k}{k!}\left(\frac{D}{\sqrt{c_2}}\right)^k\right)^m\right)\phi(x)$$

$$= \left(1 + \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{k_1,\cdots,k_m \geq 3} \frac{(-1)^{k_1+\cdots+k_m} c_{k_1}\cdots c_{k_m}}{k_1!\cdots k_m!}\left(\frac{D}{\sqrt{c_2}}\right)^{k_1+\cdots+k_m}\right)\phi(x)$$

$$= \left(1 + \sum{}^* \frac{c_{k_1}\cdots c_{k_m}}{m!k_1!\cdots k_m!}\left(\frac{1}{\sqrt{c_2}}\right)^n H_n(x)\right)\phi(x),$$

where $\sum{}^*$ means $\sum_{n=3}^{\infty} \sum_{m=1}^{[n/3]} \sum_{k_1+\cdots+k_m=n,k_i\geq 3}$. We use the following relationship

$$H_n(x)\phi(x) = (-1)^n D^n \phi(x),$$

in the last equality. Then, equation (2.7) can be written as

$$e^{itc_1} \int_{-\infty}^{\infty} e^{i\sqrt{c_2}tx}\phi(x)dx$$

$$+ e^{itc_1} \int_{-\infty}^{\infty} e^{i\sqrt{c_2}tx} \sum{}^* \frac{c_{k_1}\cdots c_{k_m}}{m!k_1!\cdots k_m!}\left(\frac{1}{\sqrt{c_2}}\right)^n H_n(x)\phi(x)dx. \qquad (2.8)$$

By using the inverse Fourier transforms of both equations (2.6) and (2.8), and by changing the relevant variable, we obtain the following Gram–Charlier expansion around the mean

$$f(x) = \frac{1}{\sqrt{c_2}}\phi\left(\frac{x-c_1}{\sqrt{c_2}}\right) + \frac{1}{\sqrt{c_2}}\sum{}^* \frac{c_{k_1}\cdots c_{k_m}}{m!k_1!\cdots k_m!}\left(\frac{1}{\sqrt{c_2}}\right)^n H_n\left(\frac{x-c_1}{\sqrt{c_2}}\right)\phi\left(\frac{x-c_1}{\sqrt{c_2}}\right).$$

The proof of (ii) is straightforward by using (i) and the properties of Chebyshev–Hermite polynomials. □

The advantage of the Gram–Charlier expansion is that it is written in additive form and hence the coefficients $q_n$ are easily expressed by the given cumulants as follows [1]

$$q_0 = 1, \quad q_1 = q_2 = 0, \quad q_3 = \frac{c_3}{3!c_2^{3/2}}, \quad q_4 = \frac{c_4}{4!c_2^2},$$

$$q_5 = \frac{c_5}{5!c_2^{5/2}}, \quad q_6 = \frac{c_6 + 10c_3^2}{6!c_2^3}, \quad q_7 = \frac{c_7 + 35c_3c_4}{7!c_2^{7/2}}. \tag{2.9}$$

The cumulants, $c_j$, can be calculated from the moments, $\mu_j$, around zero.[2]

Suppose that we know the $j$-th cumulant, $c_j$, of the underlying swap value at expiry $T_0$ under the $T_0$-forward measure that is associated with the option expiry with conditioned on $\mathcal{F}_t$. Then, the swaption value is obtained from equation (2.4) as

$$\begin{aligned}
SOV(t) &= P(t, T_0)\mathbb{E}_{\mathbb{P}_{T_0}}\left[1_{\{SV(T_0)>0\}}SV(T_0) \mid \mathcal{F}_t\right] \\
&= P(t, T_0)\bigg[c_1\Phi\left(\frac{c_1}{\sqrt{c_2}}\right) + \sqrt{c_2}\phi\left(\frac{c_1}{\sqrt{c_2}}\right) \\
&\quad + \sqrt{c_2}\phi\left(\frac{c_1}{\sqrt{c_2}}\right)\sum_{k=3}^{\infty}(-1)^k q_k H_{k-2}\left(\frac{c_1}{\sqrt{c_2}}\right)\bigg],
\end{aligned}$$

and therefore the truncated sum yields an approximation of the swaption value.

---

[1] In this context, it is well-known that $3!q_3$ represents skewness and $4!q_4$ represents excess kurtosis.

[2] See Stuart and Ord (1994). For example,

$$c_1 = \mu_1, \quad c_2 = \mu_2 - \mu_1^2, \quad c_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3,$$

$$c_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4,$$

$$c_5 = \mu_5 - 5\mu_1\mu_4 - 10\mu_2\mu_3 + 20\mu_1^2\mu_3 + 30\mu_1\mu_2^2 - 60\mu_1^3\mu_2 + 24\mu_1^5,$$

$$\begin{aligned}
c_6 = {}&\mu_6 - 6\mu_1\mu_5 - 15\mu_2\mu_4 + 30\mu_1^2\mu_4 - 10\mu_3^2 + 120\mu_1\mu_2\mu_3 - 120\mu_1^3\mu_3 \\
&+ 30\mu_2^3 - 270\mu_1^2\mu_2^2 + 360\mu_1^4\mu_2 - 120\mu_1^6,
\end{aligned}$$

$$\begin{aligned}
c_7 = {}&\mu_7 - 7\mu_1\mu_6 - 21\mu_2\mu_5 - 35\mu_3\mu_4 + 140\mu_1\mu_3^2 - 630\mu_1\mu_2^3 + 210\mu_1\mu_2\mu_4 \\
&- 1260\mu_1^2\mu_2\mu_3 + 42\mu_1^2\mu_5 + 2520\mu_1^3\mu_2^2 - 210\mu_1^3\mu_4 + 210\mu_2^2\mu_3 + 840\mu_1^4\mu_3 \\
&- 2520\mu_1^5\mu_2 + 720\mu_1^7.
\end{aligned}$$

**Proposition 2.3** (*Tanaka, Yamada and Watanabe (2010)*) *The swaption value is approximated as*

$$SOV(t) \approx P(t, T_0)\left[c_1\Phi\left(\frac{c_1}{\sqrt{c_2}}\right) + \sqrt{c_2}\phi\left(\frac{c_1}{\sqrt{c_2}}\right)\right.$$
$$\left. + \sqrt{c_2}\phi\left(\frac{c_1}{\sqrt{c_2}}\right)\sum_{k=3}^{L}(-1)^k q_k H_{k-2}\left(\frac{c_1}{\sqrt{c_2}}\right)\right]. \quad (2.10)$$

*We refer to this expression as the L-th order approximated price, $GC(L)$.*

From above, the calculation of the swaption is reduced to that of cumulants $c_j$ of the underlying swap. Since there exists a one-to-one correspondence between cumulants and moments, it is sufficient to calculate moments as given by,

$$\mathbb{E}_{\mathbb{P}_{T_0}}\left[(SV(T_0))^m \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{P}_{T_0}}\left[\left(\sum_{i=0}^{N} a_i P(T_0, T_i)\right)^m \Bigg| \mathcal{F}_t\right]$$
$$= \sum_{0 \le i_1, \dots, i_m \le N} a_{i_1} \cdots a_{i_m} \mu_{T_0}\left(t, T_0, \{T_{i_1}, \dots, T_{i_m}\}\right),$$

where $T_0$ is the expiry date of the swaption, $T_{i_1}, \dots, T_{i_m}$ are the coupon payment dates, and

$$\mu_T(t, T_0, \{T_1, \dots, T_m\}) := \mathbb{E}_{\mathbb{P}_T}\left[\prod_{i=1}^{m} P(T_0, T_i) \Bigg| \mathcal{F}_t\right], \quad (2.11)$$

which are called the bond moments by Collin-Dufresne and Goldstein (2002). As shown subsequently, $\mu_{T_0}(t, T_0, \{T_1, \dots, T_m\})$ can be calculated analytically for particular classes of interest rate models. In Section 2.4, numerical examples are presented with multi-factor interest rate models for which the bond moments are solvable either analytically or numerically.

It is worth clarifying the difference between the Gram–Charlier and Edgeworth expansions and their applications for swaption pricing. The seventh order Edgeworth expansion is equivalent to the seventh order Gram–Charlier expan-

sion, after an appropriate calculation.

$$
f(x) \approx \frac{1}{\sqrt{c_2}} \exp\left(\sum_{k=3}^{7} \frac{(-1)^k c_k}{k!} D^k\right) \phi\left(\frac{x - c_1}{\sqrt{c_2}}\right)
$$

$$
\approx \frac{1}{\sqrt{c_2}} \left(1 + \sum_{k=3}^{7} \frac{(-1)^k c_k}{k!} D^k + \frac{1}{2}\left(\left(\frac{c_3}{3!} D^3\right)^2 - 2\frac{c_3}{3!}\frac{c_4}{4!} D^3 D^4\right)\right) \phi\left(\frac{x - c_1}{\sqrt{c_2}}\right)
$$

$$
= \frac{1}{\sqrt{c_2}} \left(1 + q_3 H_3 + q_4 H_4 + q_5 H_5 + q_6 H_6 + q_7 H_7\right) \phi\left(\frac{x - c_1}{\sqrt{c_2}}\right),
$$

where $H_k = H_k\left(\frac{x - c_1}{\sqrt{c_2}}\right)$. A swaption is equivalent to an option on a coupon-bearing bond. In existing studies, the option on a coupon-bearing bond is decomposed into weighted cash-flows based on the exercise probabilities under the forward measures associated with the cash-flow timing,

$$
SOV(t) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^{T_0} r_u \, du}(CB(T_0) - K)\mathbf{1}_{\{CB(T_0) > K\}} \mid \mathcal{F}_t\right]
$$

$$
= \sum_{i=1}^{N} C_i P(t, T_i) \mathbb{P}_{\mathbb{P}_{T_i}}\left(CB(T_0) > K \mid \mathcal{F}_t\right)
$$

$$
- KP(t, T_0) \mathbb{P}_{\mathbb{P}_{T_0}}\left(CB(T_0) > K \mid \mathcal{F}_t\right), \qquad (2.12)
$$

$$
CB(T_0) := \sum_{i=1}^{N} C_i P(T_0, T_i),
$$

where $C_i$ is the cash-flow at $T_i$, and $K$ is the strike price. When calculating the probability of ending up in-the-money under the forward measure, Collin-Dufresne and Goldstein (2002) used a seventh-order Edgeworth expansion. In their approach, the probability of ending up in-the-money under each forward measure can be approximated with proper functions $\lambda(c_1, c_2)$ and $\gamma(c_1, \ldots, c_7)$.

$$
\mathbb{P}_{\mathbb{P}_{T_i}}\left(CB(T_0) > K \mid \mathcal{F}_t\right) \approx \sum_{j=1}^{7} \lambda_j \gamma_j \quad i = 0, 1, \ldots, N.
$$

See Collin-Dufresne and Goldstein (2002) for more details. On the other hand, Tanaka, Yamada and Watanabe (2010) applied the Gram–Charlier expansion and derived equation (2.10). Since each probability is expanded in Collin-Dufresne and Goldstein (2002) as in equation (2.12), and a swaption price is expressed with the sum of probabilities, cumulative truncation error would be bigger than the truncation error in equation (2.10) where a swaption price is expressed with one expectation.

## 2.2.4   The Generalized Edgeworth Expansion

In this subsection, another approximation formula is proposed using a generalized Edgeworth expansion. Standard normal distributions are used in the Gram–Charlier expansion and its key strength, roughly speaking, lies in the ease of computing higher-order derivatives of normal distributions. If the density of a target deviates from normality, which is the case where jump terms are included in a model, the accuracy of the Gram–Charlier expansion deteriorates. An arbitrary distribution can be used in the generalized Edgeworth expansion. In the generalized Edgeworth expansion, a target density is approximated with an arbitrary density and their cumulants, and accordingly an option on an underlying variable is expressed in terms of options on another variable and adjustment terms. We propose a method to approximate swaption prices based on this expansion. Since the expansion has additive form, we arrive at a decomposition for swaption prices. By adopting the tradable instrument as an approximation variable, the decomposition is indicative of a new hedging strategy using a tradable instrument. The coefficients in the decomposition are simple functions of the cumulants of the target and approximation variables.

The generalized Edgeworth expansion was used to approximate stock option prices in Jarrow and Rudd (1982). It approximates the distribution of the stock price at option expiry driven by a jump diffusion, or a constant elasticity of variance diffusion, with the Black–Scholes type lognormal distribution. Turnbull and Wakeman (1991) used a lognormal distribution to approximate the sum of lognormally distributed variables and derive an approximation for the price of an Asian option on a stock.

**Theorem 2.2** *Assume that random variables $F$ and $G$ have density functions $f$ and $g$ respectively and have cumulants $c_k(f)$ and $c_k(g)$ $(k \geq 1)$ respectively, all of which are finite and known. Then the following hold:*

*(i) (Jarrow and Rudd (1982)) $f$ can be expanded with the generalized Edgeworth expansion as*

$$f(x) = \sum_{k=0}^{\infty} \frac{(-1)^k I_k}{k!} D^k g(x),$$

29

*where $I_0 = 1$, $I_1 = c_1(f) - c_1(g)$, $I_2 = c_2(f) - c_2(g) + (c_1(f) - c_1(g))^2$, $I_3 = c_3(f) - c_3(g) + 3(c_1(f) - c_1(g))(c_2(f) - c_2(g)) + (c_1(f) - c_1(g))^2, \ldots$, and $D$ is the differential operator, $d/dx$.*

(ii) *Assume $\lim_{x \to \infty} x D^k g(x) = 0$, for any non-negative integer $k$. Then, for any $a \in \mathbb{R}$,*

$$\mathbb{E}[1_{\{F>a\}}F] = \mathbb{E}[1_{\{G>a\}}G] - I_1(-ag(a) - \mathbb{E}[1_{\{G>a\}}]) + \frac{I_2}{2!}(-ag'(a) + g(a))$$
$$+ \sum_{k=3}^{\infty} \frac{(-1)^k I_k}{k!}(-aD^{k-1}g(a) + D^{k-2}g(a)). \quad (2.13)$$

**Proof** (The proof of $(i)$) By the definition of cumulants, the characteristic function $\phi_F$ of a random variable $F$ is expressed with the $k$-th order cumulants $c_k(f)$ as

$$\phi_F(t) = \exp\left(\sum_{k=1}^{\infty} \frac{c_k(f)}{k!}(it)^k\right).$$

The analogous expression holds for a random variable $G$ with $c_k(g)$. Moreover,

$$\log \frac{\phi_F(t)}{\phi_G(t)} = \sum_{k=1}^{\infty} \frac{c_k(f) - c_k(g)}{k!}(it)^k.$$

Expand both sides and use the Taylor expansion:

$$\phi_F(t)$$
$$= \exp\left(\sum_{k=1}^{\infty} \frac{c_k(f) - c_k(g)}{k!}(it)^k\right)\phi_G(t)$$
$$= \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!}\left(\sum_{k=1}^{\infty} \frac{c_k(f) - c_k(g)}{k!}(it)^k\right)^n\right)\phi_G(t)$$
$$= \left(1 + (c_1(f) - c_1(g))(it) + \frac{c_2(f) - c_2(g) + (c_1(f) - c_1(g))^2}{2!}(it)^2 + \cdots\right)\phi_G(t)$$
$$= \sum_{k=0}^{\infty} \frac{I_k}{k!}(it)^k \phi_G(t),$$

where $I_0 = 1$, $I_1 = c_1(f) - c_1(g)$, $I_2 = c_2(f) - c_2(g) + (c_1(f) - c_1(g))^2$, $I_3 = c_3(f) - c_3(g) + 3(c_1(f) - c_1(g))(c_2(f) - c_2(g)) + (c_1(f) - c_1(g))^2, \ldots$. Using the

inverse Fourier expansion on both sides, we obtain

$$f(x) = \sum_{k=0}^{\infty} \frac{I_k}{k!} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx}(it)^k \phi_G(t)dt$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k I_k}{k!} D^k g(x).$$

(The proof of $(ii)$) The truncated expectation with respect to $F$ can be computed with a random variable $G$ as

$$\mathbb{E}\big[1_{\{F>a\}}F\big] = \int_a^{\infty} xf(x)\,dx = \int_a^{\infty} x \sum_{k=0}^{\infty} \frac{(-1)^k I_k}{k!} D^k g(x)\,dx.$$

Apply the integration by parts formula to give

$$\mathbb{E}\big[1_{\{F>a\}}F\big] = \int_a^{\infty} xg(x)\,dx - I_1\left(xg(x)\big|_a^{\infty} - \int_a^{\infty} g(x)\,dx\right)$$

$$+ \sum_{k=2}^{\infty} \frac{(-1)^k I_k}{k!}\left(xD^{k-1}g(x)\big|_a^{\infty} - \int_a^{\infty} D^{k-1}g(x)\,dx\right).$$

Assume $\lim_{x\to\infty} xD^k g(x) = 0$ for any non-negative integer $k$,

$$\mathbb{E}\big[1_{\{F>a\}}F\big] = \mathbb{E}\big[1_{\{G>a\}}G\big] - I_1\left(-ag(a) - \mathbb{E}\big[1_{\{G>a\}}\big]\right)$$

$$+ \sum_{k=2}^{\infty} \frac{(-1)^k I_k}{k!}\left(-aD^{k-1}g(a) + D^{k-2}g(a)\right).$$

$\square$

The target density is expanded with a linear combination of $g(x)$ and its derivatives. Higher-order terms adjust the gap between $f(x)$ and $g(x)$. In fact, the Gram–Charlier and generalized Edgeworth expansions are not always distinct. The generalized Edgeworth expansion is the same as the Gram–Charlier expansion provided that $g(x)$ follows the standard normal density, $g(x) = \phi\big((x - c_1(f))/\sqrt{c_2(f)}\big)$.

The choice of $G$ plays a decisive role in determining the accuracy and tractability of the approximation method. We explain how to choose $G$ in the case of a receiver's swaption. Let $F$ be a receiver's swap value at option expiry $T_0$, and accordingly let $f$ be the density function of it under the $T_0$-forward measure conditioned on $\mathcal{F}_t$. Considering the right-hand-side of equation (2.13), a random

variable $G$ that is tractable for calculations is preferable. Based on the observation that the swap value largely depends on the random variable $P(T_0, T_N)$ because its coefficient is the biggest in equation (2.1), i.e. $\delta\kappa \ll 1$, we define $G$ as containing only one random variable, $P(T_0, T_N)$, and as being a linear function of it, that is to say, $G = N_1 P(T_0, T_N) - N_2$ where $N_1$ and $N_2$ are assumed constant. $N_1$ and $N_2$ are determined to equate the first and second order moments of $G$ to those of $F$, respectively. In other words, $I_1 = 0$ and $I_2 = 0$ in equation (2.13). As a result, $G$ is defined as

$$G := N_1 P(T_0, T_N) - N_2, \tag{2.14}$$

$$N_1 := \sqrt{\frac{c_2(f)}{c_2\left(P(T_0, T_N)\right)}}, \tag{2.15}$$

$$N_2 := N_1 c_1\left(P(T_0, T_N)\right) - c_1(f), \tag{2.16}$$

where $c_1(P(T_0, T_N))$ and $c_2(P(T_0, T_N))$ are the first and second order cumulants of $P(T_0, T_N)$ under $\mathbb{P}_{T_0}$ conditioned on $\mathcal{F}_t$, respectively.

By substituting $G$ into equation (2.13), the receiver's swaption price is obtained as

$$
\begin{aligned}
SOV(t) &= P(t, T_0)\mathbb{E}_{\mathbb{P}_{T_0}}[1_{\{SV(T_0)>0\}}SV(T_0) \mid \mathcal{F}_t] \\
&= P(t, T_0)\Big(\mathbb{E}_{\mathbb{P}_{T_0}}\big[1_{\{N_1 P(T_0,T_N)-N_2>0\}}(N_1 P(T_0, T_N) - N_2) \mid \mathcal{F}_t\big] \\
&\qquad\qquad + \sum_{k=3}^{\infty} \frac{(-1)^k I_k}{k!} D^{k-2} g(0)\Big) \\
&= P(t, T_0)\Big(N_1 \mathbb{E}_{\mathbb{P}_{T_0}}\Big[1_{\{P(T_0,T_N)>\frac{N_2}{N_1}\}}\Big(P(T_0, T_N) - \frac{N_2}{N_1}\Big)\mid \mathcal{F}_t\Big] \\
&\qquad\qquad + \sum_{k=3}^{\infty} \frac{(-1)^k I_k}{k!} D^{k-2} g(0)\Big), \tag{2.17}
\end{aligned}
$$

where $g$ is the density function of $G$ under $\mathbb{P}_{T_0}$. Note that the first term in equation (2.17) corresponds to $N_1$ contracts of the call option on $P(T_0, T_N)$, and its strike price is $N_2/N_1$. Since neither the call option value nor the higher-order derivatives of the density have closed-form solutions for the multi-factor models we consider, we approximate them with the Gram–Charlier expansion. To apply the Gram–Charlier expansion to evaluate them, we need the cumulants of $G$ and

they are calculated from the moments of $G$, where the $m$-th moment is

$$\mathbb{E}_{\mathbb{P}_{T_0}}[G^m \,|\, \mathcal{F}_t] = \mathbb{E}_{\mathbb{P}_{T_0}}\left[ (N_1 P(T_0, T_N) - N_2)^m \,|\, \mathcal{F}_t \right]$$

$$= \sum_{\substack{m_1 + m_2 = m \\ m_1, m_2 \geq 0}} \binom{m}{m_1} N_1^{m_1} (-N_2)^{m_2} \mathbb{E}_{\mathbb{P}_{T_0}}[P(T_0, T_N)^{m_1} \,|\, \mathcal{F}_t].$$

Since the moments of $G$ are expressed with the bond moments, (2.11), calculating the call option price and higher-order derivatives is reduced to calculating the bond moments.

The truncated sum in equation (2.17) yields an approximation of the swaption value and we have the following Proposition.

**Proposition 2.4** *The receiver's swaption value is approximated by*

$$SOV(t) \approx C(A) + P(t, T_0) \sum_{k=3}^{L} \frac{(-1)^k I_k}{k!} D^{k-2} g(0). \tag{2.18}$$

*where*

$$C(A) := P(t, T_0) N_1 \mathbb{E}_{\mathbb{P}_{T_0}}\left[ 1_{\{P(T_0, T_N) > \frac{N_2}{N_1}\}} \left( P(T_0, T_N) - \frac{N_2}{N_1} \right) \Big| \mathcal{F}_t \right]$$

$$\approx P(t, T_0) \left[ c_1(g) \Phi\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right) + \sqrt{c_2(g)} \phi\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right) \right.$$

$$\left. + \sum_{k=3}^{L_0} (-1)^k q_k \sqrt{c_2(g)} H_{k-2}\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right) \phi\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right) \right].$$

$C(A)$ *represents the $N_1$ contracts of the call options evaluated at $t$ whose underlying variable is $P(T_0, T_N)$ and strike price is $N_2/N_1$. $D^i g(0)$ is the $i$-th order derivative of the density function of $G$ evaluated at $0$ under $\mathbb{P}_{T_0}$.*

$$D^i g(0) \approx \sum_{k=0}^{L_i} \frac{(-1)^k q_k}{\left( \sqrt{c_2(g)} \right)^{i+1}} H_{k+i}\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right) \phi\left( \frac{c_1(g)}{\sqrt{c_2(g)}} \right).$$

$I_i$, $i \geq 3$, *are expressed with cumulants;*

$$I_3 = c_3(f) - c_3(g), \quad I_4 = c_4(f) - c_4(g) + 3(c_2(f) - c_3(g))^2, \ldots,$$

*where $c_m(f)$ is the $m$-th order cumulant of the swap value which is a function of up to the $m$-th order moments of it, i.e.*

$$\mathbb{E}_{\mathbb{P}_{T_0}}[F^m \,|\, \mathcal{F}_t] = \mathbb{E}_{\mathbb{P}_{T_0}}\left[ \left( -1 + \delta \kappa \sum_{i=1}^{N} P(T_0, T_i) + P(T_0, T_N) \right)^m \Big| \mathcal{F}_t \right],$$

and $c_m(g)$ is the m-th order cumulants of G which is a function of up to the m-th order moments of it:

$$\mathbb{E}_{\mathbb{P}_{T_0}}[G^m \,|\, \mathcal{F}_t] = \mathbb{E}_{\mathbb{P}_{T_0}}[(N_1 P(T_0, T_N) - N_2)^m \,|\, \mathcal{F}_t].$$

Finally set $q_k = q_k(c_1(g), \ldots, c_k(g))$.

**Remark 2.1** *For a payer's swaption, G is defined as $N_2 - N_1 P(T_0, T_N)$. The payer's swaption value is, consequently, decomposed into the put options on the zero-coupon bond and higher-order derivatives of G.*

We refer to this expression in which a swaption price is expanded up to the L-th order, the call option up to the $L_0$-th order and i-th order derivative up to the $L_i$-th order as the $(L, L_0, L_1 \ldots, L_{L-2})$-th order approximation price, $\text{GEW}(L, L_0, L_1, \cdots, L_{L-2})$. While calculation of the swaption price with the Gram–Charlier expansion is reduced to that of the cumulants of the underlying swap value, with the generalized Edgeworth expansion it is reduced to that of the cumulants of the underlying swap value and that of the cumulants of the approximating variable.

Proposition 2.4 represents that the receiver's swaption is decomposed into the call options on the zero-coupon bond and the higher-order derivatives of G. Hence, it turns out a swaption is partially hedged with call options and this decomposition is model-free. Moreover, the ATMF swaption is decomposed into the ATM call options and the higher-order derivatives since in this case, from equations (2.15) and (2.16), the strike price of the call coincides with the forward price of the underlying zero-coupon bond;

$$\frac{N_2^{ATMF}}{N_1^{ATMF}} = c_1\big(P(T_0, T_N)\big) = \mathbb{E}_{\mathbb{P}_{T_0}}[P(T_0, T_N) \,|\, \mathcal{F}_t] = \frac{P(t, T_N)}{P(t, T_0)}.$$

Similarly, the ITM and OTM swaptions are expressed with the ITM and OTM calls respectively. We carry out Monte Carlo studies later, and examine the proximity between a swaption price and a call option price for some specific models.

Several ways have been suggested to compute a call option price for multi-factor models. One of them is to make use of the Fourier inversion theorem with

characteristic functions. Studies of computing probabilities from characteristic functions include Heston (1993), Scott (1997) and Pan (2002). Carr and Madan (1999) found an alternative representation for the European call option price with Fourier inversion. Lee (2004) conducted an in-depth discussion of Carr and Madan's approach and proposed an algorithm to minimize truncation and discretization errors in computing Fourier inversions numerically. A drawback of this approach is, however, that the optimal discretization step and integration interval depend not only on a model and a payoff function but also moneyness for high-precision numerical integration. Implementation is therefore difficult. Another approach to price an option is using saddlepoint approximations proposed by Glasserman and Kim (2009). They applied the saddlepoint approximation method and proposed its improvement in pricing options.

## 2.3  Affine Jump-Diffusion Models

Our methods for approximating swaptions are independent of any model if bond moments can be obtained. In this section, we review the basics of the multi-factor AJD process and its term structure. Multi-factor models are increasingly popular with practitioners as they can describe various types of yield curve shifts. We assume the short rate is of the form

$$r_t = \rho_0 + \rho_1^\top X_t, \quad \rho_0 \in \mathbb{R}, \quad \rho_1 \in \mathbb{R}^n,$$

where $X$ is assumed to be an $n \times 1$ vector Markov process under $\mathbb{Q}$ satisfying a stochastic differential equation,

$$dX_t = K(\theta - X_t)\,dt + \Sigma D(X_t)\,dW_t^\mathbb{Q} + dZ_t^\mathbb{Q}, \tag{2.19}$$

where

$$K \in \mathbb{R}^{n \times n}, \quad \theta \in \mathbb{R}^n, \quad p_i \in \mathbb{R}, \quad q_i \in \mathbb{R}^n, \quad i = 1, 2, \ldots, n,$$

$$D(x) = \operatorname{diag}\left[\sqrt{p_1 + q_1^\top x}, \cdots, \sqrt{p_n + q_n^\top x}\right], \quad x \in \mathbb{R}^n,$$

and $\Sigma \in \mathbb{R}^{n \times n}$ is a matrix such that $\Sigma\Sigma^\top$ is positive definite. $W^\mathbb{Q}$ is an $n$-dimensional standard Brownian motion and $Z^\mathbb{Q}$ is a pure jump process whose jumps have a fixed probability distribution $\nu$ on $\mathbb{R}^n$ and time-$t$ arrival intensity

$\lambda(X_t)$. We assume the successive jumps of $Z^{\mathbb{Q}}$ are non-negative, independent and identically distributed. Note that the stochastic integral with respect to the pure jump process is a finite sum and equation (2.19) is equivalent to

$$X_t = X_0 + \int_0^t K(\theta - X_s)\, ds + \int_0^t \Sigma D(X_s)\, dW_s^{\mathbb{Q}} + \sum_{i=1}^{N_t^{\mathbb{Q}}} J_i^{\mathbb{Q}},$$

where $N_t^{\mathbb{Q}}$ is a counting process expressing the number of jumps having occurred up to time $t$ and has arrival intensity $\lambda(X_t)$. $J_i^{\mathbb{Q}}$ follows the distribution $\nu$. The functional form of $\lambda(X_t)$ is affine with coefficients:

$$\lambda(x) = l_0 + l_1^\top x, \quad l_0 \in \mathbb{R}^n, \quad l_1 \in \mathbb{R}^{n \times n}.$$

To apply the moment expansions methods to pricing swaptions, Fourier-type transforms need to be calculated and is given in Propositon 1 in Duffie, Pan and Singleton (2000).

**Proposition 2.5** *(Propositon 1 in Duffie, Pan and Singleton (2000)) Suppose* $(K, \theta, l, \theta(c), \rho)$ *in* $X$ *are well-behaved, then*

$$\mathbb{E}\left[\exp\left(-\int_t^T r(X_s)ds\right)e^{u^\top X_T}\Big|\mathcal{F}_t\right] = e^{\alpha(t,T)+\beta(t,T)^\top X_t}.$$

$\alpha(t,T)$ *and* $\beta(t,T)$ *satisfy the following ODEs:*

$$\frac{d}{dt}\alpha(t,T) = -(K\theta)^\top \beta(t,T) - \frac{1}{2}\sum_{j=1}^n (\Sigma^\top \beta(t,T))_j^2 p_j - l_0\big(\theta(\beta(t,T)) - 1\big) + \rho_0,$$

$$\tag{2.20}$$

$$\frac{d}{dt}\beta(t,T) = K^\top \beta(t,T) - \frac{1}{2}\sum_{j=1}^n (\Sigma^\top \beta(t,T))_j^2 q_j - l_1\big(\theta(\beta(t,T)) - 1\big) + \rho_1,$$

$$\tag{2.21}$$

*where* $\theta(c)$ *is the jump transform defined as:*

$$\theta(c) = \int_{\mathbb{R}^n} e^{cz}\, d\nu(z).$$

*The terminal conditions are* $\alpha(T,T) = 0$ *and* $\beta(T,T) = u$.

See Duffie, Pan and Singleton (2000) for the proof. Proposition 2.5 holds under the following technical conditions,

**Definition 2.7** *A characteristic* $(K, \theta, l, \theta(c), \rho)$ *is well-behaved if equations (2.20) and (2.21) are solved uniquely by $\alpha$ and $\beta$ and if*

*1.* $\mathbb{E}\left[\int_0^T |\gamma_t| \, dt\right] < \infty$, *where* $\gamma_t = \Psi_t(\theta(\beta(t,T)) - 1)\lambda(X_t)$,

*2.* $\mathbb{E}\left[\left(\int_0^T \eta_t \cdot \eta_t \, dt\right)^{1/2}\right] < \infty$, *where* $\eta_t = \Psi_t \beta(t,T)^\top \Sigma D(X_t)$, *and*

*3.* $\mathbb{E}[|\Psi_T|] < \infty$,

*where* $\Psi_t = \exp\left(-\int_0^t r(X_s) \, ds\right) e^{\alpha(t,T) + \beta(t,T)^\top X_t}$.

The time-$t$ price of a zero-coupon bond maturing at $T$ is expressed in the form of an exponentially affine function,

$$P(t,T) = \exp(A(t,T) + B(t,T)^\top X_t),$$

of a vector of state variables. Proposition 2.5 yields the following system of ODEs with respect to $A(t,T)$ and $B(t,T)$:

$$\frac{d}{dt}A(t,T) = -(K\theta)^\top B(t,T) - \frac{1}{2}\sum_{j=1}^n (\Sigma^\top B(t,T))_j^2 p_j - l_0\left(\theta(B(t,T)) - 1\right) + \rho_0,$$

$$(2.22)$$

$$\frac{d}{dt}B(t,T) = K^\top B(t,T) - \frac{1}{2}\sum_{j=1}^n (\Sigma^\top B(t,T))_j^2 q_j - l_1\left(\theta(B(t,T)) - 1\right) + \rho_1.$$

$$(2.23)$$

The terminal conditions are

$$A(T,T) = 0$$

and

$$B(T,T) = 0. \tag{2.24}$$

This system can be solved in closed-form for special cases such as the model with Gaussian-type volatility, an exponentially distributed jump size and constant jump intensity or the model with CIR-type volatility, an exponentially distributed jump size and constant jump intensity, while it can be solved numerically in many other cases.

Similar to a zero-coupon bond price, the bond moments are expressed in exponentially affine form by applying Proposition 2.5:

$$
\mu_T\big(t, T_0, \{T_1, \ldots, T_m\}\big)
$$

$$
= \mathbb{E}_{\mathbb{P}_T}\left[\prod_{i=1}^{m} P(T_0, T_i)\,\big|\,\mathcal{F}_t\right]
$$

$$
= \frac{1}{P(t,T)}\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\prod_{i=1}^{m} P(T_0, T_i)\,\big|\,\mathcal{F}_t\right]
$$

$$
= \frac{1}{P(t,T)}\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^{T_0} r_u\,du}P(T_0, T)\prod_{i=1}^{m} P(T_0, T_i)\,\big|\,\mathcal{F}_t\right]
$$

$$
= \frac{1}{P(t,T)}\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^{T_0} r_u\,du}e^{\sum_{i=1}^{m} A(T_0,T_i)+A(T_0,T)+(\sum_{i=1}^{m} B(T_0,T_i)+B(T_0,T))^\top X_{T_0}}\,\big|\,\mathcal{F}_t\right]
$$

$$
= \frac{\exp(\bar{A}(t) + \bar{B}(t)^\top X_t)}{P(t,T)},
$$

where $\bar{A}(t) := \bar{A}(t, T, T_0, \{T_1, \ldots, T_m\})$ and $\bar{B}(t) := \bar{B}(t, T, T_0, \{T_1, \ldots, T_m\})$ satisfy the same system of ODEs as (2.22) and (2.23), but with different terminal conditions,

$$
\bar{A}(T_0) = \sum_{i=1}^{m} A(T_0, T_i) + A(T_0, T)
$$

and

$$
\bar{B}(T_0) = \sum_{i=1}^{m} B(T_0, T_i) + B(T_0, T). \tag{2.25}
$$

These equations are derived by Collin-Dufresne and Goldstein (2002) in the case of the affine diffusion model with the Feynman–Kac formula.

## 2.4 Numerical Examples

In this section, we carry out numerical studies using AJDs. AJDs governed either by three-factor Gaussian- or two-factor CIR-type volatilities are considered with exponential, normal or log-normal jumps. We compute the receiver's swaption prices for various strikes by using the Gram–Charlier and generalized Edgeworth expansions and compare them with prices calculated by Monte Carlo simulation with respect to accuracy and computational burden. Two approximation methods based on the Gram–Charlier expansion are considered; the

first three moments, GC(3), and the first seven moments, GC(7). In addition, two approximation methods based on the generalized Edgeworth expansion are considered; the first two moments, GEW(2,7), and the first three moments, GEW(3,7,7). With the GEW(2, 7) method, the swaption price is approximated only with the call options and the call option is approximated with the Gram–Charlier expansion up to the seventh moment. With the GEW(3, 7, 7) method, the swaption price is approximated with the call options and the first order derivative of $G$ and both are approximated with the Gram–Charlier expansion up to the seventh moment.

We use the one-into-ten swaption, the swaption expiring one year later and the underlying swap maturity is ten years, as the target. Strike rates are set from ATMF$-2.0\%$ to ATMF$+2.0\%$ in steps of $0.01\%$ and ATMF is defined in equation (2.2). The payment frequency is set semi-annual ($\delta = 1/2$). Throughout these examples, the jump intensity is assumed constant.

### 2.4.1 Monte Carlo Simulation

To confirm the accuracy of the proposed methods, we compare the swaption prices calculated by the moment expansion methods with those computed by Monte Carlo simulation. We explain how to apply Monte Carlo simulation for pricing swaptions in this subsection. In problems such as pricing zero-coupon bonds, expectations under the risk-neutral measure are easy to compute. In other cases the calculation under the risk-neutral measure might be complex and calculation under a forward measure might make problems easier. This is the case for pricing swaptions. Pricing swaptions under a forward measure needs the underlying process changed to the process under the measure. The Radon-Nikodým derivative, when restricted to $\mathcal{F}_t$, satisfies, for every $t \in [0, T_0]$,

$$\frac{d\mathbb{P}_{T_0}}{d\mathbb{Q}}\bigg|_{\mathcal{F}_t} = \frac{P(t,T)}{D_t P(0,T)}.$$

The change of measure to $\mathbb{P}_{T_0}$ from $\mathbb{Q}$ associates a Brownian motion and pure jump process with

$$dW_t^{T_0} = dW_t^{\mathbb{Q}} - \left((B(t,T_0)^\top \Sigma D(X_t))\right)^\top dt,$$
$$dN_t^{T_0} = dN_t^{\mathbb{Q}} - \theta\big(B(t,T_0)\big)\lambda\, dt.$$

39

Thus the dynamics of $X$ under the $T_0$-forward measure are given by

$$X_t = X_0 + \int_0^t \left( K\theta - KX_s + \Sigma D(X_s)D(X_s)^\top \Sigma^\top B(s, T_0) \right) ds$$

$$+ \int_0^t \Sigma D(X_s)\, dW_s^{T_0} + \sum_{i=1}^{N_t^{T_0}} J_i^{T_0},$$

where the jump intensity and jump transform under the $T_0$-forward measure are given, respectively, by

$$\lambda^{T_0} = \lambda\theta\big(B(t, T_0)\big), \tag{2.26}$$

$$\theta^{T_0}(c) = \frac{\theta\big(c + B(t, T_0)\big)}{\theta\big(B(t, T_0)\big)}. \tag{2.27}$$

Hence although constant intensity and independent jump-size are assumed under the risk-neutral measure, they are time-dependent under the $T_0$-forward measure. For the change of measure technique of a jump diffusion process, see Appendix C in Duffie, Pan and Singleton (2000).

For the negative jumps, the new jump term $-d\tilde{Z}_t$ is added to equation (2.19). The corresponding jump intensity and its transform under the forward measure are respectively written as

$$\lambda^{T_0} = \lambda\theta\big(-B(t, T_0)\big),$$

$$\theta^{T_0}(c) = \frac{\theta\big(c - B(t, T_0)\big)}{\theta\big(-B(t, T_0)\big)}.$$

A Monte Carlo algorithm to price swaptions using the forward measure technique is now summarized.

**Step 1:** Generate the underlying process at option expiry, $X_{T_0}$.

> **Step 1.1:** Simulate jump times. $j$-th jump time of $i$-th factor, $\tau_{i,j}$, $j \geq 1$, $i = 1, 2, \ldots, n$ is simulated with an exponentially distributed random variable $\xi$ whose mean is one.
>
> $$\tau_{i,j+1} = \inf\left\{ s \geq 0 : \int_{\tau_{i,j}}^s \lambda_i^{T_0}(u)\, du = \xi \right\}, \quad j \geq 0, \quad \tau_{i,0} = t. \tag{2.28}$$
>
> $\tau_{i,1} > T_0$ indicates the $i$-th factor has no jump until option expiry. The integral $\int \lambda_i^{T_0}(u)\, du$ can be solved analytically for some models.

**Step 1.2:** Between each jump, $\tau_{i,j+1} - \tau_{i,j}$, $j \geq 0$, and between the last jump and option maturity, the underlying process follows diffusion,

$$X_{\tau_{i,j+1-}} = X_{\tau_{i,j}} + \int_{\tau_{i,j}}^{\tau_{i,j+1}} \left( K\theta - KX_s + \Sigma D(X_s)D(X_s)^\top \Sigma^\top B(s,T_0) \right) ds$$
$$+ \int_{\tau_{i,j}}^{\tau_{i,j+1}} \Sigma D(X_s) \, dW_s^{T_0}.$$

For Gaussian- and CIR-type volatilities, the process between $\tau_{i,j}$ and $\tau_{i,j+1}$ can be simulated without discretizing the time step. The former follows the normal distribution and the latter the non-central chi-squared distribution. See for example, Sections 2.3 and 3.4 in Glasserman (2004).

**Step 1.3:** At jump time $\tau_{i,j}$, the $i$-th factor jumps additively.

$$X_{\tau_{i,j}} = X_{\tau_{i,j-}} + \Delta Z,$$

where $\Delta Z$ follows the jump size distribution changed into the one under the $T_0$-forward measure, $\nu_i^{T_0}(x)$.

**Step 2:** Calculate the swap value at option expiry with simulated value $X_{T_0}$,

$$SV(T_0) = -1 + \delta\kappa \sum_{i=1}^{N} P(T_0, T_i) + P(T_0, T_N),$$

where $P(T_0, T_i) = \exp(A(T_0, T_i) + B(T_0, T_i)^\top X_{T_0})$, $i = 1, 2, \ldots, N$. The components $A(T_0, T_i)$ and $B(T_0, T_i)$ are the solutions of the ODE system (2.22), (2.23) and (2.24).

**Step 3:** Repeat Steps 1 and 2 independently $M$ times so that independent samples $SV_i$, $i = 1, 2, \ldots, M$ are obtained. The swaption price is calculated by the experimental average with discount,

$$SOV^{MC} = \frac{P(t, T_0)}{M} \sum_{i=1}^{M} 1_{\{SV_i(T_0) \geq 0\}} SV_i(T_0).$$

## 2.4.2 Jump Specifications for Affine Jump-Diffusion models

We consider three types of jump size distribution: exponential, normal and truncated normal. Obtaining closed-form solutions or numerical solutions for

the systems of ODEs depends on the assumption regarding the volatility and distributional characteristics of the jump size.

An exponentially distributed jump size is widely used in jump-diffusion models to express skewness in interest rates. For example, Das and Foresi (1996) derived the zero-coupon bond price for the one-factor Gaussian model enhanced with jump term. In their model, the jump size follows an exponential distribution and the jump sign is determined from a Bernoulli distribution. Chacko and Das (2002) combined the exponentially distributed jump size in the Gaussian short rate model and derived option prices for several types of payoff functions; an option on the interest rate, an option on a zero-coupon bond and an Asian option on the interest rate. The density function and the jump transform are defined, respectively, as

$$\nu(x) = \frac{1}{\eta} e^{-\eta x} 1_{\{x \geq 0\}},$$
$$\theta(c) = \int_0^\infty e^{cx} \, d\nu(x) = \frac{1}{1 - \eta c}, \tag{2.29}$$

where $\eta$ is the mean of an exponential distribution. Since exponential distributions take only positive values, they represent asymmetric jumps. In order to express negative jumps, we introduce another jump process which has a negative sign.

The second example is a normal distribution. While a jump transform has a closed-form solution, the system of ODEs does not. Numerical approaches are applied to solve the ODEs. Baz and Das (1996) and Durham (2006) used the Taylor expansion to approximate the jump transform and derived the ODEs which can be calculated without numerical integration. The density and jump transform of a normally distributed jump size are defined, respectively, as

$$\nu(x) = \frac{1}{\sigma_J} \phi\left(\frac{x - \mu_J}{\sigma_J}\right) = \frac{1}{\sqrt{2\pi}\sigma_J} \exp\left(\frac{-(x - \mu_J)^2}{2\sigma_J^2}\right),$$
$$\theta(c) = \int_{-\infty}^\infty e^{cx} \, d\nu(x) = \exp\left(c\mu_J + \frac{c^2\sigma_J^2}{2}\right), \tag{2.30}$$

where $\mu_J$ and $\sigma_J^2$ are the mean and variance respectively. Since negative jumps would cause negative interest rates, we do not implement this jump setting in the case of CIR-type volatilities. The process with Gaussian-type volatility and normally distributed jump size is considered.

42

The third example is a normal distribution truncated to take only positive values. The jump transform has a closed-form solution, containing the cumulative normal density, which can be computed numerically. The density function of the truncated normal distribution is given by

$$\nu(x) = \frac{1}{1 - \Phi\left(-\frac{\mu_J}{\sigma_J}\right)} \frac{1}{\sigma_J} \phi\left(\frac{x - \mu_J}{\sigma_J}\right) 1_{\{x \geq 0\}},$$

where $1/(1 - \Phi(-\mu_J/\sigma_J))$ is the normalizing constant, which adjusts the total probability to one. The jump transform can be written as

$$\theta(c) = \int_0^\infty e^{cx} \, d\nu(x) = \frac{1 - \Phi\left(-\frac{\mu_J + c\sigma_J^2}{\sigma_J}\right)}{1 - \Phi\left(-\frac{\mu_J}{\sigma_J}\right)} \exp\left(c\mu_J + \frac{c^2\sigma_J^2}{2}\right). \qquad (2.31)$$

The mean and variance are respectively given as

$$\mathbb{E}[x \mid x > 0] = \mu_J + \frac{\phi\left(-\frac{\mu_J}{\sigma_J}\right)}{1 - \Phi\left(-\frac{\mu_J}{\sigma_J}\right)} \sigma_J,$$

$$\mathbb{V}\mathrm{ar}[x \mid x > 0] = \sigma_J^2 \left( 1 - \frac{\phi\left(-\frac{\mu_J}{\sigma_J}\right)}{1 - \Phi\left(-\frac{\mu_J}{\sigma_J}\right)} \left( \frac{\phi\left(-\frac{\mu_J}{\sigma_J}\right)}{1 - \Phi\left(-\frac{\mu_J}{\sigma_J}\right)} + \frac{\mu_J}{\sigma_J} \right) \right).$$

### 2.4.3 Parameters

The coefficients of an $n$-factor model are given by

Gaussian-type volatility:

$$\rho_1 = 1_n, \quad K = \mathrm{diag}[K_1, \ldots, K_n], \quad \theta = (\theta_1, \ldots, \theta_n)^\top,$$

$$\Sigma = \mathrm{diag}[\sigma_1, \ldots, \sigma_n]V, \quad \text{where } VV^\top = (\rho_{ij})_{ij},$$

$$D(X_t) = I_n, \quad \lambda(X_t) = \mathrm{diag}[l_1, \ldots, l_n].$$

CIR-type volatility:

$$\rho_1 = 1_n, \quad K = \mathrm{diag}[K_1, \ldots, K_n], \quad \theta = (\theta_1, \ldots, \theta_n)^\top, \quad (\theta_j > 0),$$

$$\Sigma = \mathrm{diag}[\sigma_1, \ldots, \sigma_n], \quad D(X_t) = \mathrm{diag}\left[\sqrt{X_{t,1}}, \ldots, \sqrt{X_{t,n}}\right],$$

$$\lambda(X_t) = \mathrm{diag}[l_1, \ldots, l_n].$$

The selected parameters are shown in Tables 2.1 and 2.2 . The drift and diffusion parameters are the same as those in Collin-Dufresne and Goldstein Collin-Dufresne and Goldstein (2002).

Table 2.1: Drift and diffusion parameters

|  | Gaussian | CIR |
|---|---|---|
| $\rho_0$ | 0.06 | 0.02 |
| $\rho_1$ | [1 1 1] | [1 1] |
| $K$ | [1 0.2 0.5] | [0.02 0.02] |
| $\theta$ | [0.0 0.0 0.0] | [0.03 0.01] |
| $\Sigma$ | diag[0.01 0.005 0.002]$V$ | diag[0.04 0.02]$V$ |
| $VV^\top$ | $\begin{pmatrix} 1 & -0.2 & -0.1 \\ -0.2 & 1 & 0.3 \\ -0.1 & 0.3 & 1 \end{pmatrix}$ | $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ |
| $X_0$ | [0.01 0.005 -0.02] | [0.04 0.02] |

Table 2.2: Jump parameters

|  | Gaussian | CIR |
|---|---|---|
| $l^u$, positive jump | [0 2 3] | [2 3] |
| $l^d$, negative jump | [0 2 3] |  |
| Exponential, positive $\eta^u$ | [0.00 0.0025 0.0005] | [0.0025 0.001] |
| negative $\eta^d$ | [0.00 0.0025 0.0005] |  |
| Normal, $\mu_J$ | [0 0.0025 -0.001] |  |
| $\sigma_J$ | [0 0.004 0.001] |  |
| Truncated normal, $\mu_J$ | [0 0.0025 -0.001] | [0.00125 -0.001] |
| $\sigma_J$ | [0 0.004 0.001] | [0.004 0.001] |

### 2.4.4 Gaussian-type Volatility Models

We consider the Gaussian-type volatility and jump. According to equation (2.23), the solutions of $B(t, T)$ and $\bar{B}(t)$ for the models containing the constant jump intensity are the same as those for the models without jumps. The solutions of $A(t, T)$ and $\bar{A}(t)$ can be split into two parts; the part containing the characteristics of the Gaussian diffusion, and the part containing the characteristics of the jump. $A(t, T)$ and $\bar{A}(t)$ are therefore written as

$$A(t, T) = A^G(t, T) + A^J(t, T),$$
$$\bar{A}(t) = \bar{A}^G(t) + \bar{A}^J(t),$$

where the superscripts $G$ and $J$ represent the part of the Gaussian diffusion and that of the jump respectively.

In the case of Gaussian-type volatility without jump, the zero-coupon bond price and the bond moments have closed-form solutions. The solutions of equations (2.22), (2.23) and (2.24) are given by

$$A^G(t,T) = - \rho_0(T-t) - \sum_{i=1}^{n}\left((T-t) - \frac{1-e^{-K_i(T-t)}}{K_i}\right)\theta_i$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n}\frac{\rho_{ij}\sigma_i\sigma_j}{K_iK_j}\left((T-t) - \frac{1-e^{-K_i(T-t)}}{K_i}\right.$$

$$\left. - \frac{1-e^{-K_j(T-t)}}{K_j} + \frac{1-e^{-(K_i+K_j)(T-t)}}{K_i+K_j}\right),$$

$$B_i^G(t,T) = - \frac{1-e^{-K_i(T-t)}}{K_i}, \quad i=1,2,\dots,n. \tag{2.32}$$

For bond moments, the solutions of equations (2.22), (2.23) and (2.25) are also given analytically by

$$\bar{A}^G(t) = A^G(t,T_0) + F_0 + \sum_{i=1}^{n}\theta_iF_i(1-e^{-K_i\tau})$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n}\rho_{ij}\sigma_i\sigma_j\left(F_iF_j\frac{1-e^{-(K_i+K_j)\tau}}{K_i+K_j}\right.$$

$$+ \frac{F_i}{K_j}\left(\frac{1-e^{-(K_i+K_j)\tau}}{K_i+K_j} - \frac{1-e^{-K_i\tau}}{K_i}\right)$$

$$\left. + \frac{F_j}{K_i}\left(\frac{1-e^{-(K_i+K_j)\tau}}{K_i+K_j} - \frac{1-e^{-K_j\tau}}{K_j}\right)\right),$$

$$\bar{B}_i^G(t) = B_i^G(t,T_0) + F_i\exp(-K_i\tau), \quad i=1,2,\dots,n, \tag{2.33}$$

where $\tau = T_0 - t$, $F_0 = \sum_{i=1}^{m}A^G(T_0,T_i)$ and $F_i = \sum_{j=1}^{m}B_i^G(T_0,T_j)$.

### Exponentially Distributed Jump Size

We assume the Gaussian-type volatility and exponentially distributed jump size, combined with the constant jump intensity. Two jumps in each factor are considered; one is positive with intensity denoted by $l^u$ and the mean of jump size denoted by $\eta^u$, the other is negative with properties $l^d$ and $\eta^d$ respectively.

**Moment expansions** Because of the constant jump intensity, the solutions for $B(t,T)$ and $\bar{B}(t)$ are given in equations (2.32) and (2.33) respectively and they are the same as for the case of the Gaussian-type volatility without jump. As stated above, the solutions for $A(t,T)$ and $\bar{A}(t)$ are split into two parts; the volatility parts are given above and jump parts are given by

$$
\begin{aligned}
A^J(t,T) = &-\sum_{i=1}^{n} l_i^u \left( (T-t) - \frac{1}{K_i + \eta_i^u} \log \left| \frac{(K_i + \eta_i^u)e^{K_i(T-t)} - \eta_i^u}{K_i} \right| \right) \\
&-\sum_{i=1}^{n} l_i^d \left( (T-t) - \frac{1}{K_i - \eta_i^d} \log \left| \frac{(K_i - \eta_i^d)e^{K_i(T-t)} + \eta_i^d}{K_i} \right| \right), \\
\bar{A}^J(t) = &-\sum_{i=1}^{n} l_i^u \left( \tau - \frac{1}{K_i + \eta_i^u} \log \left| \frac{(K_i + \eta_i^u)e^{K_i\tau} - \eta_i^u(KF_i + 1)}{K_i(1 - \eta_i^u F_i)} \right| \right) \\
&-\sum_{i=1}^{n} l_i^d \left( \tau - \frac{1}{K_i - \eta_i^d} \log \left| \frac{(K_i - \eta_i^d)e^{K_i\tau} + \eta_i^d(KF_i + 1)}{K_i(1 + \eta_i^d F_i)} \right| \right),
\end{aligned}
$$

where $F_0$ and $F_i$ being contained in $\bar{A}(t)$ and $\bar{B}(t)$ are respectively given by $F_0 = \sum_{i=1}^{m} A(T_0, T_i)$ and $F_i = \sum_{j=1}^{m} B_i(T_0, T_j)$.

**Monte Carlo simulation** The integrated jump intensity for positive jump used in simulating jump times has a closed-form solution and is given from equations (2.26) and (2.29) by,

$$
\begin{aligned}
\int_{\tau_{i,j}}^{t} \lambda_i^{T_0}(u)\,du &= \int_{\tau_{i,j}}^{t} \frac{l_i^u}{1 - \eta_i^u B_i(u, T_0)}\,du \\
&= \frac{-l_i^u}{K_i + \eta_i^u} \log \left| \frac{(K_i + \eta_i^u)e^{K_i(T_0-t)} - \eta_i^u}{(K_i + \eta_i^u)e^{K_i(T_0-\tau_{i,j})} - \eta_i^u} \right|, \quad i = 1, 2, \ldots, n.
\end{aligned}
$$

The jump transform for positive jumps under $\mathbb{P}_{T_0}$ determines the jump size distribution under the measure and it is given from equations (2.27) and (2.29) by

$$
\theta_i^{T_0}(c) = \frac{1}{1 - c\frac{\eta_i^u}{1 - \eta_i^u B_i(t, T_0)}}, \quad i = 1, 2, \ldots, n.
$$

Hence, the jump size under the $T_0$-forward measure follows an exponential with mean $\eta_i/(1 - \eta_i B_i(t, T_0))$. Since the underlying process follows a Gaussian between jumps, simulating a sample path between jumps does not entail a discretization error. For a negative jump, the jump intensity and jump transform

46

under the forward measure are defined respectively as

$$\lambda_i^{T_0}(t) = \frac{l_i^d}{1 + \eta_i^d B_i(t, T_0)},$$

$$\theta_i^{T_0}(c) = \frac{1}{1 - c\frac{\eta_i^d}{1 + \eta_i^d B_i(t, T_0)}}, \quad i = 1, 2, \ldots, n.$$

**Results** The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7), and that of the corresponding call options used in the generalized Edgeworth expansion are shown in Figure 2.1. All results point to a good match. While each jump makes the distribution of $X$ asymmetric, density functions show symmetry. This is caused by the assumption of two asymmetric jumps; one is positive and the other negative. The price differences, defined as

Figure 2.1: The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option used in the generalized Edgeworth expansion. Three-factor Gaussian-type volatility with exponentially distributed jump size.
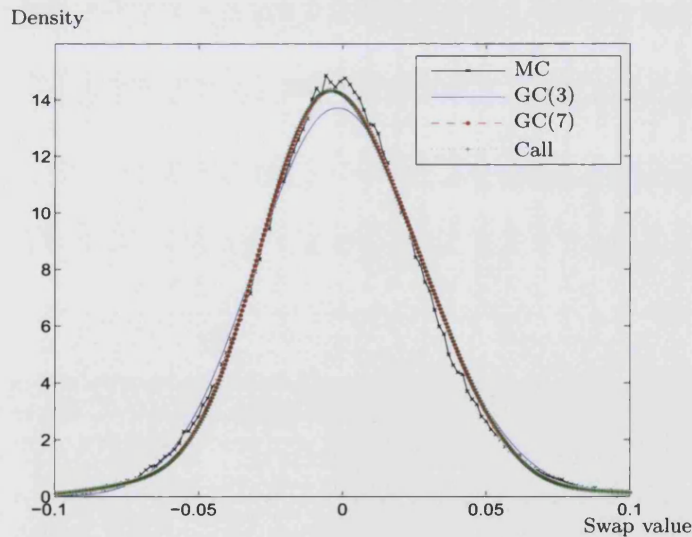


*Monte Carlo price − approximation price*, are presented in Figure 2.2 and prices are shown in Table 2.3. Monte Carlo results are obtained with four million simulations. GC(3) has the worst performance on average. This is because GC(3) captures only the first three moments of the swap value. The higher the

moment order, the greater the accuracy of the Gram–Charlier expansion, though the accuracy differs across the strike rates. In contrast to GC(3), GEW(2,7) and GEW(3,7,7) perform well although they capture only the first two and three moments of the swap value respectively. This is because in the generalized Edgeworth expansion, $P(T_0, T_N)$ is used as an approximating function which partially captures the influence of higher-order moments.

Figure 2.2: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Three-factor Gaussian-type volatility with exponentially distributed jump size.
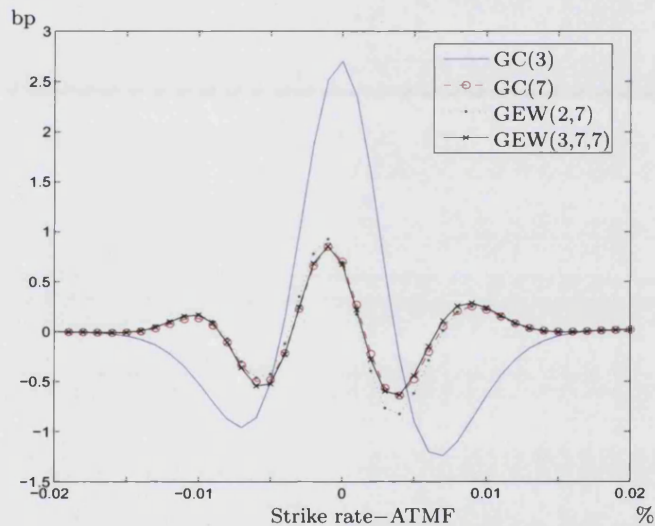


In the generalized Edgeworth expansion, the ITM swaption requires a greater number of call options than OTM does as shown in the $N_1$ row in Table 2.3. $N_2/N_1$ represents the strike price of a call option and from Table 2.3, it shows that ITM, ATMF and OTM swaptions are decomposed into ITM, ATM and OTM call options and adjustment terms respectively. The call option prices come close to the true swaption prices. These results support the idea that a swaption price can be mainly hedged only with options on a zero-coupon bond. Overall these approximation methods work well, but comparing the results with those under the diffusion process, approximated prices under the AJD perform worse by one digit as shown in Collin-Dufresne and Goldstein (2002, Exhibit 2)

and Tanaka, Yamada and Watanabe (2010, Figure 1).

Table 2.3: One-into-ten swaption prices by Monte Carlo simulation, GC(3), GC(7), corresponding call option price and GEW(3, 3, 7) over strike rate. ATM forward rate is 6.00%. Three-factor Gaussian-type volatility with exponentially distributed jump size.

| Strike rate, ATMF+$\kappa$% | −0.015 | −0.01 | −0.005 | 0.0 | 0.005 | 0.01 | 0.015 |
|---|---|---|---|---|---|---|---|
| MC | 0.04 | 0.93 | 13.54 | 107.33 | 366.79 | 706.50 | 1058.17 |
| GC(3) | 0.00 | 0.42 | 13.04 | 110.03 | 365.90 | 705.85 | 1058.13 |
| GC(7) | 0.04 | 1.07 | 13.06 | 108.03 | 366.31 | 706.71 | 1058.17 |
| GEW(2,7) | 0.04 | 1.11 | 13.12 | 108.03 | 366.17 | 706.71 | 1058.17 |
| GEW(3,7,7) | 0.04 | 1.10 | 13.02 | 108.00 | 366.35 | 706.73 | 1058.17 |
| $N_1$ | 1.39 | 1.43 | 1.48 | 1.52 | 1.57 | 1.61 | 1.65 |
| $N_2/N_1$ | 0.63 | 0.60 | 0.58 | 0.55 | 0.53 | 0.51 | 0.49 |

Table 2.4: Running time by Monte Carlo simulation with 200,000 sample paths, GC(3), GC(7), and GEW(3,7,7). Three-factor Gaussian-type volatility with exponentially distributed jump size.

| | Monte Carlo | GC(3) | GC(7) | GEW(3,7,7) |
|---|---|---|---|---|
| seconds | 179.654 | 0.015 | 3.328 | 0.031 |

Table 2.4 illustrates the CPU time in seconds to calculate one swaption price by Monte Carlo simulation, GC(3), GC(7) and GEW(3,7,7). All computations were performed on a PC with Intel(R) Core 2 CPU 2.40 GHz with 3 GB RAM. Since an iterative calculation is required to calculate the moments of swaps, the higher the moment order, the greater the CPU time. All approximation methods run much faster than Monte Carlo simulation.

**Normally Distributed Jump Size**

The model for the Gaussian-type volatility with normally distributed jump size is now considered.

**Moment expansions** Due to the constant jump intensity, the solutions of $B(t,T)$ and $\bar{B}(t)$ are the same as those for the model without jumps and are given in equations (2.32) and (2.33) respectively. The ODEs for $A(t,T)$ and $\bar{A}(t)$ do not have closed-form solutions since the ODE satisfied by $A(t,T)$ becomes

$$\frac{d}{dt}A(t,T) = -(K\theta)^{\top}B(t,T) - \frac{1}{2}\sum_{i=1}^{n}(\Sigma^{\top}B(t,T))_i^2$$

$$- \sum_{i=1}^{n}l_i\big(e^{\mu_{J,i}B_i(t,T)+\sigma_{J,i}^2 B_i(t,T)^2/2} - 1\big) + \rho_0.$$

The terminal condition is $A(T,T) = 0$. $\bar{A}(t)$ satisfies the same ODE in which $\bar{B}(t)$ is used instead of $B(t,T)$ and the terminal condition is $\bar{A}(T_0) = \sum_{i=1}^{m} A(T_0,T_i)$. These ODEs are solvable with numerical methods such as the Runge-Kutta method and solutions contain discretization errors accordingly.

**Monte Carlo simulation** The jump intensity under the $T_0$-forward measure is given from equations (2.26) and (2.30) by

$$\lambda_i^{T_0} = l_i \exp\Big(\mu_{J,i}B_i(t,T_0) + \frac{\sigma_{J,i}^2 B_i(t,T_0)^2}{2}\Big), \quad i = 1,2,\ldots,n. \qquad (2.34)$$

Since the integral in equation (2.28) combined with equation (2.34) cannot be obtained in closed-form, the trapezium rule with equi-distant time steps is used in equation (2.28), hence including discretization error. The jump transform under the $T_0$-forward measure is from equations (2.27) and (2.30):

$$\theta_i^{T_0}(c) = \exp\Big(c(\mu_{J,i} + \sigma_{J,i}^2 B(t,T_0)) + \frac{\sigma_{J,i}^2 c^2}{2}\Big), \quad i = 1,2,\ldots,n.$$

Hence, the jump size distribution under the $T_0$-forward measure follows the normal distribution whose mean and variance are $\mu_J + \sigma_J^2 B(t,T_0)$ and $\sigma_J^2$ respectively. Monte Carlo simulation under this setting causes discretization error in sampling jump times though no discretization error in sampling diffusion parts. We discretize option expiry into ten thousand of equal length.

**Results** The density of swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option computed by GEW(2,7) are shown

Figure 2.3: The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option used in the generalized Edgeworth expansion. Three-factor Gaussian-type volatility with normally distributed jump size.



Figure 2.4: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Three-factor Gaussian-type volatility with normally distributed jump size.



51
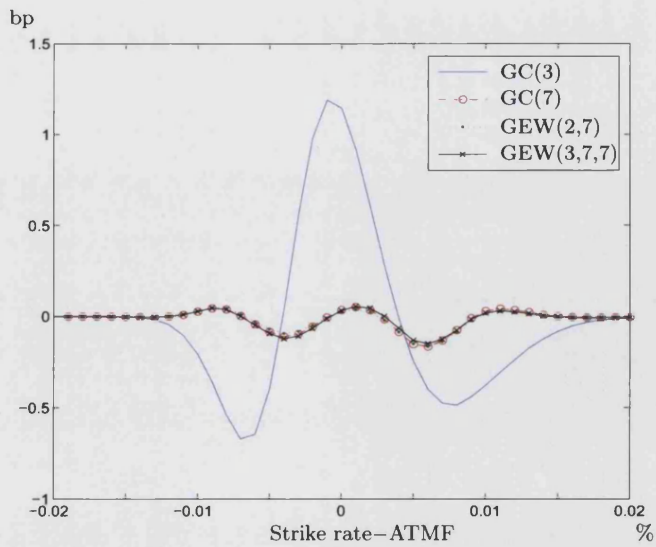
in Figure 2.3. Due to the assumption that jump size distributions follow a normal with non-zero mean, the densities show asymmetry.

Figure 2.4 presents the price differences and prices are shown in Table 2.5. For the Monte Carlo simulation, we use four million sample paths. While a similar tendency is observed as in the case with exponential jump size, overall pricing errors decrease. With the Gram–Charlier expansion, the accuracy improves as the higher-order moments are used though the accuracy differs across the strike rates. GEW$(2,7)$ and GEW$(3,7,7)$ perform almost the same as GC$(7)$. This is because in the generalized Edgeworth expansion, $P(T_0, T_N)$ is used as an approximating function which partially captures the influence of higher-order moments.

Table 2.5: One-into-ten swaption prices by Monte Carlo simulation, GC$(3)$, GC$(7)$, corresponding call option price and GEW$(3, 3, 7)$ over strike rate. ATM forward rate is 7.04%. Three-factor Gaussian-type volatility with normally distributed jump size.

| Strike rate, ATMF+$\kappa$% | $-0.015$ | $-0.01$ | $-0.005$ | 0.0 | 0.005 | 0.01 | 0.015 |
|---|---|---|---|---|---|---|---|
| MC | 0.00 | 0.10 | 7.61 | 101.45 | 354.70 | 679.02 | 1016.29 |
| GC(3) | 0.00 | -0.11 | 7.21 | 102.60 | 354.45 | 678.65 | 1016.21 |
| GC(7) | 0.00 | 0.13 | 7.52 | 101.49 | 354.55 | 679.05 | 1016.29 |
| GEW(2,7) | 0.00 | 0.13 | 7.54 | 101.48 | 354.55 | 679.04 | 1016.28 |
| GEW(3,7,7) | 0.00 | 0.13 | 7.51 | 101.49 | 354.57 | 679.04 | 1016.28 |
| $N_1$ | 1.51 | 1.56 | 1.60 | 1.65 | 1.70 | 1.74 | 1.79 |
| $N_2/N_1$ | 0.57 | 0.54 | 0.52 | 0.50 | 0.47 | 0.45 | 0.44 |

In the generalized Edgeworth expansion, a similar tendency is observed as the case with the exponential distribution setting. The ITM swaption requires a greater number of call options than OTM does as shown in the $N_1$ row in Table 2.5. ITM, ATMF and OTM swaptions are decomposed into ITM, ATM and OTM call options and adjustment terms respectively, as shown in the $N_2/N_1$ row in Table 2.5. The call option prices come close to the true swaption prices. These results again support the idea that a swaption price can be mainly hedged only with an option on a zero-coupon bond.

Table 2.6: Running time by Monte Carlo simulation with 200,000 sample paths, GC(3), GC(7), and GEW(3,7,7). Three-factor Gaussian-type volatility with normally distributed jump size.

|  | Monte Carlo | GC(3) | GC(7) | GEW(3,7,7) |
|---|---|---|---|---|
| seconds | 2841.310 | 0.244 | 140.077 | 1.544 |

Table 2.6 illustrates the CPU time in seconds to calculate one swaption price by Monte Carlo simulation, GC(3), GC(7) and GEW(3,7,7) with the same PC as in the exponentially distributed jump size case. Approximation methods run faster than Monte Carlo simulation. The price by Monte Carlo simulation is computed with the equi-spaced time discretization of $10^{-5}$. Due to the time discretization, Monte Carlo simulation under the normal jump size is more time consuming than that under the exponential jump size. The ODE is computed numerically with time discretization of $2 \times 10^{-2}$ and therefore it makes the CPU time increase.

**Truncated Normally Distributed Jump Size**

Gaussian-type volatility with truncated normally distributed jump size is considered in this example.

**Moment expansion** The solutions for $B(t,T)$ and $\bar{B}(t)$ are the same as those for the diffusion case because of the constant intensity and are given in equations (2.32) and (2.33) respectively. The ODE for $A(t,T)$ is given by

$$\frac{d}{dt}A(t,T) = -(K\theta)^\top B(t,T) - \frac{1}{2}\sum_{i=1}^{n}\left(\Sigma^\top B(t,T)\right)_i^2$$

$$-\sum_{i=1}^{n}l_i\left(\frac{1 - \Phi\left(-\frac{\mu_{J,i}+B_i(t,T)\sigma_{J,i}^2}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i}}{\sigma_{J,i}}\right)}e^{\mu_{J,i}B_i(t,T)+\frac{\sigma_{J,i}^2 B_i(t,T)^2}{2}} - 1\right) + \rho_0,$$

with the terminal condition $A(T,T) = 0$. $\bar{A}(t)$ satisfies the same equation as $A(t)$ does in which $\bar{B}(t)$ is used instead of $B(t,T)$. The terminal condition is $\bar{A}(T_0) = \sum_{i=1}^{m} A(T_0, T_i)$. Like the normal jump setting, these equations are

solved numerically via the Runge-Kutta method, thereby including discretization error.

**Monte Carlo simulation**  Jump intensity under the $T_0$-forward measure used in a jump time simulation is given from equations (2.26) and (2.31) by

$$\lambda_i^{T_0} = l_i \frac{1 - \Phi\left(-\frac{\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T_0)}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i}}{\sigma_{J,i}}\right)} e^{\mu_{J,i} B_i(t,T_0) + \frac{\sigma_{J,i}^2 B_i(t,T_0)^2}{2}}, \quad i = 1, 2, \ldots, n.$$

$$(2.35)$$

Since the integral in equation (2.28) combined with (2.35) cannot be obtained in closed-form, it is performed numerically by adopting the trapezium rule with equi-distant time step in equation (2.28), thereby including discretization error. The jump transform under the $T_0$-forward measure takes the form from equations (2.27) and (2.31)

$$\theta_i^{T_0}(c) = \frac{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma_{J,i}^2 + c\sigma_{J,i}^2}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma_{J,i}^2}{\sigma_{J,i}}\right)} e^{(\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T_0))c + \frac{\sigma_{J,i}^2 c^2}{2}}, \quad i = 1, 2, \ldots, n.$$

By comparing this with equation (2.31), this equation indicates that the jump size distribution under the $T_0$-forward measure follows $N(\mu_{J,i} + \sigma_{J,i}^2 B_i(t, T_0), \sigma_{J,i}^2)$ which is truncated at $\sigma_{J,i}^2 B_i(t, T_0)$. Hence, we deduce under the $T_0$-forward measure, the jump size follows, for $i = 1, 2, \ldots, n$,

$$\nu_i^{T_0}(x) =$$

$$\frac{1}{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma_{J,i}^2}{\sigma_{J,i}}\right)} \frac{1}{\sigma_{J,i}} \phi\left(\frac{x - \left(\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T_0)\right)}{\sigma_{J,i}}\right) 1_{\{x \geq B_i(t,T_0)\sigma_{J,i}^2\}}.$$

**Results**  The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and the density function of the call options used in the generalized Edgeworth expansion are shown in Figure 2.5. The density functions show asymmetry since only positive jumps are included. Even under the asymmetric jumps, approximation methods work well.

Figure 2.6 presents the price differences and prices are shown in Table 2.7. For the Monte Carlo simulation, we use four million sample paths. With the

Figure 2.5: The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option used in the generalized Edgeworth expansion. Three-factor Gaussian-type volatility with truncated normally distributed jump size.
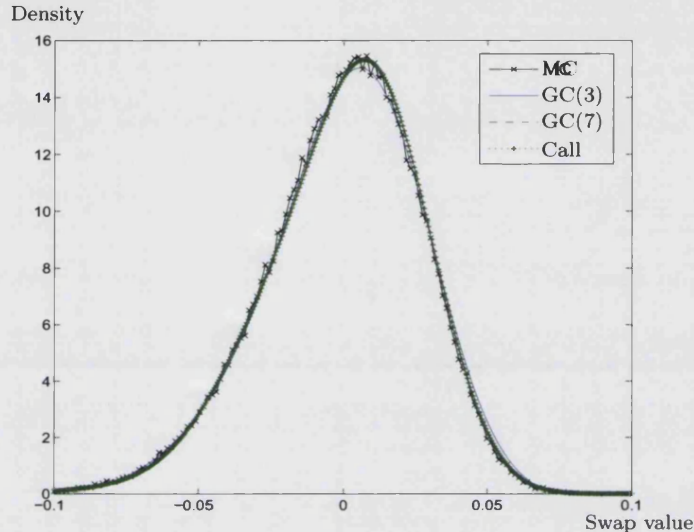


Gram–Charlier expansion, the accuracy improves as the higher-order moments are used, though the accuracy differs across the strike rates. $GEW(2,7)$ and $GEW(3,7,7)$ perform almost the same as $GC(7)$. This indicates that $GEW(2,7)$ and $GEW(3,7,7)$ capture the influence of the higher-order moments though they are set to equate up to second and third moments, respectively, to those of the swap value. This is because in the generalized Edgeworth expansion, $P(T_0, T_N)$ is used as an approximating function which partially captures the influence of higher-order moments.

Table 2.8 illustrates the CPU time in seconds to calculate one swaption price by Monte Carlo simulation, GC(3), GC(7) and GEW(3,7,7), again with the same PC as in the exponentially distributed jump size case. Approximation methods run faster than Monte Carlo simulation. The price by Monte Carlo simulation is computed with the equi-spaced time discretization of $10^{-5}$. Like the normal jump setting, time discretization makes Monte Carlo simulation time-consuming. The ODE is computed numerically with time discretization of $2 \times 10^{-2}$, and
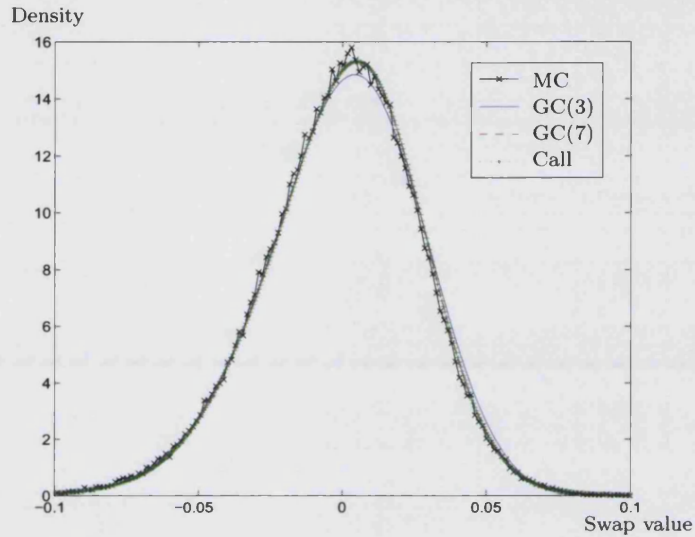
Figure 2.6: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Three-factor Gaussian-type volatility with truncated normally distributed jump size.



Table 2.7: One-into-ten swaption prices by Monte Carlo simulation, GC(3), GC(7), corresponding call option price and GEW(3, 3, 7) over strike rate. ATM forward rate is 8.92%. Three-factor Gaussian-type volatility with truncated normally distributed jump size.

| Strike rate, ATMF+$\kappa$% | $-0.015$ | $-0.01$ | $-0.005$ | 0.0 | 0.005 | 0.01 | 0.015 |
|---|---|---|---|---|---|---|---|
| MC | 0.00 | 0.04 | 8.07 | 101.29 | 332.25 | 628.22 | 938.78 |
| GC(3) | -0.01 | -0.29 | 8.18 | 102.07 | 332.46 | 627.88 | 938.65 |
| GC(7) | 0.00 | 0.07 | 8.10 | 101.45 | 332.29 | 628.23 | 938.81 |
| GEW(2,7) | 0.00 | 0.06 | 8.12 | 101.47 | 332.29 | 628.20 | 938.80 |
| GEW(3,7,7) | 0.00 | 0.06 | 8.08 | 101.48 | 332.26 | 628.21 | 938.80 |
| $N_1$ | 1.75 | 1.80 | 1.86 | 1.91 | 1.96 | 2.01 | 2.06 |
| $N_2/N_1$ | 0.46 | 0.44 | 0.43 | 0.41 | 0.39 | 0.37 | 0.36 |

Table 2.8: Running time by Monte Carlo simulation with 200,000 sample paths, GC(3), GC(7), and GEW(3,7,7). Three-factor Gaussian-type volatility with truncated normally distributed jump size.

| | Monte Carlo | GC(3) | GC(7) | GEW(3,7,7) |
|---|---|---|---|---|
| seconds | 3812.484 | 0.468 | 251.740 | 3.372 |

therefore it makes the CPU time increase.

### 2.4.5 CIR-type Volatility Models

In this subsection, we consider the CIR-type volatility. Well-known properties of the CIR process include that it is non-negative and mean-reverting because of the drift term. For a strictly positive interest rate, jump size distributions taking only positive values are implemented. Like the Gaussian-type volatility setting, the solutions of $A(t,T)$ and $\bar{A}(t)$ can be split into volatility and jump parts. Thus in the case of a jump-diffusion setting, $A(t,T)$ and $\bar{A}(t)$ are written as

$$A(t,T) = A^C(t,T) + A^J(t,T),$$
$$\bar{A}(t,T) = \bar{A}^C(t,T) + \bar{A}^J(t,T).$$

In the case of the independent CIR-type volatility without jump terms, the system of ODEs has analytical solutions. We denote the solutions of equations (2.22), (2.23) and (2.24) by $A^C(t,T)$ and $B^C(t,T)$ and those of equations (2.22), (2.23) and (2.25) by $\bar{A}^C(t)$ and $\bar{B}^C(t)$. These are given by

$$A^C(t,T) = -\rho_0(T-t)$$
$$- \sum_{i=1}^{n} K_i \theta_i \left[ \frac{2}{\sigma_i^2} \ln \frac{(K_i + \gamma_i)(e^{\gamma_i(T-t)} - 1) + 2\gamma_i}{2\gamma_i} + \frac{2}{K_i - \gamma_i}(T-t) \right],$$
$$B_i^C(t,T) = \frac{-2(e^{\gamma_i(T-t)} - 1)}{(K_i + \gamma_i)(e^{\gamma_i(T-t)} - 1) + 2\gamma_i}, \quad i = 1, 2, \ldots, n, \tag{2.36}$$

$$\bar{A}^C(t) = F_0 - \rho_0 \tau - \sum_{i=1}^{n} K_i \theta_i \left[ \frac{2}{\sigma_i^2} \ln \frac{(K_i + \gamma_i - \sigma_i^2 F_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}{2\gamma_i} \right.$$

57

$$+ \frac{(K_i + \gamma_i)F_i + 2}{K_i - \gamma_i - \sigma_i^2 F_i} \tau \Bigg],$$

$$\bar{B}_i^C(t) = \frac{-\left((K_i - \gamma_i)F_i + 2\right)(e^{\gamma_i \tau} - 1) + 2\gamma_i F_i}{(K_i + \gamma_i - \sigma_i^2 F_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}, \quad i = 1, 2, \dots, n, \qquad (2.37)$$

where $F_0 = \sum_{i=1}^{m} A^C(T_0, T_i)$, $F_i = \sum_{j=1}^{m} B_i^C(T_0, T_j)$, $\gamma_i = \sqrt{K_i^2 + 2\sigma_i^2}$ and $\tau = T_0 - t$.

## Exponentially Distributed Jump Size

The CIR-type volatility with the exponentially distributed jump-size is considered in this example.

**Moment expansion** Due to the constant intensity, the solutions for $B(t,T)$ and $\bar{B}(t)$ are the same as those without jumps and are given in equations (2.36) and (2.37) respectively. The jump parts of $A(t,T)$ and $\bar{A}(t)$ are respectively given by

$$A^J(t,T) = -\sum_{i=1}^{n} l_i \left( (T - t) + \frac{2\eta_i}{\bar{\gamma}_i} \log \left| \frac{\gamma_i^*(e^{\gamma_i(T-t)} - 1) + 2\gamma_i}{2\gamma_i e^{\gamma_i(T-t)/2\eta_i}} \right| \right),$$

$$\bar{A}^J(t) = -\sum_{i=1}^{n} l_i \left( \tau + \frac{2\eta_i}{\bar{\gamma}_i} \log \frac{(\gamma_i^* - \hat{\gamma}_i F_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i(1 - \eta_i F_i)}{2\gamma_i(1 - \eta_i F_i)e^{\hat{\gamma}_i \tau/2\eta_i}} \right),$$

where $\bar{\gamma}_i = \sigma_i^2 - 2K\eta_i - 2\eta_i^2$, $\gamma_i^* = K_i + \gamma_i + 2\eta_i$, $\hat{\gamma}_i = \sigma_i^2 - \eta_i K_i + \eta_i \gamma_i$. The terminal conditions used in the bond moments are $F_0 = \sum_{i=1}^{m} A(T_0, T_i)$ and $F_i = \sum_{j=1}^{m} B_i(T_0, T_j)$.

**Monte Carlo simulation** The integral in equation (2.28) used to simulate jump times under the $T_0$-forward measure is given from equations (2.26) and (2.29) by

$$\int_{\tau_{i,j}}^{t} \lambda_i^{T_0}(u)\, du$$

$$= \int_{\tau_{i,j}}^{t} \frac{l_i}{1 - \eta_i B_i(u, T_0)}\, du$$

$$= l_i \left( \frac{2\eta_i}{\bar{\gamma}_i} \log \left| \frac{\gamma_i^*(e^{\gamma_i(T_0-t)} - 1) + 2\gamma_i}{\gamma_i^*(e^{\gamma_i(T_0-\tau_{i,j})} - 1) + 2\gamma_i} \right| - \frac{K_i - \gamma_i}{K_i - \gamma_i + 2\eta_i}(t - \tau_{i,j}) \right),$$

$i = 1, 2, \ldots, n$, and therefore it satisfies the nonlinear equation, which can be solved by means of the bisection algorithm. Since the underlying process follows CIR between jumps, simulating a sample path without time-discretization is possible. The jump transform under the $T_0$-forward measure from equations (2.27) and (2.29) is

$$\theta_i^{T_0}(c) = \frac{1}{1 - c\frac{\eta_i}{1-\eta_i B_i(t,T_0)}}, \quad i = 1, 2, \ldots, n.$$

Consequently, we deduce at each jump time that an additive jump occurs and its jump size follows an exponential with mean

$$\frac{\eta_i}{1 - \eta_i B_i(\tau_{i,j}, T_0)}, \quad i = 1, 2, \ldots, n.$$

Figure 2.7: The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option used in the generalized Edgeworth expansion. Two-factor CIR-type volatility with exponentially distributed jump size.
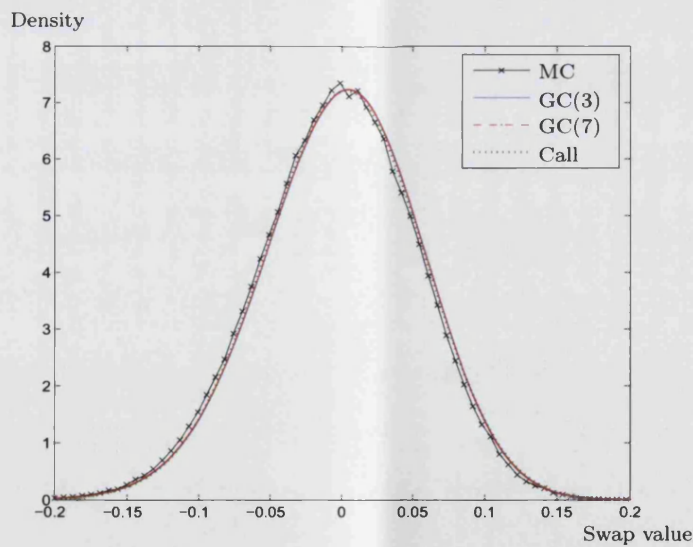


**Results** The densities of swap values computed by Monte Carlo simulation, GC(3) and GC(7) and that of the corresponding call option used in GEW(2,7) and GEW(3,7,7) are shown in Figure 2.7. It may seem that all methods work well. Figure 2.8 presents the price differences and prices are shown in Table 2.9.

Figure 2.8: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Two-factor CIR-type volatility with exponentially distributed jump size.
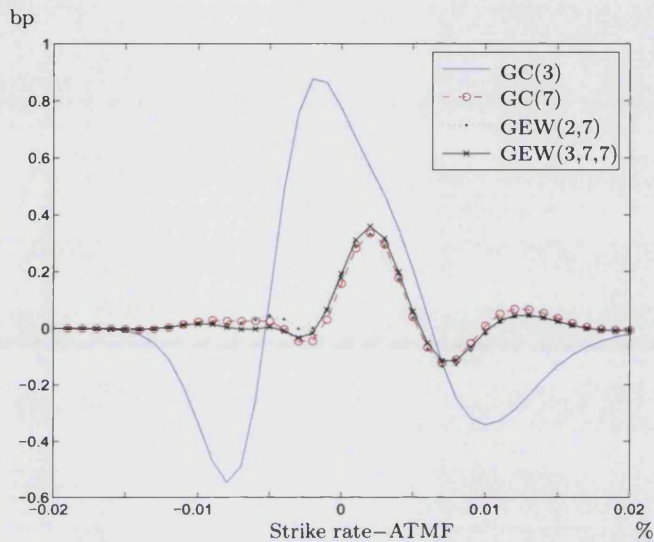


Table 2.9: One-into-ten swaption prices by Monte Carlo simulation, $GC(3)$, $GC(7)$, corresponding call option price and $GEW(3,3,7)$ over strike rate. ATM forward rate is 11.75%. Two-factor CIR-type volatility with exponentially distributed jump size.

| Strike rate, ATMF+$\kappa$% | $-0.015$ | $-0.01$ | $-0.005$ | 0.0 | 0.005 | 0.01 | 0.015 |
|---|---|---|---|---|---|---|---|
| MC | 6.30 | 28.70 | 88.37 | 202.72 | 374.52 | 592.11 | 838.80 |
| $GC(3)$ | 6.23 | 28.74 | 88.48 | 202.87 | 374.67 | 592.20 | 838.78 |
| $GC(7)$ | 6.30 | 28.70 | 88.36 | 202.73 | 374.53 | 592.09 | 838.78 |
| $GEW(2,7)$ | 6.84 | 29.71 | 89.32 | 202.91 | 373.80 | 590.96 | 837.76 |
| $GEW(3,7,7)$ | 6.25 | 28.75 | 88.46 | 202.81 | 374.63 | 592.20 | 838.82 |
| $N_1$ | 1.92 | 1.96 | 2.01 | 2.05 | 2.10 | 2.14 | 2.19 |
| $N_2/N_1$ | 0.35 | 0.33 | 0.32 | 0.30 | 0.29 | 0.27 | 0.26 |

Monte Carlo prices are computed with four million sample paths. In contrast to the Gaussian-type volatility setting, an approximation only with the call option performs less well. This is caused by the possibility that $G$ does not capture the majority of the influence on swaption price. Since $G$ contains one random variable $P(T_0, T_N)$ and has the same first and second order cumulants as the swap value, it cannot capture all the influence of more than third order cumulants of the swap value. In the case of a higher swap rate or higher volatility of the underlying interest rates and accordingly larger values of more than third order cumulants, $G$ cannot adequately capture the influence of more than third order moments, and this is the case for the CIR-type volatility setting (see the parameters in Table 3.3). This hypothesis is numerically investigated later. Nevertheless, GEW(3,7,7) captures the influence of up to third order moments and therefore the results improve. Though the approximation with only call price performs worst, its overall mean absolute error is less than 0.5 bp. This result supports the idea that a swaption can be hedged with the corresponding call option.

Table 2.9 shows also the strike rates and the number of contracts of call options used in the generalized Edgeworth expansion. ITM, ATMF and OTM swaptions are decomposed into ITM, ATM and OTM call options and adjustment terms respectively.

Table 2.10: Running time by Monte Carlo simulation with 2,000,000 sample paths, GC(3), GC(7), and GEW(3,7,7). Two-factor CIR-type volatility with exponentially distributed jump size.

| | Monte Carlo | GC(3) | GC(7) | GEW(3,7,7) |
|---|---|---|---|---|
| seconds | 160.061 | 0.015 | 5.124 | 0.046 |

Table 2.10 illustrates the CPU time in seconds to calculate one swaption price by Monte Carlo simulation, GC(3), GC(7) and GEW(3,7,7). All computations were performed on a PC with Intel(R) Core 2 CPU 2.40 GHz with 3 GB RAM. All approximation methods run much faster than Monte Carlo simulation. It indicates that these approximation methods can be used for practical implementation.

Table 2.11: New parameters. Others are the same as in Tables 2.1 and 2.2.

| $\rho_0$ | $\Sigma$ | $X_0$ |
|---|---|---|
| $-0.02$ | $\mathrm{diag}[0.02\,0.01]V$ | $[0.02\,0.01]$ |

To investigate the hypothesis, we set parameters so that they have the lower swap rate and the lower volatility shown in Table 2.11. The price differences are presented in Figure 2.9. As inferred, the patterns of the price differences are similar to that under the Gaussian-type volatility settings. These results support the hypothesis that in the case of a lower swap rate and a lower volatility, the lower order approximation with the generalized Edgeworth expansion will produce accurate prices. However, the overall approximation errors deteriorated compared with Figure 2.8. This is because the jump parameters are unchanged and consequently the density of the swap value deviates from the normal distribution.

Figure 2.9: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Two-factor CIR-type volatility with exponentially distributed jump size.
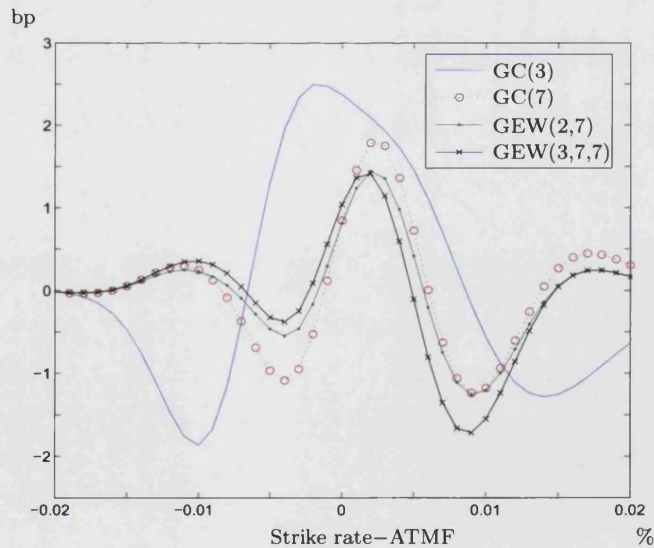
**Truncated Normally Distributed Jump Size**

In this example, we consider the CIR-type volatility and normally distributed jump size truncated to take only positive values.

**Moment expansion**  Due to the constant intensity, the solutions for $B(t, T)$ and $\bar{B}(t)$ are the same as those of the CIR-type volatility without jump and they are given by equations (2.36) and (2.37). The ODE satisfied by $A(t, T)$ becomes

$$\frac{d}{dt} A(t, T) = -(K\theta)^{\top} B(t, T)$$

$$- \sum_{i=1}^{n} l_i \left( \frac{1 - \Phi\left(-\frac{\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T)}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i}}{\sigma_{J,i}}\right)} e^{\mu_{J,i} B_i(t,T) + \frac{\sigma_{J,i}^2 B_i(t,T)^2}{2}} - 1 \right) + \rho_0,$$

with the terminal condition $A(T, T) = 0$. The ODE for $\bar{A}(t)$ is the same as that for $A(t, T)$ in which $\bar{B}(t)$ is used instead of $B(t, T)$ and its terminal condition is $\bar{A}(T_0) = \sum_{i=1}^{m} A(T_0, T_i)$. These equations are solved numerically by means of the Runge-Kutta method.

**Monte Carlo simulation**  The jump intensity under the $T_0$-forward measure is given from equations (2.26) and (2.31) by

$$\lambda_i^{T_0}(t) = l_i \frac{1 - \Phi\left(-\frac{\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T_0)}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i}}{\sigma_{J,i}}\right)} e^{\mu_{J,i} B_i(t,T_0) + \frac{\sigma_{J,i}^2 B_i(t,T_0)^2}{2}}, \quad i = 1, 2, \dots, n.$$

(2.38)

Since the integral in equation (2.28) with (2.38) has no analytical solution, we adopt the trapezium rule with equidistant time step. The jump transform under the $T_0$-forward measure is given from equations (2.27) and (2.31) by

$$\theta_i^{T_0}(c) = \frac{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma_{J,i}^2 + c\sigma_{J,i}^2}{\sigma_{J,i}}\right)}{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma_{J,i}^2}{\sigma_{J,i}}\right)} e^{(\mu_{J,i} + \sigma_{J,i}^2 B_i(t,T_0))c + \frac{\sigma_{J,i}^2 c^2}{2}}, \quad i = 1, 2, \dots, n,$$

and therefore, like the Gaussian-type volatility and truncated normal jump size setting, the jump size distribution for $i$-th factor under $\mathbb{P}_{T_0}$ follows $N(\mu_{J,i} +$

$\sigma^2_{J,i} B_i(t,T_0), \sigma^2_{J,i})$ being truncated at $\sigma^2_{J,i} B_i(t,T_0)$,

$$\nu_i^{T_0}(x) =$$

$$\frac{1}{1 - \Phi\left(-\frac{\mu_{J,i} + B_i(t,T_0)\sigma^2_{J,i}}{\sigma_{J,i}}\right)} \frac{1}{\sigma_{J,i}} \phi\left(\frac{x - \left(\mu_{J,i} + \sigma^2_{J,i} B_i(t,T_0)\right)}{\sigma_{J,i}}\right) 1_{\{x \geq \sigma^2_{J,i} B_i(t,T_0)\}}.$$

**Results**   Figure 2.10 shows the density functions of swap values computed in Monte Carlo simulation, GC(3), GC(7) and that of the corresponding call options computed in GEW. All results point to a good match. Figure 2.11 presents

Figure 2.10: The density functions of the swap value computed by Monte Carlo simulation, GC(3) and GC(7) and that of the call option used in the generalized Edgeworth expansion. Two-factor CIR-type volatility with truncated normally distributed jump size.
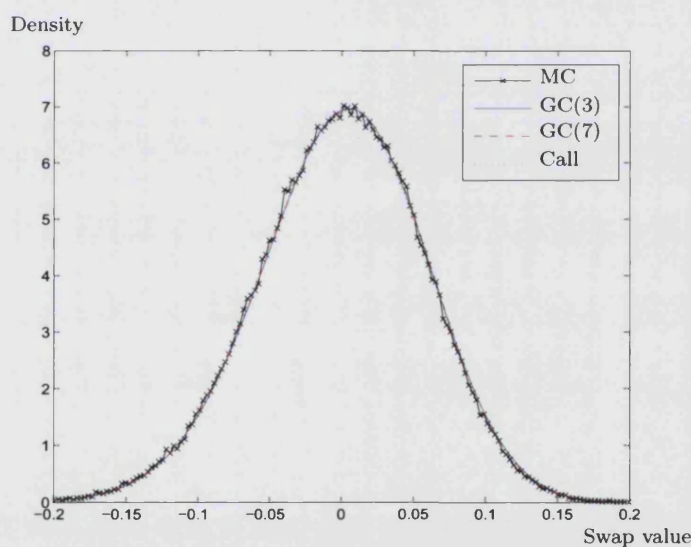


the price differences and prices are shown in Table 2.12. Monte Carlo prices are computed with four million sample paths. Similar to the CIR-type volatility and exponentially distributed jump-size setting, GEW(2, 7) perform the worst but GEW(3,7,7) perform well.

Table 2.13 illustrates the CPU time in seconds to calculate one swaption price by Monte Carlo simulation, GC(3), GC(7) and GEW(3,7,7), again with the same

Figure 2.11: Price differences between approximation methods and Monte Carlo simulation for various strike rates. Two-factor CIR-type volatility with the truncated normally distributed jump size.
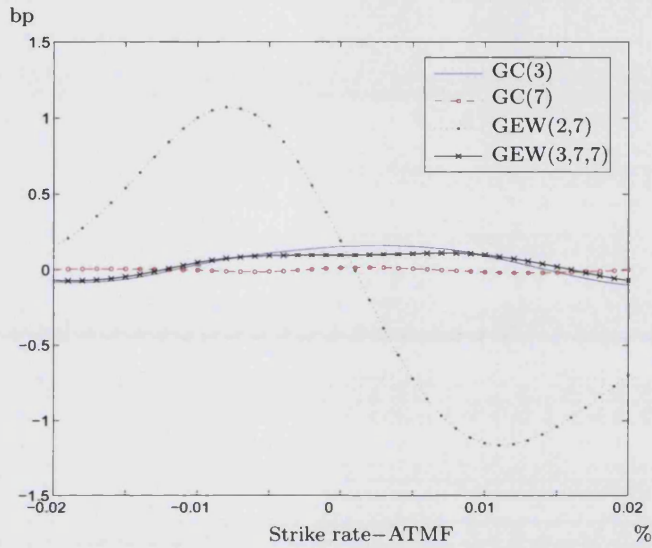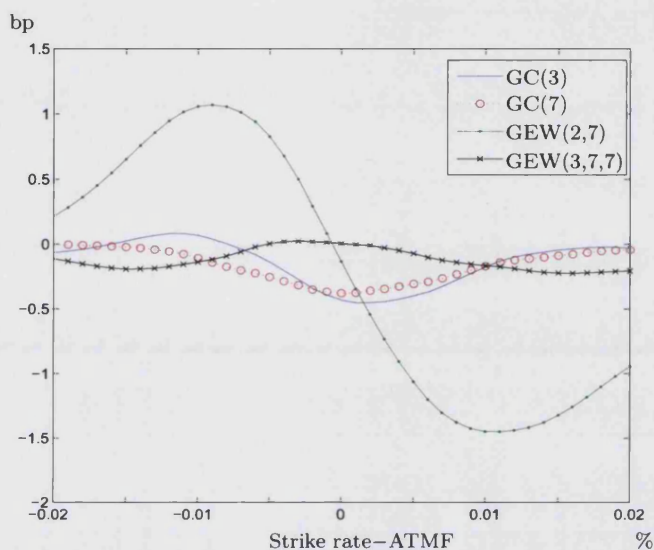


Table 2.12: One-into-ten swaption prices by Monte Carlo simulation, GC(3), GC(7), corresponding call option price and GEW(3, 3, 7) over strike rate. ATM forward rate is 12.17%. Two-factor CIR-type volatility with truncated normally distributed jump size.

| Strike rate, ATMF+$\kappa$% | −0.015 | −0.01 | −0.005 | 0.0 | 0.005 | 0.01 | 0.015 |
|---|---|---|---|---|---|---|---|
| MC | 7.99 | 33.32 | 96.13 | 211.10 | 379.50 | 590.90 | 830.74 |
| GC(3) | 8.01 | 33.39 | 95.99 | 210.67 | 379.10 | 590.72 | 830.70 |
| GC(7) | 7.97 | 33.22 | 95.88 | 210.72 | 379.20 | 590.73 | 830.65 |
| GEW(2,7) | 8.64 | 34.38 | 96.96 | 210.97 | 378.44 | 589.45 | 829.42 |
| GEW(3,7,7) | 7.80 | 33.18 | 96.13 | 211.11 | 379.43 | 590.73 | 830.52 |
| $N_1$ | 1.98 | 2.03 | 2.07 | 2.12 | 2.17 | 2.21 | 2.26 |
| $N_2/N_1$ | 0.33 | 0.32 | 0.30 | 0.29 | 0.27 | 0.26 | 0.25 |

Table 2.13: Running time by Monte Carlo simulation with 2,000,000 sample paths, GC(3), GC(7), and GEW(3,7,7). Two-factor CIR-type volatility with truncated normally distributed jump size.

|          | Monte Carlo | GC(3) | GC(7)   | GEW(3,7,7) |
|----------|-------------|-------|---------|------------|
| seconds  | 3747.446    | 0.256 | 142.669 | 1.620      |

PC as in the exponentially distributed jump size case. The price by Monte Carlo simulation is computed with 2,000,000 sample paths and time discretization of equal space $10^{-5}$. Due to the time discretization, Monte Carlo simulation is time consuming. Since each ODE used in the Gram–Charlier and generalized Edgeworth expansions are computed numerically via the Runge–Kutta method with time discretization of $2 \times 10^{-2}$, it makes the CPU time much longer than those of exponentially distributed jump size.

## 2.5 Concluding Remarks

In this chapter, we investigate the validity of two methods to evaluate swaption prices under the multi-factor affine jump-diffusion process. The density function of the underlying swap is approximated with more tractable densities and their higher-order cumulants. Two methods based upon moment expansions, the Gram–Charlier expansion and the generalized Edgeworth expansion, are examined.

The Gram–Charlier expansion uses normal distributions to approximate the swap value. The pricing of a swaption is reduced to calculating higher-order moments of the swap value. For some models such as the Gaussian-type volatility with exponential jump and the CIR-type independent volatility with exponential jump, moments of swap value can be calculated analytically. An arbitrary distribution can be used to approximate the swap value in a generalized Edgeworth expansion. We adopt one zero-coupon bond at swap maturity as the approximating random variable. It turns out that swaption prices can be expressed in terms of the options on a zero-coupon bond and adjustment terms. This decomposition indicates that a swaption can be partially hedged with options on

66

zero-coupon bonds. In this method, swaption pricing is reduced to calculating higher-order moments of the swap value and the approximating variable.

Based on the numerical examples, the higher-order approximation yields more accurate prices on average for both the Gram–Charlier and the generalized Edgeworth expansions in any jump setting. For some special cases, such as low swap rate and low volatility, the generalized Edgeworth expansion outperforms the Gram–Charlier expansion in the sense that it can attain the same level of pricing error but requires a lower degree of moments and less CPU time. The idea that the options on the zero-coupon bond can be used for hedging swaptions is supported numerically. For the normally distributed and the truncated normally distributed jump sizes, the ordinary differential equations are solved numerically. Solving the ordinary differential equations numerically works well although it is more time-consuming than an analytical approach. Even so, approximation methods run much faster than Monte Carlo simulation.

# Chapter 3

# The Spot and Forward Relationship in Carbon Emissions Markets I

## 3.1 Introduction

The Kyoto Protocol was adopted in December 1997. To achieve its commitment, signatory nations committed to reduce carbon dioxide ($CO_2$) emissions. The European Union (EU) put a new trading program called the EU Emission Trading Scheme (EU ETS) into practice in order to achieve this and consequently in January 2005, a new market in which European Union $CO_2$ allowances (EUAs) are traded was created. The EU ETS operates a cap–and–trade scheme under which the total quantity to be allocated to each EU member state is defined in the National Allocation Plan (NAP). Companies in specified industrial sectors or with combustion facilities, identified by member states under the terms of the NAP, have to comply with this requirement. They receive free emission permits or allowances annually, which represent the right to emit a specific amount, one allowance is equivalent to one tonne of $CO_2$, enabling them to emit $CO_2$ up to the assigned tonnage. The total allowance issued is thus set as a cap on total permitted emissions, and by setting the cap level tighter over time, total emissions are controlled.

In the market, companies assigned emission allowances can bilaterally trade these, to offset any excess or shortage. Companies that emit beyond their allowance must purchase allowances to cover their excess emissions. This scheme is designed to offer companies incentives for achieving emission reductions. Since companies that are able to keep emissions below their assignment are free to sell excess quotas, they are incentivised to develop and deploy cleaner technologies.

Actual trading is organised in several phases. Phase I (2005–2007) is regarded as a pilot period, set to run up to the beginning of Phase II (2008–2012), the Kyoto commitment period. The details of the trading scheme in Phase III (2013–2020) are still under review. In the market the spot contracts, as well as the futures contracts, are widely traded and other derivatives such as options have emerged. See Capoor and Ambrosi (2008) for recent developments in these markets.

EUA contracts are traded with several maturities. The EUA price of the current year can be viewed as the spot price and the EUA price for the next year as the forward price. Regulators announce verified emissions data every year and unused allowances are to be surrendered at the end of every trading year. Under the EU ETS, companies that are not able to obtain sufficient allowances to cover their emissions for the year have to pay a penalty (in Phase I, 40 EUR per tonne; in Phase II, 100 EUR per tonne). Moreover, any emissions not covered by an allowance have to be made up in the next year. Consequently penalty payments still force companies to obtain these allowances. This market regulation provides a direct relationship between spot and forward prices. However, only the situation where market participants need allowances, but are unable to acquire them, brings about such a direct relationship. In this sense, the factor representing the market condition has to be taken into account in expressing the spot and forward relationship. We, therefore, introduce the difference between the total allocation and emissions (referred to as the net position of the zone), for expressing the spot and forward relationship. It plays a crucial role in evaluating the relationship. The spot price depends not only on the forward price but also the net position of the zone, since only the short position of the zone, when the total allocation is below actual emissions, leads the spot price to the forward price plus the penalty. The spot and forward relationship in this market differs due to these regulations and market design from other markets such as zero-coupon bond or stock markets.

Bankability is taken into consideration in our framework. Within Phase I and Phase II, excess allowances can be transferred for use during the following year, which is called banking. Banking between Phase I and Phase II is forbidden by most countries, so that the unused allowances allocated during Phase I are useless after Phase I. Considering the lack of bankability, if it were certain that

the net position would be long, in other words the carbon allowances were in a state of excess supply, the spot price at the end of Phase I would be zero. When banking is possible, the unused allowances can be used for next year and therefore even if the market is long, the spot price at the end of each trading year is still greater than zero.

The objective of this article is to extend the framework proposed by Çetin and Verschuere (2009) for the spot and forward relationship in Phase I to that in Phase II. We make allowances for the possibility of banking in building the framework. The possibility of banking is evaluated in the form of an expectation, conditioned on the net position of the zone being long. Our approach is based upon the no-arbitrage principle. The spot price is considered as a derivative contract whose underlying variables are the forward price and the net position of the zone. This is pricing in an incomplete market since the net position of the zone is not tradable. The locally risk-minimization technique and associated minimal martingale measure are implemented. The forward process is assumed to be driven by a Brownian motion and its drift term is modelled to be a function of the net position of the zone which is assumed to follow a finite-state Markov chain.

Two settings with respect to the information levels market participants observe are considered, as in Çetin and Verschuere (2009): complete information in which the forward price and the net position of the zone are observed continuously; and incomplete information in which the forward price is observed continuously and the net position of the zone is observed only at periods of emission announcement. The relationship in the complete information setting is derived and the analytical arbitrage-free price is obtained. The relationship in the incomplete information setting is also investigated, which leads to an application of a filtering technique. The net position of the zone is observed only through the fluctuations of the forward price and the actual value is periodically announced. Thus its projection onto the observable information is implemented. The setting where the unobserved process follows a Markov chain leads to the use of the Wonham filter. The arbitrage-free price is obtained using Monte Carlo simulation. We also consider parameter estimation. The framework is that of a hidden Markov model (HMM). A recursive expectation-maximization (EM)

algorithm is employed to obtain the parameters of the future price and the net position of the zone.

Several papers deal with characteristics of prices in carbon emissions markets. Carmona, Fehr, Hinz and Porchet (2009) studied cap–and–trade schemes in an equilibrium setting and gave several qualitative properties of equilibrium prices. Seifert, Uhrig-Homburg and Wagner (2008) assumed every market participant is risk-neutral or the existence of a representative agent with a logarithmic utility, thereby reducing the problem to that of a representative agent who aims to maximize the total profit of all agents. Chesney and Taschini (2009) constructed an endogenous model for describing the emission allowance spot price dynamics that accounts for the potential presence of asymmetric information in the market. Several papers investigate the price process from econometric viewpoints; Daskalakis, Psychoyios and Markellos (2009) studied spot prices and the stochastic convenience yield with several jump-diffusion processes. Benz and Trück (2009) suggested the use of AR-GARCH and Markov regime-switching models for explaining log returns of EUA prices. Goodness-of-fit tests on in-sample data and forecasting analysis with out-of-sample data were conducted. Paolella and Taschini (2006) undertook an econometric analysis of emission allowance spot market returns and found that the unconditional tails can be well represented by a Pareto distribution while the conditional dynamics can be approximated by a new GARCH-type structure. There is a growing number of papers in the area of stochastic filtering in finance. For example, Elliott and van der Hoek (1997) applied a filtering technique for the optimal asset allocation problem. Landén (2000) considered zero-coupon bond pricing in which the drift term in the underlying diffusion process is modulated with a Markov chain. Several papers use a filtering technique to model the imperfect information investors can observe. For example, Çetin, Jarrow, Protter and Yildirim (2004), Jeanblanc and Valchev (2005) and Duffie, Eckner, Horel and Saita (2009) considered incomplete information in reduced-form modelling of credit risk. Lakner (1998) and Sass and Haussmann (2004) applied filtering theory to portfolio optimization problems under partial information.

The outline of this chapter is as follows. After this introductory section, Section 3.2 introduces the models used for pricing. We present models for the

forward price process and the net position of the zone and then derive the spot and forward relationship under both complete and incomplete information settings. Estimated parameters are given in Section 3.3, along with the EM algorithm for an HMM. Numerical examples are provided in Section 3.4, where the price differences are highlighted in the two information settings with respect to each parameter. Of interest are the jumps in the process of the conditional expectation coming from the announcement of the actual value and consequent changes in the spot price. Concluding remarks and several possible extensions are discussed in Section 3.5.

## 3.2 Models for EUA Prices

The framework for exploring the spot and forward relationship in the EU ETS market is introduced in this section. For simplicity, assume two EUAs are traded in the market: an EUA for the current year denoted by EUA0, and an EUA for the following year denoted by EUA1. We will first introduce the underlying processes and subsequently define the equation which the EUA0 price satisfies.

### 3.2.1 The spot and forward relationship in EU ETS markets

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a stochastic space with a filtration $\mathcal{F}_t$ and all stochastic processes defined on this space are assumed to be $\mathbb{F}$-adapted, $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}$, where $T^*$ is a fixed but arbitrary time horizon. We adopt a similar approach to Çetin and Verschuere (2009). Let the positive-valued EUA1 price process, $S$, be a continuous-time process satisfying the stochastic differential equation (SDE),

$$dS_t = (\mu + \alpha \theta_t) S_t \, dt + \sigma S_t \, dW_t, \quad S_0 = s_0, \tag{3.1}$$

where $\mu$, $\alpha$ and $\sigma$ are assumed constant, $W_t$ is a one-dimensional standard Brownian motion. $\theta_t$ is assumed to be a finite-state Markov chain expressing the net position of the zone whose details are explained later. The pair, $(S_t, \theta_t)$, is a vector-valued Markov process. From equation (3.1) it follows that the drift term of the EUA1 price has a different expected mean depending on the value of the net position of the zone. The choice of the drift coefficient is meant to reflect the changing demand for allowances depending on the net position of the market.

EU ETS provides companies with fixed, free allowances on an annual basis at the beginning of each trading year and announces the total allowances submitted by the companies at the end of each trading year. Companies which receive allowances can emit $CO_2$ up to their allocated allowance and in the market companies can trade allowances to offset any excess or shortage, as explained in Section 3.1. While total allowances are capped by regulators, the total amount of emissions is uncertain so that market conditions are either in excess or in a shortage. The net position of the zone models the difference between emissions and allocations, thereby are certain to be either long or short. If the total allowances were over-allocated, or companies abated $CO_2$ emissions, the net position of the zone is long, meaning that the total allowances would be in excess. On the other hand, if numerous companies emit more than their allocations, or they abate less, the net position of the zone is short, meaning that the total allowances would result in a shortage.

The net position of the zone, $\theta_t$, is assumed to follow a continuous-time Markov chain taking values in $E := \{-1, 1\}$. $\theta_t = -1$ and $\theta_t = 1$ correspond to the net position of the zone at $t$ when short and long, respectively. $\theta_t$ is assumed to have the time-homogeneous generator matrix defined by

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_{-1} & -\lambda_{-1} \end{pmatrix}, \tag{3.2}$$

where $\lambda_1$ and $\lambda_{-1}$ are assumed to be positive constant. This indicates $\theta_t$ stays in each state for an exponentially-distributed amount of time, with the exponential distribution which determines the transition from state $-1$ to state $1$, having parameter $\lambda_1$ and that from state $1$ to state $-1$, $\lambda_{-1}$. We assume the Markov chain is independent of the Brownian motion.

The spot and forward relationship under the assumption of no banking, which was in effect in 2007 during the transition from Phase I to Phase II, was analysed in Çetin and Verschuere (2009). Under the assumption of no banking, the EUA0 price would go to zero as the expiry date approaches, provided that the net position of the zone is long. This is because there are companies which have allowances that more than cover actual emissions and the excess allowances are worthless after expiry, which therefore drives the price to zero. If the net position of the zone is short, EUA1 turns to EUA0 next year by paying a penalty, denoted

by $K$, for every excess tonne of emissions, because of regulations.

Considering the above points, the value of EUA0 at the end of Phase I, $T$, denoted by $P_T$ is assumed to satisfy the following relationship:

**Definition 3.1** *(Çetin and Verschuere (2009)). The EUA0 price at the end of Phase I, $T$, is defined as*

$$P_T = \begin{cases} S_T + K & \text{if} \quad \theta_T = -1, \\ 0 & \text{if} \quad \theta_T = 1. \end{cases}$$

In this chapter, we discuss the relationship when banking is in effect. Phase I consists of three years, from 2005-2007, and Phase II consists of five years, from 2008-2012, but for simplicity in what follows, we assume that there are only two periods to trade, one being $[0, T]$ and the other $[T, 2T]$, where $2T < T^*$, and there is no constraint on intra-period banking between $[0, T]$ and $[T, 2T]$, and $2T$ is the maturity of the phase when banking for next year is banned. At $2T$, since banking is not allowed, the unused permits are worthless provided the market is long. We assume that at $2T$ the analogous relationship will hold as at $T$ in Çetin and Verschuere (2009).

$$P_{2T} = \begin{cases} S_{2T} + K & \text{if} \quad \theta_{2T} = -1, \\ 0 & \text{if} \quad \theta_{2T} = 1. \end{cases}$$

We consider the relationship at $T$. Now that intra-period banking is in effect, unused permits are no longer worthless. Banking is in effect so that unused permits can be transferred and used later, provided that the market is long. We assume in this case that EUA0 is evaluated as the expectation of the future value. The price of an EUA0 contract, in the case that the market is short, is equal to that of EUA1 plus a penalty as the market is short at $2T$. The relation at $T$ is therefore

$$P_T = \begin{cases} S_T + K & \text{if} \quad \theta_T = -1, \\ \mathbb{E}\big[1_{\{\theta_{2T}=-1\}}(S_{2T} + K)\big|S_T, \theta_T = 1\big] & \text{if} \quad \theta_T = 1. \end{cases}$$

where $1_{\{\theta_t=i\}}$ is an indicator function which returns 1 if $\theta_t = i$, 0 otherwise and $\mathbb{E}[\cdot]$ is the expectation operator under a pricing measure explained later. Considering the relationship at $T$, it turns out that EUA0 is regarded as an option on EUA1 as well as on the net position of the zone. In the next two

subsections, we discuss the pricing method based on a local risk-minimizing approach.

## 3.2.2 Pricing under Complete Information

We consider no-arbitrage pricing in the sense that there exist equivalent martingale measures and the fair price is calculated as an expectation of future cash-flows under these measures. In this subsection, pricing is done under the complete information setting: market participants can access full information, $\mathbb{F}$.

The pricing problem in our setting is considered as that in the incomplete market. Incompleteness arises in the presence of the net position of the zone, since it is not tradable. The uncertainty in the EUA0 price caused by $W$ and $\theta$ cannot be hedged perfectly by trading only $S$. It means that there are cash-flows that are unable to be replicated by a self-financing trading strategy. The no-arbitrage assumption consequently provides infinitely many equivalent martingale measures, with each measure leading to a different price. Since there is no exact replication to provide a unique price and there are several equivalent martingale measures, we need to fix one equivalent martingale measure to price and hedge derivatives in this incomplete market setting. Currently there is no general agreement on how to choose a specific equivalent martingale measure towards pricing derivatives and a variety of alternatives are considered in the literature.

We adopt a local risk-minimizing strategy first proposed by Föllmer and Sondermann (1986) for the case where an underlying process is a martingale and extended to the case where an underlying process follows a general semi-martingale by Föllmer and Schweizer (1991). Under this approach, the change of measure technique renders tradable instruments as martingales and the orthogonal parts to tradable instruments remain unchanged. The associated martingale measure is called the minimal martingale measure. A more detailed description of the local risk-minimizing strategy and related topics can be found, for instance, in Schweizer (2001) or Föllmer and Schied (2004).

**Definition 3.2** *Let $X$ be a continuous semi-martingale with the canonical decomposition $X = X_0 + M + A$ where $M$ is a martingale and $A$ is adapted, continuous and of finite variation. A probability measure $\hat{\mathbb{P}}$, which is equivalent to $\mathbb{P}$, is called the minimal martingale measure if $X$ follows a martingale under $\hat{\mathbb{P}}$, $\hat{\mathbb{P}} = \mathbb{P}$ on $\mathcal{F}_0$ and any square integrable martingale orthogonal to $M$ remains a martingale under $\hat{\mathbb{P}}$.*

The minimal martingale measure, $\hat{\mathbb{P}}$, in our setting, is identified by the following Radon-Nikodým derivative,

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}} = \hat{Z}_{2T}, \tag{3.3}$$

where

$$\hat{Z}_t = \exp\left(-\int_0^t \frac{\mu + \alpha\theta_s}{\sigma}\, dW_s - \frac{1}{2}\int_0^t \left(\frac{\mu + \alpha\theta_s}{\sigma}\right)^2 ds\right), \quad 0 \le t \le 2T.$$

The Girsanov-Meyer theorem (see for example Protter (2004, Theorem 39)) yields $\hat{W}$ defined as

$$\hat{W}_t = W_t + \int_0^t \frac{\mu + \alpha\theta_s}{\sigma}\, ds,$$

where $\hat{W}_t$ is a standard Brownian motion under $\hat{\mathbb{P}}$, for $0 \le t \le 2T$.

The next step is to derive the underlying processes under the minimal martingale measure. The EUA1 price process under $\hat{\mathbb{P}}$ follows, along with $\hat{W}$, the SDE,

$$\begin{aligned}
dS_t &= (\mu + \alpha\theta_t)S_t\, dt + \sigma S_t\left(d\hat{W}_t - \frac{\mu + \alpha\theta_t}{\sigma}\, dt\right)\\
&= \sigma S_t\, d\hat{W}_t, \quad S_0 = s_0. 
\end{aligned} \tag{3.4}$$

Since $\theta$ is assumed to be independent of $W$ and the minimal martingale measure leaves the orthogonal martingale to $W$ unchanged, the law for $\theta_t$ is invariant after the change of measure. For any state $i \in E$, we denote $(Q)(i) = \sum_{j\in E} \lambda_{ij} j$, where $\lambda_{ij}$ is the $(i,j)$-th entry in $Q$. The continuous-time Markov chain is characterised by a semi-martingale representation (see for example Jacod (1975)) as

$$\theta_t = \theta_0 + A_t + N_t,$$
$$A_t = \int_0^t (Q_s)(\theta_s)\, ds = \int_0^t \left((1 - \theta_s)\lambda_1 + (-1 - \theta_s)\lambda_{-1}\right) ds,$$

where $A_t$ is a finite variation process with $A_0 = 0$ and $N_t$ is a martingale process with $N_0 = 0$ and this decomposition is unique.

Since we adopt the minimal martingale measure approach, the unused permit at $2T$ is evaluated under this measure. We can now give a precise definition of the EUA0 price under complete information.

**Definition 3.3** *Under the minimal martingale measure, the EUA0 price in Phase II contracts at $T$ is defined as*

$$P_T = \begin{cases} S_T + K & \text{if } \theta_T = -1, \\ \hat{\mathbb{E}}\big[1_{\{\theta_{2T}=-1\}}(S_{2T} + K)\big| S_T, \theta_T = 1\big] & \text{if } \theta_T = 1, \end{cases} \tag{3.5}$$

*where $\hat{\mathbb{E}}[\cdot]$ is the expectation operator under $\hat{\mathbb{P}}$.*

With the minimal martingale measure, derivatives prices are determined from Föllmer and Schweizer (1991) as follows:

**Theorem 3.1** *(Theorem 3.14 in Föllmer and Schweizer (1991))*
*Let $C \in L^2(\Omega, \mathcal{F}_{2T}, \mathbb{P})$ and let $\hat{\mathbb{P}}$ be the unique minimal martingale measure for $S$ given by equation (3.3). Then there exists a unique $\xi^C$ such that*

$$V_t = \hat{\mathbb{E}}[C] + \int_0^t \xi_s^C \, dS_s + L_t^C,$$

*where the process $L^C = \{L_t^C, 0 \le t \le 2T\}$ is a square integrable $\hat{\mathbb{P}}$-martingale with $L_0^C = 0$ and orthogonal to $S$.*

See, for the proof, Föllmer and Schweizer (1991). In our problem, $C$ is defined from equation (3.5) as

$$C = 1_{\{\theta_T=-1\}}(S_T + K) + 1_{\{\theta_T=1\}}\hat{\mathbb{E}}\big[1_{\{\theta_{2T}=-1\}}(S_{2T} + K) \mid S_T, \theta_T = 1\big].$$

The requirement of $\mathcal{F}_T$- and $\mathcal{F}_{2T}$-measurability and square integrability for the payoff $C$ is met in our setting.

Before stating the main result, we introduce a fundamental property with respect to the Markov chain.

**Proposition 3.1** *The process $M$, given by the formula*

$$M_t = \left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\lambda_1 + \lambda_{-1}}\right)e^{(\lambda_1 + \lambda_{-1})t},$$

*follows a $\hat{\mathbb{P}}$-martingale orthogonal to $S$.*

**Proof** $M_t$ follows a $\hat{\mathbb{P}}$-martingale since

$$\left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\lambda_1 + \lambda_{-1}}\right)e^{(\lambda_1 + \lambda_{-1})t} = \left(\theta_0 - \frac{\lambda_1 - \lambda_{-1}}{\lambda_1 + \lambda_{-1}}\right) + \int_0^t e^{(\lambda_1 + \lambda_{-1})s}\,d\theta_s$$

$$+ \int_0^t \left(\theta_s - \frac{\lambda_1 - \lambda_{-1}}{\lambda_1 + \lambda_{-1}}\right)(\lambda_1 + \lambda_{-1})e^{(\lambda_1 + \lambda_{-1})s}\,ds$$

$$= \left(\theta_0 - \frac{\lambda_1 - \lambda_{-1}}{\lambda_1 + \lambda_{-1}}\right) + \int_0^t e^{(\lambda_1 + \lambda_{-1})s}\,dN_s.$$

The last equation follows from the semi-martingale representation,

$$d\theta_t = (-(\lambda_1 + \lambda_{-1})\theta_t + \lambda_1 - \lambda_{-1})\,dt + dN_t.$$

Since the minimal martingale measure preserves orthogonality and $\theta$ is independent of $W$, the process $M$ is orthogonal to $\hat{W}$. $\qquad\square$

From Proposition 3.1, as well as equation (3.4), it follows that the EUA0 price under the complete information setting is calculated as per Theorem 3.2.

**Theorem 3.2** *The EUA0 contract price at $t < T$, (3.5), under complete information is given by*

$$P_t = (S_t + K)\left(1 - \frac{1}{2}\left(\frac{2\lambda_1}{\bar{\lambda}} + \left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)e^{-\bar{\lambda}(T-t)}\right)\left(\frac{\lambda_1}{\bar{\lambda}} + \frac{\lambda_{-1}}{\bar{\lambda}}e^{-\bar{\lambda}T}\right)\right),$$

$$(3.6)$$

*where $\bar{\lambda} = \lambda_1 + \lambda_{-1}$.*

**Proof** With the minimal martingale approach, the EUA0 prices can be written as

$$P_t = \hat{\mathbb{E}}\left[(S_T + K)1_{\{\theta_T = -1\}} + \hat{\mathbb{E}}\left[1_{\{\theta_{2T} = -1\}}(S_{2T} + K)\big|\, S_T, \theta_T = 1\right]1_{\{\theta_T = 1\}}\big|\,\mathcal{F}_t\right].$$

Since $\theta_t$ takes the value either 1 or $-1$ at $t = T$ and $2T$, the EUA0 price is written as

$$P_t = \hat{\mathbb{E}}\left[(S_T + K)\frac{1 - \theta_T}{2} + \hat{\mathbb{E}}\left[\frac{1 - \theta_{2T}}{2}(S_{2T} + K)\big|\, S_T, \theta_T = 1\right]\frac{1 + \theta_T}{2}\big|\,\mathcal{F}_t\right].$$

The inner expectation can be written as

$$\hat{\mathbb{E}}\left[\frac{1-\theta_{2T}}{2}(S_{2T}+K)\Big|\, S_T, \theta_T = 1\right]$$

$$= \hat{\mathbb{E}}\left[\frac{S_{2T}+K}{2}\left(1 - M_{2T}e^{-2\bar{\lambda}T} - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)\Big|\, S_T, \theta_T = 1\right]$$

$$= \frac{S_T + K}{2}\left(1 - \left(1 - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)e^{-\bar{\lambda}T} - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)$$

$$= (S_T + K)\frac{\lambda_{-1}}{\bar{\lambda}}(1 - e^{-\bar{\lambda}T}). \tag{3.7}$$

The second equality comes from the fact that when $S_{2T}$ and $M_{2T}$ are orthogonal martingales and conditioned on $\theta_T = 1$, $M_T = (1-(\lambda_1-\lambda_{-1})/\bar{\lambda})e^{\bar{\lambda}T}$. The EUA0 price is, therefore, given by

$$P_t = \hat{\mathbb{E}}\left[(S_T+K)\frac{1-\theta_T}{2} + (S_T+K)\frac{\lambda_{-1}}{\bar{\lambda}}(1 - e^{-\bar{\lambda}T})\frac{1+\theta_T}{2}\Big|\, \mathcal{F}_t\right]$$

$$= \hat{\mathbb{E}}\left[(S_T+K)\left(1 - \frac{1+\theta_T}{2}\right) + (S_T+K)\frac{\lambda_{-1}}{\bar{\lambda}}(1 - e^{-\bar{\lambda}T})\frac{1+\theta_T}{2}\Big|\, \mathcal{F}_t\right]$$

$$= \hat{\mathbb{E}}\left[(S_T+K) + (S_T+K)\frac{1+\theta_T}{2}\left(-1 + \frac{\lambda_{-1}}{\bar{\lambda}}(1 - e^{-\bar{\lambda}T})\right)\Big|\, \mathcal{F}_t\right].$$

Noting

$$\hat{\mathbb{E}}\left[(S_T+K)\frac{1+\theta_T}{2}\right] = \hat{\mathbb{E}}\left[\frac{S_T+K}{2}\left(1 + M_Te^{-\bar{\lambda}T} + \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)\Big|\, \mathcal{F}_t\right]$$

$$= \frac{S_t + K}{2}\left(\frac{2\lambda_1}{\bar{\lambda}} + M_te^{-\bar{\lambda}T}\right)$$

$$= \frac{S_t + K}{2}\left(\frac{2\lambda_1}{\bar{\lambda}} + \left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)e^{-\bar{\lambda}(T-t)}\right), \tag{3.8}$$

the EUA0 price is then

$$P_t = (S_t + K)\left(1 - \frac{1}{2}\left(\frac{2\lambda_1}{\bar{\lambda}} + \left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)e^{-\bar{\lambda}(T-t)}\right)\left(\frac{\lambda_1}{\bar{\lambda}} + \frac{\lambda_{-1}}{\bar{\lambda}}e^{-\bar{\lambda}T}\right)\right).$$

$\square$

Note that

$$\frac{\lambda_{-1}}{\bar{\lambda}}(1 - e^{-\bar{\lambda}T})$$

in equation (3.7) and

$$\frac{1}{2}\left(\frac{2\lambda_1}{\bar{\lambda}} + \left(\theta_t - \frac{\lambda_1 - \lambda_{-1}}{\bar{\lambda}}\right)e^{-\bar{\lambda}(T-t)}\right)$$

in equation (3.8) correspond to the probability $\hat{\mathbb{P}}(\theta_{2T} = -1 \mid \theta_T = 1)$ and $\hat{\mathbb{P}}(\theta_T = 1 \mid \theta_t)$, $\theta_t \in E$, respectively. Since the EUA0 price is expressed as

$$P_t = (S_t + K)\left(1 - \hat{\mathbb{P}}(\theta_{2T} = 1 \mid \theta_T = 1)\hat{\mathbb{P}}(\theta_T = 1 \mid \theta_t)\right),$$

the extension to a multi-period setting is straightforward. The details of several possible extensions are discussed in Section 3.5.

### 3.2.3 Pricing under Incomplete Information

In this subsection, we consider the pricing problem in a more realistic setting. The process $\theta_t$ is, in fact, not always observable to market participants. The information available to market participants is that generated by $S$ and only at $T$ and $2T$, which are announcement times, by $\theta$. By setting up a new filtration in such a way that information for $\theta_t$ is included only at $T$ and $2T$ and pricing is done with the new filtration, this aspect is included in the pricing. We define the new filtration, $\mathcal{G}$, expressing this situation where $\mathcal{G}$ is defined as

$$\mathcal{G}_t = \begin{cases} \mathcal{F}_t^S & \text{for } t < T, \\ \mathcal{F}_t^S \vee \sigma(\theta_T) & \text{for } T \leq t < 2T, \\ \mathcal{F}_t^S \vee \sigma(\theta_T) \vee \sigma(\theta_{2T}) & \text{for } t = 2T. \end{cases}$$

The new filtration $\mathcal{G}$ represents the following:

– At $t < T$, only $S_t$ is observed.

– At $t = T$, both $S_t$ and $\theta_t$ are observed.

– At $T < t < 2T$, only $S_t$ is observed.

– At $t = 2T$, both $S_t$ and $\theta_t$ are observed.

Due to the definition of filtration (see, for example, Section 1.1 in Protter (2004)), it has an increasing family of $\sigma$-algebras, meaning that the filtration $\mathcal{G}_t$ includes all information before $t$. In other words, once observed, the values of $\theta_T$ and $\theta_{2T}$ are held in the filtration.

$\theta$ is no longer an adapted process with respect to $\mathcal{G}$. It is however partially observed through the fluctuations of $S$, from equation (3.1), thereby allowing use of the projection of $\theta$. Under this setting, the distribution of $\theta$ restricted

to $\mathcal{G}$ is implemented. For $0 \leq t \leq 2T$, the conditional expectation process of $\theta$ under $\mathcal{G}$, denoted by $\bar{\theta}$, is defined by

$$\bar{\theta}_t = \mathbb{E}[\theta_t \mid \mathcal{G}_t].$$

We consider the spot and forward relationship under the new filtration $\mathcal{G}$ with $\bar{\theta}$. We start by deriving the dynamics of $\bar{\theta}$.

**Theorem 3.3** *Define the innovation process* $\overline{W}$ *by*

$$\overline{W}_t = \int_0^t \frac{dS_s - (\mu + \alpha\bar{\theta}_s)S_s \, ds}{\sigma S_s}. \tag{3.9}$$

*Then,* $\overline{W}$ *is a* $\mathcal{G}$*-Brownian motion.*

**Proof** From standard filtering theory, see for example Liptser and Shiryaev (2001, Theorem 8.1), $\overline{W}_t$ is a $\mathcal{G}$-Brownian motion in $0 \leq t < T$ and in $T < t < 2T$. $\overline{W}_t$ stays a martingale at $T$ since for $t < T$,

$$\mathbb{E}[\overline{W}_T \mid \mathcal{G}_t] = \overline{W}_t.$$

By Lévy's characterisation theorem of Brownian motion, see for example (Protter, 2004, Theorem 39), $\overline{W}$ is a $\mathcal{G}$-Brownian motion at $T$. An analogous relation holds for $\overline{W}$ at $2T$. $\qquad\qquad\square$

Using the $\mathcal{G}$-Brownian motion $\overline{W}_t$, the EUA1 price process is written as

$$dS_t = (\mu + \alpha\bar{\theta}_t)S_t \, dt + \sigma S_t \, d\overline{W}_t, \quad S_0 = s_0. \tag{3.10}$$

The SDE for $\bar{\theta}_t$ is derived from the conditional probability of $\theta_t$ on $\mathcal{G}_t$. Define $\pi_i(t) := \mathbb{P}(\theta_t = i \mid \mathcal{G}_t)$, for any $i \in E$, which is given for $0 \leq t < T$ by

$$\pi_i(t) = \pi_i(0) + \int_0^t \lambda_k - (\lambda_1 + \lambda_{-1})\pi_i \, du + \int_0^t \frac{2i\alpha}{\sigma}\pi_i(1 - \pi_i) \, d\overline{W}_t. \tag{3.11}$$

and for $T \leq t < 2T$ by

$$\pi_i(t) = \delta_i(T) + \int_T^t \lambda_k - (\lambda_1 + \lambda_{-1})\pi_i \, du + \int_T^t \frac{2i\alpha}{\sigma}\pi_i(1 - \pi_i) \, d\overline{W}_t. \tag{3.12}$$

where $\{k\} = E\backslash\{i\}$ and $\delta_i(T) = 1$ if $\theta_T = i$, 0 otherwise. See for example Liptser and Shiryaev (2001, Theorems 9.1 and 9.3) for the derivation and related topics.

Note that at $T$ the information with respect to $\theta_T$ is revealed and therefore $\pi_i$ will jump and these processes follow the same SDE with different initial values. This system is referred to as the Wonham filter. The dynamics of $\bar{\theta}_t$ are therefore given from equations (3.11) and (3.12) and $\pi_1(t) + \pi_{-1}(t) = 1$ that for $0 \leq t < T$,

$$d\bar{\theta}_t = -\big(\lambda_1(\bar{\theta}_t - 1) + \lambda_{-1}(\bar{\theta}_t + 1)\big)\,dt + \frac{\alpha}{\sigma}(1 - \bar{\theta}_t^2)\,d\overline{W}_t, \quad \bar{\theta}_0 = 2\pi_1(0) - 1,$$

$$(3.13)$$

and for for $T \leq t < 2T$,

$$d\bar{\theta}_t = -\big(\lambda_1(\bar{\theta}_t - 1) + \lambda_{-1}(\bar{\theta}_t + 1)\big)\,dt + \frac{\alpha}{\sigma}(1 - \bar{\theta}_t^2)\,d\overline{W}_t, \quad \bar{\theta}_T = \theta_T, \quad (3.14)$$

where the initial values are given by $\bar{\theta}_0 = \pi_1(0) - \pi_{-1}(0) = 2\pi_1(0) - 1$ and at $T$ $\bar{\theta}_T = \delta_1(T) - \delta_{-1}(T) = \theta_T$. Since the actual value of the net position of the zone is revealed at time $T$, $\bar{\theta}_t$ will jump to $\theta_t$ at $t = T$. Also at time $2T$, $\bar{\theta}_{2T}$ will jump to $\theta_{2T}$.

Now we consider pricing under $\mathcal{G}$. Since $\bar{\theta}$ is not tradable, this pricing is again considered as being in an incomplete market and the local risk-minimizing approach is capable of dealing with this setting. The minimal martingale measure $\hat{\mathbb{P}}$ used in the complete information setting would not coincide with the martingale measure with respect to the new filtration $\mathcal{G}$. However, there is still a unique minimal martingale measure $\mathbb{P}^*$, which is to be distinguished from $\hat{\mathbb{P}}$. The Radon-Nikodým derivative is given by

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = Z_{2T}^*,$$

where

$$Z_t^* = \exp\left(-\int_0^t \frac{\mu + \alpha\bar{\theta}_s}{\sigma}\,d\overline{W}_s - \frac{1}{2}\int_0^t \left(\frac{\mu + \alpha\bar{\theta}_s}{\sigma}\right)^2 dt\right), \quad 0 \leq t \leq 2T.$$

From Girsanov-Meyer theorem, under $\mathbb{P}^*$ the process $W_t^*$ is a $\mathcal{G}$-Brownian motion where

$$W_t^* = \overline{W}_t + \int_0^t \frac{\mu + \alpha\bar{\theta}_s}{\sigma}\,ds. \tag{3.15}$$

Underlying processes under $\mathbb{P}^*$ are written, along with $W^*$, as

$$dS_t = \sigma S_t \, dW_t^*, \quad S_0 = s_0, \tag{3.16}$$

$$d\bar{\theta}_t = -\left(\lambda_1(\bar{\theta}_t - 1) + \lambda_{-1}(\bar{\theta}_t + 1) + \frac{\alpha}{\sigma^2}(1 - \bar{\theta}_t^2)(\mu + \alpha\bar{\theta}_t)\right) dt$$
$$+ \frac{\alpha}{\sigma}(1 - \bar{\theta}_t^2) \, dW_t^*, \tag{3.17}$$

$$\bar{\theta}_0 = 2\pi_1(0) - 1, \tag{3.18}$$

$$\bar{\theta}_T = \theta_T, \tag{3.19}$$

$$\bar{\theta}_{2T} = \theta_{2T}. \tag{3.20}$$

The EUA0 price is defined under the filtration $\mathcal{G}$ as follows:

**Proposition 3.2** *The price of the EUA0 contract at time $t < T$ under incomplete information is written as*

$$P_t = \mathbb{E}^*\left[1_{\{\theta_T=-1\}}(S_T + K) + 1_{\{\theta_T=1\}} \mathbb{E}^*\left[1_{\{\theta_{2T}=-1\}}(S_{2T} + K)\big| S_T, \theta_T = 1\right] \big| \mathcal{G}_t\right], \tag{3.21}$$

*where $\mathbb{E}^*[\cdot]$ is the expectation operator under $\mathbb{P}^*$.*

Calculation of equation (3.21), along with the processes (3.16)-(3.20), requires the joint distribution of $(S_T, \theta_T)$ given $\mathcal{G}_t$ and that of $(S_{2T}, \theta_{2T})$ given $\mathcal{G}_T$ since $\bar{\theta}_t$ will be equal to $\theta_t$ at $T$ and $2T$ and these values are used.

The distribution of $\theta_T$ given $\mathcal{G}_t$, $0 \leq t < T$, under the minimal martingale measure is given as

$$\pi_i(T, t) := \mathbb{P}^*(\theta_T = i \mid \mathcal{G}_t)$$
$$= \mathbb{E}^*[1_{\{\theta_T=i\}} \mid \mathcal{G}_t]$$
$$= \mathbb{E}^*[\mathbb{E}^*[1_{\{\theta_T=i\}} \mid \mathcal{F}_T^S] \mid \mathcal{G}_t], \quad \forall i \in E.$$

The second equality comes from the definition of $\mathcal{G}_t$, $t < T$ and the tower property. Since the inner expectation is written as the limit of $\pi(t)$ as

$$\mathbb{E}^*\left[1_{\{\theta_T=i\}} \mid \mathcal{F}_T^S\right] = \lim_{u \to T} \mathbb{E}^*\left[1_{\{\theta_u=i\}} \mid \mathcal{F}_u^S\right] = \lim_{u \to T} \mathbb{E}^*\left[1_{\{\theta_u=i\}} \mid \mathcal{G}_u\right] = \lim_{u \to T} \pi_i(u).$$

Consequently, the distribution is given as

$$\pi_i(T, t) = \mathbb{E}^*[\lim_{u \to T} \pi_i(u) \mid \mathcal{G}_t].$$

83

Analogous relations hold for the distribution of $\theta_{2T}$ given $\mathcal{G}_T$ while the process $\pi_i(t)$ for $T \le t < 2T$ is governed by equation (3.12).

It turns out that these extrapolating distributions are obtained simply by simulating the conditional probability. For $0 \le t < T$, $\pi_i(t)$ under $\mathbb{P}^*$ is given from equations (3.11) and (3.15) as,

$$d\pi_i(t) = \left( \lambda_k - (\lambda_1 + \lambda_{-1})\pi_i(t) - \frac{2\left(\mu + (\pi_i(t) + \pi_k(t))\,\alpha\right)i\alpha}{\sigma^2}\pi_i(t)(1 - \pi_i(t)) \right) dt$$
$$+ \frac{2i\alpha}{\sigma}\pi_i(t)\left(1 - \pi_i(t)\right) dW_t^*, \quad \pi_i(0) = \mathbb{P}^*(\theta_0 = i \mid \mathcal{G}_0), \quad (3.22)$$

and for $T \le t < 2T$, it is given from equations (3.12) and (3.15) as,

$$d\pi_i(t) = \left( \lambda_k - (\lambda_1 + \lambda_{-1})\pi_i(t) - \frac{2\left(\mu + (\pi_i(t) + \pi_k(t))\,\alpha\right)i\alpha}{\sigma^2}\pi_i(t)(1 - \pi_i(t)) \right) dt$$
$$+ \frac{2i\alpha}{\sigma}\pi_i(t)\left(1 - \pi_i(t)\right) dW_t^*, \quad \pi_i(T) = \delta_i(T). \quad (3.23)$$

$\pi_i(t)$, for any $i \in E$, generally depends not only on $\pi_i(t)$ itself, but also on all other states, $\pi_k(t)$, $\{k\} = E \setminus \{i\}$. It therefore requires the simultaneous simulation of all states in $E$. This simulation is, however, simplified in our setting. Since $E$ consists of two states, a sample path for $\pi_i(t)$ is simply computed from that for $\bar{\theta}_t$. Simulation of $\bar{\theta}_t$ is thus implemented rather than that of $\pi_i(t)$. After simulating $\bar{\theta}_T$ up to $T-$, we use the relation that $\pi_i(T) = (i\bar{\theta}_{T-} + 1)/2$.

Consider solving numerically the expectation (3.21) with Monte Carlo simulation. We partition $[0, 2T]$ into $N$ sub-intervals of equal length $\Delta t = 2T/N$ and over each sub-interval $[t, t + \Delta t]$ simulate a process after the discretization as follows:

$$S_{t+\Delta t} = S_t \exp\left(-\frac{\sigma^2}{2}\Delta t + \sigma(W_{t+\Delta t}^* - W_t^*)\right), \quad (3.24)$$
$$\bar{\theta}_{t+\Delta t} = \bar{\theta}_t - \left( \lambda_1(\bar{\theta}_t - 1) + \lambda_{-1}(\bar{\theta}_t + 1) + \frac{\alpha}{\sigma^2}(1 - \bar{\theta}_t^2)(\mu + \alpha\bar{\theta}_t) \right)\Delta t$$
$$+ \frac{\alpha}{\sigma}(1 - \bar{\theta}_t^2)(W_{t+\Delta t}^* - W_t^*), \quad (3.25)$$

where $W_{t+\Delta t}^* - W_t^*$ follows a normal distribution with mean 0 and variance $\Delta t$. The Monte Carlo algorithm is summarized below:

**Step 1:** Simulate $S_t$ and $\bar{\theta}_t$ from $t = 0$ to $t = T$ using equations (3.24) and (3.25). The initial value for $S$ is $s_0$ and for $\bar{\theta}$ is set to a uniform random number $\xi$ by means of $2\xi - 1$.

**Step 1.1:** At time $t$, generate a normal random number $\epsilon$ and compute $S$ and $\bar{\theta}$ for time $t + \Delta t$. Note that both processes are driven by the same Brownian motion, the same $\epsilon$ is used for computing both $S_{t+\Delta t}$ and $\bar{\theta}_{t+\Delta t}$.

**Step 2:** Simulate $\theta_T$ and set $\bar{\theta}_T$ as follows:

**Step 2.1:** Generate a uniform random number $\xi$ and set $\theta_T$ by means of $\theta_T = 1$ if $\xi < (\bar{\theta}_{T-} + 1)/2$ and $\theta_T = -1$ otherwise.

**Step 2.2:** Set $\bar{\theta}_T = \theta_T$.

**Step 3:** Simulate $S_t$ and $\bar{\theta}_t$ from $t = T$ to $t = 2T$ using equations (3.24) and (3.25) in a similar manner to Step 1. The initial values are $S_T$ which is simulated in Step 1 and $\bar{\theta}_T$ set in Step 2 respectively.

**Step 4:** Simulate $\theta_{2T}$ and set $\bar{\theta}_T$ as follows:

**Step 4.1:** Generate a uniform random number $\xi$ and set $\theta_{2T}$ by means of $\theta_{2T} = 1$ if $\xi < (\bar{\theta}_{2T-} + 1)/2$ and $\theta_{2T} = -1$ otherwise.

**Step 4.2:** Set $\bar{\theta}_{2T} = \theta_{2T}$.

**Step 5:** Calculate the EUA0 price, $P_i$, from equation (3.21).

**Step 6:** Repeat Step 1 to Step 5 independently $M$ times and obtain independent samples $P_i$, $i = 1, 2, \ldots, M$. The EUA0 price is then calculated by the experimental average,

$$EUA0^{MC} = \frac{1}{M} \sum_{i=1}^{M} P_i.$$

Extension to a multi-period setting is straightforward by repeating Steps 3 and 4. The details of several possible extensions are discussed in Section 3.5.

## 3.3 Parameter Estimation

In this section, we estimate the parameters of the proposed model. There are two means of estimating parameters: one is with historical data, the other is with

contemporaneous day-to-day market data. In the former approach, parameters are estimated from time-series of associated data. An alternative is to estimate parameters with the data of a contemporaneous day, or implied data. In this approach, parameters are estimated with derivatives prices of the day. Derivatives prices are functions of variables so that the parameters are implied from their prices. Since there are not enough liquid derivatives markets for estimating all parameters, we estimate all parameters from historical data and use them for pricing.

Since an investor can observe only the EUA1 prices, we are in the classic situation of an HMM with the observable process $S_t$ and unobservable process $\theta_t$. Given the setting that the unobservable process follows a finite-state Markov chain and the conditional distribution of the log increment of the observable variable follows a normal distribution, it is possible to estimate parameters by maximizing the likelihood function. The unconditional distribution of $S_t$ follows a kind of mixture of normal distributions, but not a simple mixture because the unobservable state follows a Markov chain. In an HMM, the most popular method for estimating parameters would be the EM algorithm, which we apply here. After reviewing the EM algorithm for finding maximum likelihood estimators, we estimate parameters.

To apply the EM algorithm, we derive the discrete-time version of the SDEs. Partition the interval $[0, T]$ at the points $0 = t_0 < t_1 < \cdots < t_N = T$ where $t_k = k\Delta$, for $k = 0, 1, \ldots, N$, and $\Delta = T/N$ is the length of the discrete time interval. First we consider the transition probabilities computed with the the infinitesimal generator. Suppose the Markov chain has the transition matrix, denoted by $R$, with entries $R_{ij}(t) := \mathbb{P}(\theta_t = j \mid \theta_0 = i)$. The transition probabilities, along with the infinitesimal generator, satisfy the following Kolmogorov forward equation,

$$\frac{dR(t)}{dt} = R(t)Q, \quad R(0) = I,$$

or

$$\begin{pmatrix} R'_{-1,-1}(t) & R'_{-1,1}(t) \\ R'_{1,-1}(t) & R'_{1,1}(t) \end{pmatrix} = \begin{pmatrix} R_{-1,-1}(t) & R_{-1,1}(t) \\ R_{1,-1}(t) & R_{1,1}(t) \end{pmatrix} \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_{-1} & -\lambda_{-1} \end{pmatrix},$$

where $I$ is the identity matrix. Since $R_{-1,1}(t) = 1 - R_{-1,-1}(t)$ and $R_{1,-1}(t) =$

$1 - R_{1,1}(t)$, the forward equations simplify to

$$R'_{-1,-1}(t) = -(\lambda_1 + \lambda_{-1})R_{-1,-1}(t) + \lambda_{-1}, \quad R_{-1,-1}(0) = 1,$$
$$R'_{1,1}(t) = -(\lambda_1 + \lambda_{-1})R_{1,1}(t) + \lambda_1, \quad R_{1,1}(0) = 1.$$

These ordinary differential equations are solved as,

$$R_{-1,-1}(t) = \frac{\lambda_{-1}}{\lambda_1 + \lambda_{-1}} + \frac{\lambda_1}{\lambda_1 + \lambda_{-1}} e^{-(\lambda_1 + \lambda_{-1})t}, \tag{3.26}$$

$$R_{-1,1}(t) = 1 - R_{-1,-1}(t),$$

$$R_{1,-1}(t) = 1 - R_{1,1}(t),$$

$$R_{1,1}(t) = \frac{\lambda_1}{\lambda_1 + \lambda_{-1}} + \frac{\lambda_{-1}}{\lambda_1 + \lambda_{-1}} e^{-(\lambda_1 + \lambda_{-1})t}. \tag{3.27}$$

Next, we consider the log price of the EUA1 process (3.1). Applying Ito's formula yields

$$d\log S_t = \left(\mu + \alpha\theta_t - \frac{1}{2}\sigma^2\right) dt + \sigma \, dW_t.$$

The log increment of the EUA1 price in discrete-time is thus defined as

$$y_n := \log S_{n+1} - \log S_n = \left(\mu + \alpha\theta_n - \frac{1}{2}\sigma^2\right)\Delta + \sigma\left(W_{n+1} - W_n\right), \tag{3.28}$$

where $W_{n+1} - W_n$ follows a normal distribution with mean 0 and variance $\Delta$. Thus, the conditional increment follows a Gaussian distribution. Hence, the family of parameters for the HMM are

$$\left(R(\Delta), \, N\left(\left(\mu + \alpha\theta - \frac{1}{2}\sigma^2\right)\Delta, \sigma^2\Delta\right), p\right), \tag{3.29}$$

representing the transition probabilities, $R$, output probabilities which follow the normal distribution, $N((\mu + \alpha\theta - 1/2\sigma^2)\Delta, \sigma^2\Delta)$, and initial state probabilities, $p_i = \mathbb{P}(\theta_0 = i)$, for any $i \in E$. This system generates sequences of observations in the following way: the initial state of the unobservable variable is determined probabilistically based on $p$. At each time step, the system produces an unobservable state $\theta_k$ from $\theta_{k-1}$ one step before, according to the transition probability. Once $\theta_k$ is determined, the observation is produced according to the output probability distribution.

### 3.3.1 EM Algorithm

We explain the EM algorithm used to estimate the parameters of the models. For HMM or models with incomplete data, in general, the EM algorithm is more popular than gradient ascent approaches. The EM algorithm originally developed by Dempster, Laird and Rubin (1977) is a numerical method for calculating maximum likelihood estimates of partially observed models. It computes the incomplete data log-likelihood via the iterative maximization of the expected complete data log-likelihood, conditional on the observed data. It has the appealing property that an iteration does not decrease the value of the likelihood function. Convergence results for the EM algorithm under general conditions are investigated in Wu (1983). Suppose that we observe data $y_0, \dots, y_N$ (or in short, $y_{0:N}$), which are generated from the unobservable sequence $\theta_{0:N}$.

Let $\Lambda$ be a family of parameters. Our aim is to estimate parameters $\lambda \in \Lambda$ so as to maximize the log-likelihood function of the data $y_{0:N}$,

$$l(y_{0:N} \mid \lambda) := \log \mathbb{P}(y_{0:N} \mid \lambda).$$

Since there is an unobservable variable $\theta$ in our setting, maximizing the log-likelihood is not straightforward. The maximizing log-likelihood function has no closed-form solutions. To avoid numerical routines, we employ the EM algorithm for finding the parameters. The EM algorithm does not implement numerical routine so as to maximize the likelihood function. Instead of maximizing directly the log-likelihood function including only observable variables, in the EM algorithm, the complete data log likelihood function including both observable and unobservable variables is maximized. Here, we explain how the EM algorithm is connected to maximizing the log-likelihood. From the definition of conditional probability, trivially we have the following relation:

$$\mathbb{P}(y_{0:N} \mid \lambda) = \frac{\mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda)}{\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda)}.$$

By taking logarithms, we have

$$\log \mathbb{P}(y_{0:N} \mid \lambda) = \log \mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda) - \log \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda). \tag{3.30}$$

Taking the expectation with respect to $\theta_{0:N} | y_{0:N}, \lambda_n$ in which $\lambda_n$ are the param-

eters in the $n$-th iteration,

$$\int \log \mathbb{P}(y_{0:N} \mid \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n) =$$

$$\int \log \mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n) - \int \log \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n).$$

Since the left-hand-side of equation (3.30) is independent of $\theta_{0:N}$, it is simply written as

$$\int \log \mathbb{P}(y_{0:N} \mid \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n) = \log \mathbb{P}(y_{0:N} \mid \lambda).$$

For the right-hand-side, we define:

$$L(\lambda, \lambda_n) := \int \log \mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n),$$

$$H(\lambda, \lambda_n) := \int \log \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda) \, d\mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n).$$

Overall,

$$\log \mathbb{P}(y_{0:N} \mid \lambda) = L(\lambda, \lambda_n) - H(\lambda, \lambda_n).$$

We seek $\lambda_{n+1}$ which attains $\log \mathbb{P}(y_{0:N} \mid \lambda_{n+1}) \geq \log \mathbb{P}(y_{0:N} \mid \lambda_n)$, or equivalently,

$$L(\lambda_{n+1}, \lambda_n) - L(\lambda_n, \lambda_n) - (H(\lambda_{n+1}, \lambda_n) - H(\lambda_n, \lambda_n)) \geq 0.$$

The essence of the EM algorithm is that it can be shown by using Jensen's inequality, that for any $\lambda_{n+1}$,

$$H(\lambda_{n+1}, \lambda_n) - H(\lambda_n, \lambda_n) \leq 0.$$

For the derivation of the EM algorithm, see for example Tanner (1993). Therefore, the inequality $\log \mathbb{P}(y_{0:N}|\lambda_{n+1}) \geq \log \mathbb{P}(y_{0:N}|\lambda_n)$ is guaranteed by choosing

$$\lambda_{n+1} = \arg\max_{\lambda \in \Lambda} L(\lambda, \lambda_n).$$

Since $L$ represents a complete data log-likelihood, finding parameters that maximize the log-likelihood function is replaced by repeatedly maximizing complete data log-likelihood function. The algorithm consisting of an E step and an M step at each iteration is summarized below:

**Step 1:** Set initial value $\lambda_0 \in \Lambda$.

**Step 2: (E step)** Evaluate the expected complete data log-likelihood function.

$$L(\lambda, \lambda_n) = \mathbb{E}\left[\log \mathbb{P}\left(y_{0:N}, \theta_{0:N} \mid \lambda\right) \mid y_{0:N}, \lambda_n\right].\qquad(3.31)$$

**Step 3: (M step)** Derive $\lambda_{n+1}$ from $\lambda_n$ which takes the following form:

$$\lambda_{n+1} = \arg\max_{\lambda \in \Lambda} L(\lambda, \lambda_n).$$

**Step 4** Repeat E and M steps until a stopping criterion is satisfied.

Given initialisations of the parameters and the observed data, in the expectation part of the algorithm, the EM algorithm starts by calculating the smoothers of a particular state. With the smoothers, the algorithm constructs the expected complete data log-likelihood function. The expected complete data log-likelihood function is then maximized in the maximization part of the algorithm in order to obtain new estimates of the parameters. This is done by equating the first-order derivatives, with respect to the parameters, to zero. Using these new estimates, the algorithm returns to the expectation part again until new estimates satisfy some stopping rule. In most cases, however, the EM algorithm is not guaranteed to converge to the global optimum. Instead, it stops at a local optimum. The extent to which the algorithm correctly estimates parameters is reflected by the initial set of parameters. Initialization of the algorithm is therefore an important consideration. We run multiple procedures of the EM algorithm with different initial values and compare the likelihood of different convergences.

We derive the EM algorithm for finding the maximum likelihood estimate of the parameters of an HMM given a set of observations. This algorithm is known as the Baum–Welch algorithm proposed by Baum and Petrie (1966) and Baum, Petrie, Soules and Weiss (1970). The formal definition of an HMM is as follows:

$$\Lambda = (A, B, p).$$

$A$ is the time-homogeneous transition matrix storing the probability of state $j$ following state $i$.

$$A = \{a_{ij}\}, \quad a_{ij} = \mathbb{P}(\theta_k = j \mid \theta_{k-1} = i), \quad \forall i, j \in E.$$

$B$ consists of the time-homogeneous density functions for observations assigned to each state. They are assumed to follow the Gaussian distribution,

$$B = \{b_i(y)\}, \quad b_i = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{(y - \mu_i)^2}{2\Sigma}\right) \quad \forall i \in E,$$

where distributions have the different means, $\mu_1$ and $\mu_2$, respectively but the same variance, $\Sigma$. $p$ is the initial state probability,

$$p = \{p_i\}, \quad p_i = \mathbb{P}(\theta_0 = i), \quad \forall i \in E.$$

A direct approach for estimating $\lambda = \{A, B, p\}$ is maximizing the log-likelihood function $l$.

$$l(y_{0:N} \mid \lambda) = \log \prod_{k=0}^{N} \mathbb{P}(y_k \mid \lambda) = \sum_{k=0}^{N} \log \sum_{\theta_k} \mathbb{P}(y_k, \theta_k \mid \lambda). \tag{3.32}$$

The problem with equation (3.32) comes from the inner summation over all possible values of the unobserved data $\theta_t$, which might be difficult to compute due to the dimension and length of the data. Summation might grow exponentially and computing the gradient and likelihood in a gradient ascent maximization is infeasible. In the EM algorithm, equation (3.31) can be written as:

$$L(\lambda, \lambda_n) = \sum_{\theta_{0:N}} \log \mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda)\, \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n)$$

$$= \sum_{\theta_{0:N}} \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n) \left( \log p_{\theta_0} + \sum_{k=0}^{N-1} \log a_{\theta_k \theta_{k+1}} + \sum_{k=0}^{N} \log b_{\theta_k}(y_k) \right),$$
$$\tag{3.33}$$

where the last equation comes from the fact that the joint density function of the hidden states $\theta_{0:N}$ and associated observations $y_{0:N}$ is given by

$$\mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda) = p_{\theta_0} b_{\theta_0}(y_0) a_{\theta_0 \theta_1} \cdots a_{\theta_{N-1}\theta_N} b_{\theta_N}(y_N)$$

$$= p_{\theta_0} \prod_{k=0}^{N-1} a_{\theta_k \theta_{k+1}} \prod_{k=0}^{N} b_{\theta_k}(y_k).$$

To maximize equation (3.33), we can maximize the terms separately since it has additive form and they are not related. The first term in equation (3.33) becomes

$$\sum_{\theta_{0:N}} \mathbb{P}(\theta_{0:N} \mid y_{0:N}, \lambda_n) \log p_{\theta_0} = \sum_{i \in E} \mathbb{P}(\theta_0 = i \mid y_{0:N}, \lambda_n) \log p_i.$$

Introducing the Lagrange multiplier $\delta$ corresponding to the equality constraints $\sum_{i \in E} p_i = 1$, and setting the derivative equal to zero, it yields:

$$\frac{\partial}{\partial p_i} \left( \sum_{i \in E} \mathbb{P}(\theta_0 = i | y_{0:N}, \lambda_n) \log p_i + \delta \left( \sum_{i \in E} p_i - 1 \right) \right) = 0,$$

which is solved by

$$\hat{p}_i = \frac{\mathbb{P}(\theta_0 = i | y_{0:N}, \lambda_n)}{\sum_{i \in E} \mathbb{P}(\theta_0 = i | y_{0:N}, \lambda_n)}, \quad \forall i \in E. \tag{3.34}$$

The second term in equation (3.33) can be written as:

$$\sum_{\theta_{0:N}} \mathbb{P}(\theta_{0:N} | y_{0:N}, \lambda_n) \sum_{k=0}^{N-1} \log a_{\theta_k \theta_{k+1}}$$

$$= \sum_{i,j \in E} \sum_{k=0}^{N-1} \mathbb{P}(\theta_k = i, \theta_{k+1} = j | y_{0:N}, \lambda_n) \log a_{ij}.$$

In a similar way to the first term, the Lagrange multiplier combined with the constraint, $\sum_{j \in E} a_{ij} = 1$, yields:

$$\hat{a}_{ij} = \frac{\sum_{k=0}^{N-1} \mathbb{P}(\theta_k = i, \theta_{k+1} = j | y_{0:N}, \lambda_n)}{\sum_{k=0}^{N-1} \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n)}, \quad \forall i, j \in E. \tag{3.35}$$

The third term in equation (3.33) becomes:

$$\sum_{\theta_{0:N}} \mathbb{P}(\theta_{0:N} | y_{0:N}, \lambda_n) \sum_{k=0}^{N} \log b_{\theta_k}(y_k) = \sum_{i \in E} \sum_{k=0}^{N} \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n) \log b_i(y_k). \tag{3.36}$$

Since $b_i(y_k)$, for any $i \in E$, and $k = 0, \ldots, N$ is assumed to follow a normal distribution with mean $\mu_i$ and variance $\Sigma$, the right-hand-side of equation (3.36) can be written as:

$$\sum_{i \in E} \sum_{k=0}^{N} \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n) \left( -\frac{(y_k - \mu_i)^2}{2\Sigma} - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \Sigma \right). \tag{3.37}$$

The maximizers are given by differentiating equation (3.37) with respect to $\mu_i$ and $\Sigma$ and setting them both equal to zero. This yields

$$\hat{\mu}_i = \frac{\sum_{k=0}^{N} y_k \, \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n)}{\sum_{k=0}^{N} \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n)}, \quad \forall i \in E. \tag{3.38}$$

$$\hat{\Sigma} = \frac{\sum_{i \in E} \sum_{k=0}^{N} (y_k - \hat{\mu}_i)^2 \, \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n)}{\sum_{i \in E} \sum_{k=0}^{N} \mathbb{P}(\theta_k = i | y_{0:N}, \lambda_n)}. \tag{3.39}$$

Acquiring updated parameters $\lambda_{n+1} = (\hat{A}, \hat{B}, \hat{p})$ from $\lambda_n$ requires two quantities,

$$\mathbb{P}(\theta_k = i \mid y_{0:N}, \lambda_n) \tag{3.40}$$

$$\mathbb{P}(\theta_k = i, \theta_{k+1} = j \mid y_{0:N}, \lambda_n), \tag{3.41}$$

which appear in equations (3.34), (3.35), (3.38) and (3.39). They are referred to as smoothers, in that estimates are given based on whole observations. These quantities are obtained with the Baum–Welch algorithm including a forward–backward procedure.

In the EM algorithm, the stopping criterion used to stop the algorithm is the absolute difference of maximized log-likelihood functions from one step before being less than $10^{-4}$. That is, the algorithm is stopped if

$$\left| l(y_0, \dots y_N \mid \lambda_{n+1}) - l(y_0, \dots, y_N \mid \lambda_n) \right| < 10^{-4}.$$

**Baum–Welch Algorithm**

In the Baum–Welch algorithm, two quantities, equations (3.40) and (3.41) are efficiently computed with forward and backward procedures. For each iteration, start computing the forward procedure and then the backward one, with parameters $\lambda_n$. For the forward procedure, we define

$$\alpha_i(k) = \mathbb{P}(y_{0:k}, \theta_k = i \mid \lambda_n), \quad \forall i \in E, \quad k = 0, \dots, N,$$

which is the probability of the partial observation sequence $y_0, \dots y_k$ where $\theta$ is in state $i$ at time $k$. The key result is that $\alpha_i(k)$ can be computed recursively as:

1. $\alpha_i(0) = p_i b_i(y_0), \; \forall i \in E$.

2. $\alpha_i(k) = \left( \sum_{j \in E} \alpha_j(k-1) a_{ji} \right) b_i(y_k), \quad \forall i \in E, \quad k = 1, \dots, N$.

3. $\mathbb{P}(y_{0:N} \mid \lambda) = \sum_{i \in E} \alpha_i(N)$.

The backward procedure is analogous to the forward procedure except it starts at time $N$ and works back towards time 0.

$$\beta_i(k) = \mathbb{P}(y_{k+1:N} \mid \theta_k = i, \lambda_n), \quad \forall i \in E, \quad k = 0, \dots, N-1,$$

which is the probability of the partial sequence $y_{k+1}, \ldots, y_N$ given the starting state $i$ at time $k$.

1. $\beta_i(N) = 1, \quad \forall i \in E.$

2. $\beta_i(k) = \sum_{j \in E} a_{ij} b_j(y_{k+1}) \beta_j(k+1), \quad \forall i \in E, \quad k = N-1, \ldots, 0.$

3. $\mathbb{P}(y_{0:N} \mid \lambda) = \sum_{i \in E} p_i b_i(y_0) \beta_i(0).$

For $k = 0, \ldots, N$ and any $i \in E$, define

$$\gamma_i(k) := \mathbb{P}(\theta_k = i \mid y_{0:N}, \lambda_n),$$

which is the probability of being in state $i$ at time $k$ given the whole observation, $y_0, \ldots, y_N$. Note that

$$\mathbb{P}(\theta_k = i \mid y_{0:N}, \lambda_n) = \frac{\mathbb{P}(y_{0:N}, \theta_k = i \mid \lambda_n)}{\mathbb{P}(y_{0:N} \mid \lambda_n)} = \frac{\mathbb{P}(y_{0:N}, \theta_k = i \mid \lambda_n)}{\sum_{j \in E} \mathbb{P}(y_{0:N}, \theta_k = j \mid \lambda_n)}.$$

Also note that because of Markovian conditional independence, for any $i$ and $k$,

$$\alpha_i(k) \beta_i(k) = \mathbb{P}(y_{0:k}, \theta_k = i \mid \lambda_n) \mathbb{P}(y_{k+1:N} \mid \theta_k = i, \lambda_n) = \mathbb{P}(y_{0:N}, \theta_k = i \mid \lambda_n).$$

Consequently we can calculate $\gamma_i(k)$ in terms of $\alpha_i(k)$ and $\beta_i(k)$ as

$$\gamma_i(k) = \frac{\alpha_i(k) \beta_i(k)}{\sum_{j \in E} \alpha_j(k) \beta_j(k)}, \quad \forall i \in E, \quad k = 0, \ldots, N.$$

Next, we define

$$\xi_{ij}(k) := \mathbb{P}(\theta_k = i, \theta_{k+1} = j \mid y_{0:N}, \lambda_n), \quad \forall i, j \in E, \quad k = 0, \ldots, N-1, \quad (3.42)$$

which is the probability of being in state $i$ at time $k$ and being in state $j$ at $k + 1$. This can be computed as:

$$\xi_{ij}(k) = \frac{\mathbb{P}(\theta_k = i, \theta_{k+1} = j, y_{0:N} \mid \lambda)}{\mathbb{P}(y_{0:N} \mid \theta_k = i, \lambda)} = \frac{\alpha_i(k) a_{ij} b_j(y_{k+1}) \beta_j(k+1)}{\sum_{i,j \in E} \alpha_i(k) a_{ij} b_j(y_{t+1}) \beta_j(k+1)}.$$

Both $\gamma$ and $\xi$ appear in computing the updating parameters in equations (3.34), (3.35), (3.38) and (3.39).

### 3.3.2 Data and Estimation Results

The spot, futures and options contracts are traded both over-the-counter (OTC) and on exchanges. Daily data on closing prices of futures contracts in the EU ETS are obtained from the European Climate Exchange (ECX), the most liquid futures market. 80% of EU ETS transactions were struck OTC and 38% of OTC transactions are cleared through the ECX, see the report from Capoor and Ambrosi (2008). In the ECX, futures contracts with various maturities are traded. For example, during 2008, the first year of Phase II, the spot contracts which expire in December 2008 are traded and futures contracts expiring December 2009–2012 respectively have been traded. The time period considered is from June 2, 2006 to December 31, 2007, totally 430 trading days, for estimating the parameters of the futures contract maturing in December 2008.

Parameter estimation based on the EM algorithm described in the previous subsection is performed. We set 240 trading days per year, $\Delta = 1/240$. Estimates are presented in Table 3.1. The maximum likelihood estimates appear in the first row and the corresponding asymptotic standard deviations calculated from the outer product and Hessian matrix appear in the second row.

Table 3.1: Estimation results of equation (3.29)

| $\mu_1$ | $\mu_2$ | $\Sigma$ |
|---|---|---|
| 0.0093 | −0.0022 | 0.00057 |
| $(2.6 \times 10^{-4})$ | $(5.1 \times 10^{-4})$ | $(0.2 \times 10^{-4})$ |

$$R = \begin{pmatrix} 0.549 & 0.451 \\ 0.090 & 0.910 \end{pmatrix}, \quad p = \begin{pmatrix} 0.027 & 0.973 \end{pmatrix}.$$

The log-likelihood = 987.04.

The estimation results of the transition matrix are indicative that during the data period, the market has been almost in one state. This is consistent with some reports that the EU ETS was, during Phase I, significantly over-allocated, meaning companies received more allocations compared to the actual emissions and the market had been long. See, for example, Ellerman and Buchner (2008). Since the market was long in this period, we set the relationship between $\mu_1$ and

$\mu_2$ and $\mu + \alpha\theta$ as

$$\left(\mu + \alpha - \frac{1}{2}\sigma^2\right)\Delta t = \mu_2$$

$$\left(\mu - \alpha - \frac{1}{2}\sigma^2\right)\Delta t = \mu_1.$$

Combined with the relation that

$$\sigma^2\Delta t = \Sigma^2,$$

$\mu$, $\alpha$ and $\sigma$ are calculated. The relationship between $R$ and $\lambda$ is given by equations (3.26) and (3.27). Corresponding parameters in continuous-time are given in Table 3.2.

Table 3.2: Parameters in continuous-time

| $\mu$ | $\alpha$ | $\sigma$ | $\lambda_1$ | $\lambda_{-1}$ |
|-------|----------|----------|-------------|----------------|
| 0.92 | −1.38 | 0.37 | 155.80 | 31.08 |

## 3.4 Numerical Studies for EUA Prices

The purpose of this section is to compare price characteristics between complete information and incomplete information. Emphasis will be placed on experiments which highlight differences in key quantities such as initial prices, drift coefficients and the infinitesimal generator. The EUA0 price under the complete information setting is given in equation (3.6) and under the incomplete information setting in equation (3.21). We use the following set of parameters shown in Table 3.3. Monte Carlo prices are obtained with 100,000 simulations.[1]

Table 3.3: Parameters in numerical studies

| $s_0$ | $\alpha$ | $\mu$ | $\sigma$ | $\lambda_1$ | $\lambda_{-1}$ | $K$ | $\theta_0$ | $\pi_1(0)$ | $T$ |
|-------|----------|-------|----------|-------------|----------------|-----|------------|------------|-----|
| 30 | −2.0 | 0.5 | 0.5 | 160 | 30 | 100 | 1 | 1.0 | 1.0 |

---

[1]Due to the requirement of time discretization, a sample path of $\bar{\theta}_t$ could diverge. To avoid divergence, we discretize 1 year into 100,000 time steps.

Figures 3.1, 3.2 and 3.3 show the prices under complete information and incomplete information with respect to $s_0$, $\lambda_1$ and $\lambda_{-1}$, respectively. The overall impression is that their sensitivities have similar patterns in the sense that prices have the same sign of first and second order derivatives. The higher value in $\lambda_1$ or lower value in $\lambda_{-1}$ leads to the higher probability of $\theta = 1$, indicating lower EUA0 prices.

Prices under incomplete information are, in the sense that they depend on the coefficients of $S$, $\mu$, $\alpha$ and $\sigma$, different to those under complete information. Figures 3.4, 3.5 and 3.6 show the sensitivities with respect to $\mu$, $\alpha$ and $\sigma$, respectively. Since prices under complete information do not depend on these parameters, they are flat, while prices under incomplete information do depend on these parameters. Since equation (3.17) indicates that a higher value of $\mu$ leads to a lower value of $\bar{\theta}$ with our parameters, the higher value of $\mu$ leads to the lower EUA0 price as shown in Figure 3.4. Both $\alpha$ and $\sigma$ are included in equation (3.17) as a quadratic form, leading to the shape of quadratic function as in Figures 3.5 and 3.6.

Another interesting phenomenon is the jump at $T$ in the conditional expectation, which occurs when the verified emissions are announced. Consequently, it causes the discontinuous changes in the EUA0 price since the cash flow at $T$ is determined based on $\theta_T$ and the initial value to compute the expectation at $2T$ will vary. We confirm this on a sample path by sample path basis. At $T$, the true position of the zone is revealed and the conditional expectation $\bar{\theta}_T$ will jump to the revealed value $\theta_T$. Simulation is done as follows. The EUA1 price and process of the conditional expectation under $\mathbb{P}$ are simulated with equations (3.10) and (3.14). At each $t_i$, with $(S_{t_i}, \bar{\theta}_{t_i})$, the EUA0 price is computed using Monte Carlo simulation. Figure 3.8 presents the sample paths for the EUA0 and EUA1 prices. Assume at $T$, the verified emissions amount is revealed and the market is long. Consequently, the conditional expectation jumps as shown in Figure 3.7 and the price for EUA0 is discontinuous as shown in Figure 3.8.

97

Figure 3.1: The EUA0 price with various initial values, $s_0$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
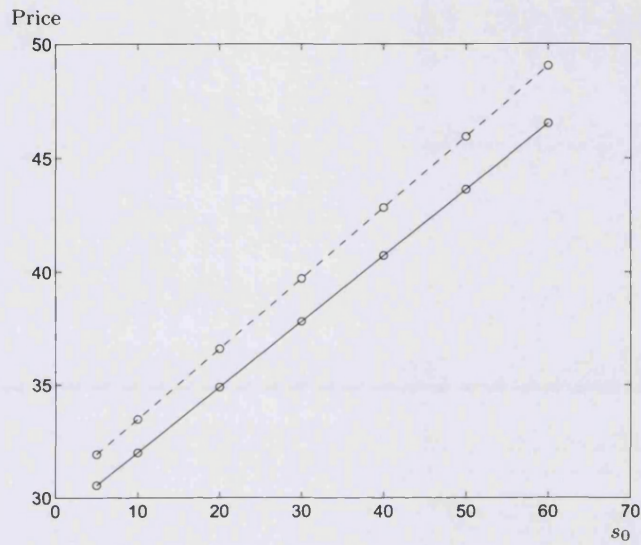


Figure 3.2: The EUA0 price with various initial values, $\lambda_1$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
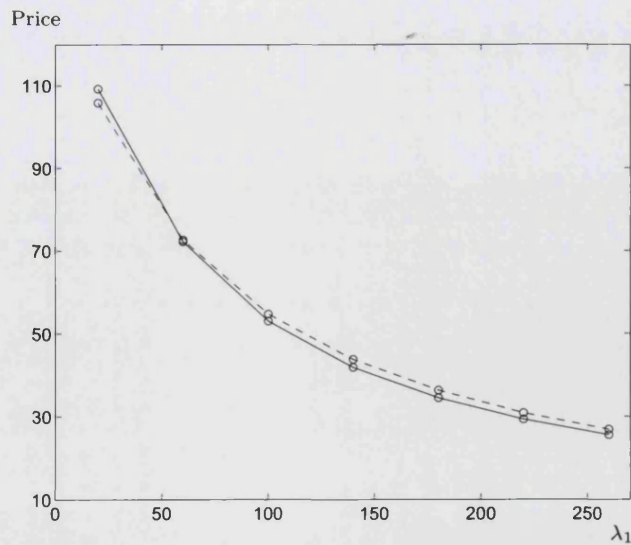
Figure 3.3: The EUA0 price with various initial values, $\lambda_{-1}$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
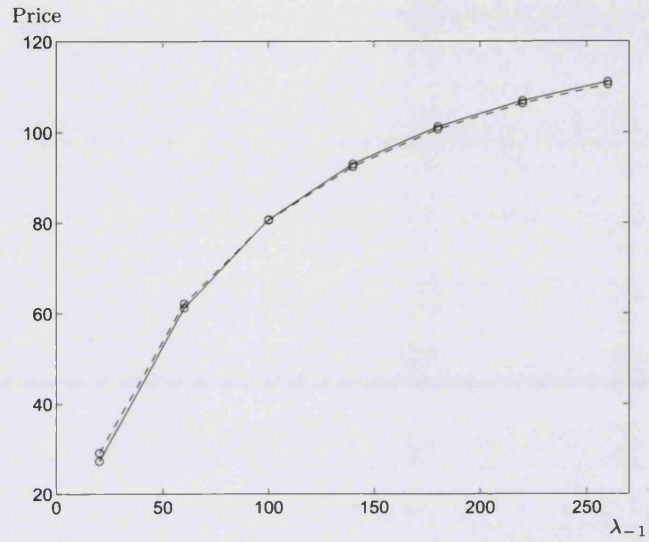


Figure 3.4: The EUA0 price with various initial values, $\mu$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.

Figure 3.5: The EUA0 price with various initial values, $\alpha$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.



Figure 3.6: The EUA0 price with various initial values, $\sigma$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
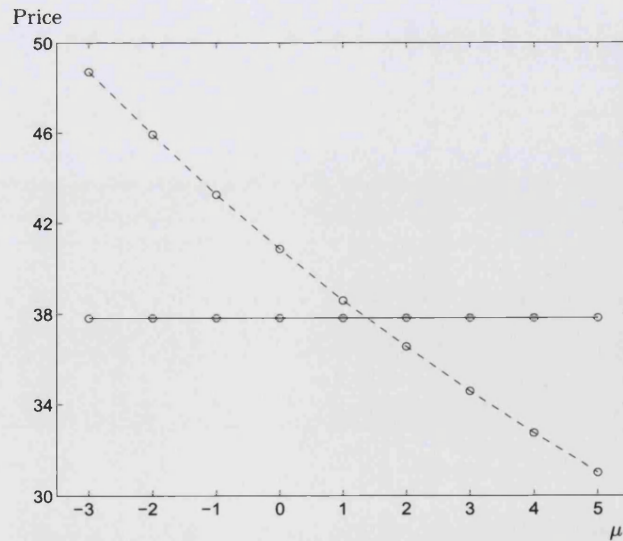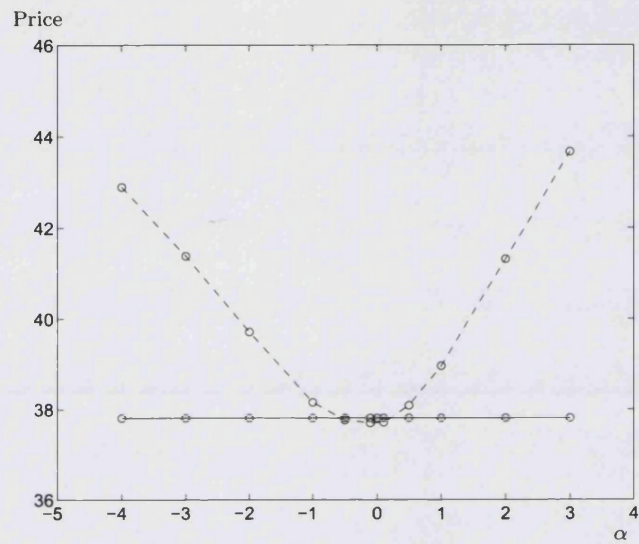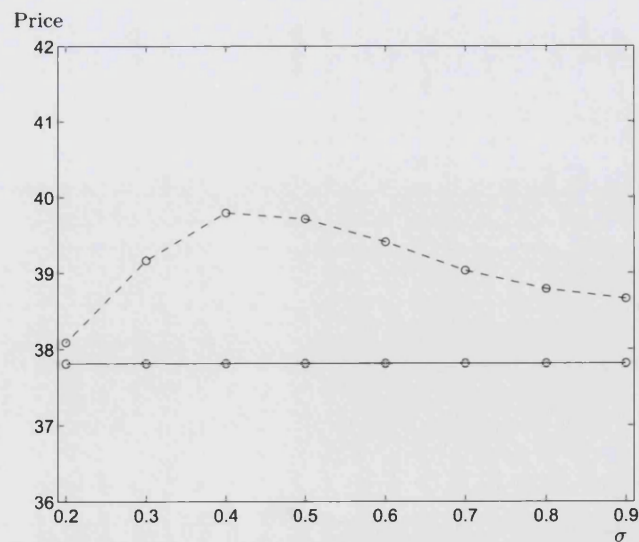
Figure 3.7: Simulated process of the conditional expectation.
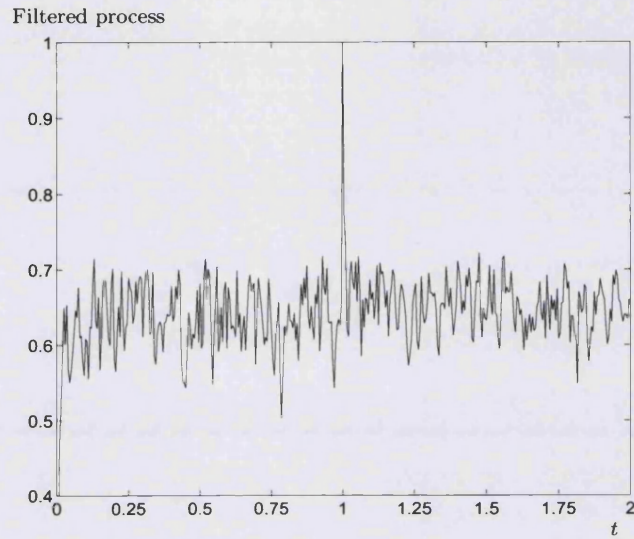


Figure 3.8: Simulated EUA0 and EUA1 prices. Solid line represents the EUA1 price and dashed line the EUA0 price.
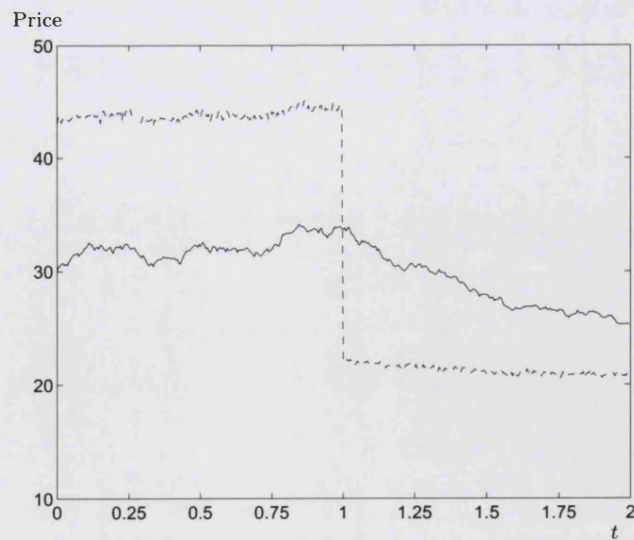
## 3.5 Concluding Remarks

In this chapter, we consider the spot and forward relationship in the carbon emissions market based upon the no-arbitrage principle. This market differs due to the trading regulations and the market design from other financial markets: the net position of the zone is taken into consideration for explaining the relationship. We assume the net position of the zone follows a continuous-time Markov chain and the drift term of the forward price process depends on the net position of the zone. Bankability is included in the framework in the form of an expectation of future cash-flows.

Two information settings are considered: complete and incomplete. The complete information setting is defined when market participants can access full information. Since the net position of the zone is not tradable, this setting is pricing in an incomplete market. We adopt the locally risk-minimization technique so as to fix the equivalent martingale measure. With the minimal martingale measure, the analytical arbitrage-free price is obtained. Under the incomplete information setting when market participants can observe the forward process and the net position of the zone only at announcement time, a filtering technique is applied to derive the optimal projection of the net position of the zone onto the new filtration. This leads to the use of the Wonham filter. The minimal martingale measure is found and the distribution of the net position of te zone restricted to the new filtration under this measure is calculated. Consequently, the arbitrage-free price is obtained using Monte Carlo simulation.

There could be several extensions without changing the framework. The assumption that there are only two trading periods can easily be extended to multi-periods. In the multi-period setting when banking is prohibited only at the end of the phase and intra-period banking is available, the payoff function has a nested structure. We denote by $T_N$ the end of the phase when the banking is prohibited and by $T_1, \ldots, T_{N-1}$ the end of each trading year when intra-banking

is available. The payoff function can be written as follows:

$$P_{T_N} = \begin{cases} S_{T_N} + K & \text{if } \theta_{T_N} = -1, \\ 0 & \text{if } \theta_{T_N} = 1. \end{cases}$$

$$P_{T_i} = \begin{cases} S_{T_i} + K & \text{if } \theta_{T_i} = -1, \\ \mathbb{E}\big[P_{T_{i+1}} \big| S_{T_i}, \theta_{T_i} = 1\big] & \text{if } \theta_{T_i} = 1, \quad i = 1, \ldots, N-1. \end{cases}$$

The extension of the assumption that the net position of the zone follows a two-state Markov chain is straightforward. With the assumption of an $N$-state Markov chain that takes values in $E = \{i_1, \ldots, i_N\}$, it expresses the extent of how long or short the market is. The semi-martingale representation of an $N$-state Markov chain is written as

$$\theta_t = \theta_0 + A_t + N_t,$$
$$A_t = \int_0^t (Q_s)(\theta_s)\, ds = \int_0^t \sum_{i \in E} (i - \theta_s)\lambda_i \, ds,$$

where $\lambda_i$ represents the jump intensity to state $i$ from other states and $N_t$ is a martingale. Computing the spot price under the complete information setting with these extensions would be complicated but straightforward. Under the incomplete information setting, the conditional expectation process of $\theta_t$ is defined as

$$\bar{\theta}_t = \sum_{i \in E} i\pi_i(t),$$
$$\pi_i(t) := \mathbb{P}(\theta_t = i \mid \mathcal{G}_t).$$

Furthermore, $\mathbb{P}(\theta_t = i \mid \mathcal{G}_t)$, $i \in E$ is obtained from filtering theory. Extrapolating probabilities between each year, $\pi_i(T_{i+1}, T_i)$, have to be derived and this is analogous to equation (3.22) because of the Markovian structure.

A discount factor, or interest rate, can be included in the framework. A deterministic or stochastic interest rate which is independent of EUA prices is easily implemented in the form of a zero-coupon bond. To implement a stochastic interest rate which is correlated with EUA prices, a change of measure technique can be used to manipulate the equations.

Since the market for the EU ETS has become considerably more active, the link between EUA contracts and other carbon contracts has deepened. For example, the EU ETS scheme not only allows companies to trade among themselves

but enables interactions with other countries not included in the EU ETS via the Clean Development Mechanism and Joint Implementation schemes. During Phase II, firms can cover their requirements by purchasing Certified Emission Reduction Credits from such projects. The amount of these contracts are growing, and therefore their pricing is important.

# Chapter 4

# The Spot and Forward Relationship in Carbon Emissions Markets II

## 4.1 Introduction

In this chapter, we deal with the same problem as in Chapter 3 in a different setting. We assume that the net position of the zone follows a linear diffusion while it is assumed to follow a Markov chain in Chapter 3. The threshold that determines the market is long or short is introduced. The forward process is assumed to be driven by a Brownian motion and its drift term is modelled to be a function of the net position of the zone which is assumed to follow a linear diffusion, driven by another Brownian motion. The setting where the unobserved process follows a linear diffusion leads to the use of the Kalman-Bucy filter. Parameter estimation is done with the Kalman filter.

The outline of this chapter is as follows. Section 4.2 introduces the models used for pricing. We present models for the forward price process and the net position of the zone and then derive the spot and forward relationship under both complete and incomplete information settings. Estimated parameters are given in Section 4.3, along with the EM algorithm for the Kalman filter. Numerical examples are provided in Section 4.4, where the price differences are highlighted in the two information settings with respect to each parameter. The jumps in the filtered process coming from the announcement of the actual value and consequent changes in the spot price are of interest. Concluding remarks and several possible extensions are discussed in Section 4.5.

## 4.2 Models for EUA Prices

The framework for exploring the spot and forward relationship in the EU ETS market is introduced in this section. For simplicity, assume two EUAs are traded in the market: an EUA for the current year denoted by EUA0, and an EUA for the following year denoted by EUA1. We will first introduce the underlying processes and subsequently define the equation which the EUA0 price satisfies.

### 4.2.1 The spot and forward relationship in EU ETS markets

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a stochastic space with a filtration $\mathcal{F}_t$ and all stochastic processes defined on this space are assumed to be $\mathbb{F}$-adapted, $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}$, where $T^*$ is a fixed but arbitrary time horizon. We adopt a similar approach to Çetin and Verschuere (2009). Let the positive-valued EUA1 price process, $S$, be a continuous-time process satisfying the stochastic differential equation (SDE),

$$dS_t = (\mu + \alpha \theta_t) S_t \, dt + \sigma_S S_t \, dW_t, \tag{4.1}$$

where $\mu$, $\alpha$ and $\sigma_S$ are assumed constant, $W_t$ is a one-dimensional standard Brownian motion. $\theta_t$ is assumed to be a linear diffusion expressing the net position of the zone whose details are explained later. The pair, $(S_t, \theta_t)$, is a vector-valued Markov process. From equation (4.1), it follows that the drift term of the EUA1 price has a different expected mean depending on the value of the net position of the zone. The choice of the drift coefficient is meant to reflect the changing demand for allowances depending on the net position of the market.

EU ETS provides companies with fixed, free allowances on an annual basis at the beginning of each trading year and announces the total allowances submitted by the companies at the end of each trading year. Companies which receive allowances can emit $CO_2$ up to their allocated allowance and in the market companies can trade allowances to offset any excess or shortage, as explained in Section 4.1. While total allowances are capped by regulators, the total amount of emissions is uncertain, indicating that every year total allowances are uncertain. The net position of the zone models the difference between emissions and allocations. If the total allowances were over-allocated, or companies abated

$CO_2$ emissions, the net position of the zone is long, meaning that the total allowances would be in excess. On the other hand, if numerous companies emit more than their allocations, or they abate less, the net position of the zone is short, meaning that the total allowances would result in a shortage.

The net position of the zone, $\theta_t$, is assumed to follow a continuous-valued linear diffusion taking values in $E := \mathbb{R}$.

$$d\theta_t = (a + b\theta_t)\, dt + \sigma_\theta\, dB_t, \qquad (4.2)$$

where $a$, $b$ and $\sigma_\theta$ are assumed constant. $B_t$ is assumed to be a one-dimensional standard Brownian motion, which is independent of $W_t$.

The spot and forward relationship under the assumption of no banking, which was in effect in 2007 during the transition from Phase I to Phase II, was analysed in Çetin and Verschuere (2009). In their approach, they modelled the net position of the zone as a Markov chain taking two values, one short and the other long. Under the assumption of no banking, the EUA0 price would go to zero as the expiry date approaches, provided that the net position of the zone is long. This is because there are companies which have allowances that more than cover actual emissions and the excess allowances are worthless after expiry, which therefore drives the price to zero. If the net position of the zone is short, EUA1 turns to EUA0 next year by paying a penalty, denoted by $K$, for every excess tonne of emissions, because of regulations.

Considering the above points, along with the assumption that $\theta_t$ follows a Markov chain with two states as in Çetin and Verschuere (2009), the value of EUA0 at the end of Phase I, $T$, denoted by $P_T$ is assumed to satisfy the following relationship:

**Definition 4.1** *(Çetin and Verschuere (2009)). The EUA0 price at the end of Phase I, $T$, is defined as*

$$P_T = \begin{cases} S_T + K & \text{if} \quad \theta_T = -1, \\ 0 & \text{if} \quad \theta_T = 1. \end{cases}$$

We modify Definition 4.1 for continuous-valued $\theta$ simply by changing the conditions with respect to $\theta_T$ and introducing the threshold $\kappa$, assumed constant. We

set $\theta_t < \kappa$ and $\theta_t \geq \kappa$ corresponding to the net position of the zone at $t$ being short and long, respectively. The higher and lower the values of $\theta_t$, the greater the degree of being long and short respectively.

In this chapter, we discuss the relationship when banking is in effect. Phase I consists of three years, from 2005-2007, and Phase II consists of five years, from 2008-2012, but for simplicity in what follows, we assume that there are only two periods to trade, one being $[0, T]$ and the other $[T, 2T]$, where $2T < T^*$, and there is no constraint on intra-period banking between $[0, T]$ and $[T, 2T]$, and $2T$ is the maturity of the phase when banking for next year is banned. At $2T$, since banking is not allowed, the unused permits are worthless provided the market is long. We assume that at $2T$ the following relationship will hold,

$$P_{2T} = \begin{cases} S_{2T} + K & \text{if} \quad \theta_{2T} < \kappa, \\ 0 & \text{if} \quad \theta_{2T} \geq \kappa. \end{cases}$$

We consider the relationship at $T$. Now that intra-period banking is in effect, unused permits are no longer worthless. Banking is in effect so that unused permits can be transferred and used later, provided that the market is long. We assume in this case that EUA0 is evaluated as the expectation of the future value. The price of an EUA0 contract, in the case that the market is short, is equal to that of EUA1 plus a penalty as the market is short at $2T$. The relation at $T$ is therefore

$$P_T = \begin{cases} S_T + K & \text{if} \quad \theta_T < \kappa, \\ \mathbb{E}\left[1_{\{\theta_{2T} < \kappa\}}(S_{2T} + K) \,\middle|\, S_T, \theta_T \geq \kappa\right] & \text{if} \quad \theta_T \geq \kappa, \end{cases}$$

where $1_{\{\theta_t = i\}}$ is an indicator function which returns 1 if $\theta_t = i$, 0 otherwise and $\mathbb{E}[\cdot]$ is the expectation operator under a pricing measure explained later. Considering the relationship at $T$, it turns out that EUA0 is regarded as an option on EUA1 as well as on the net position of the zone. In the next two subsections, we discuss the pricing method based on a local risk-minimizing approach.

### 4.2.2 Pricing under Complete Information

We consider no-arbitrage pricing in the sense that there exist equivalent martingale measures and the fair price is calculated as an expectation of future

cash-flows under these measures. In this subsection, pricing is done under the complete information setting: market participants can access full information, $\mathbb{F}$, and the expectation is taken with respect to $\mathbb{F}$.

The pricing problem in our setting is considered as that in the incomplete market. Incompleteness arises in the presence of the net position of the zone, since it is not tradable. The uncertainty in the EUA0 price caused by $W$ and $\theta$ cannot be hedged perfectly by only trading $S$. We employ a local risk-minimizing strategy for fixing one equivalent martingale measure as in Chapter 3. The minimal martingale measure, $\hat{\mathbb{P}}$, in our setting, is identified by the following Radon-Nikodým derivative,

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}} = \hat{Z}_{2T}, \tag{4.3}$$

where

$$\hat{Z}_t = \exp\left(-\int_0^t \frac{\mu + \alpha\theta_s}{\sigma_S}\, dW_s - \frac{1}{2}\int_0^t \left(\frac{\mu + \alpha\theta_s}{\sigma_S}\right)^2 ds\right), \quad 0 \le t \le 2T.$$

The Girsanov-Meyer theorem (see, for example, Protter (2004, Theorem 39)) yields $\hat{W}$ defined as

$$\hat{W}_t = W_t + \int_0^t \frac{\mu + \alpha\theta_s}{\sigma_S}\, ds,$$

where $\hat{W}_t$ is a standard Brownian motion under $\hat{\mathbb{P}}$, for $0 \le t \le 2T$.

The next step is to derive the underlying processes under the minimal martingale measure. The EUA1 price process under $\hat{\mathbb{P}}$ follows, along with $\hat{W}$, the SDE,

$$dS_t = (\mu + \alpha\theta_t)S_t\, dt + \sigma_S S_t\left(d\hat{W}_t - \frac{\mu + \alpha\theta_t}{\sigma_S}\, dt\right)$$
$$= \sigma_S S_t\, d\hat{W}_t, \tag{4.4}$$

and $\theta_t$ under $\hat{\mathbb{P}}$ has the same representation as under $\mathbb{P}$ because of orthogonality, and so

$$d\theta_t = (a + b\theta_t)\, dt + \sigma_\theta\, dB_t. \tag{4.5}$$

As a result, $S_t$ and $\theta_t$ are independent under $\hat{\mathbb{P}}$. $\theta_t$ follows a normal distribution with mean and variance, respectively, given by

$$\mathbb{E}[\theta_t \mid \mathcal{F}_s] = \theta_s e^{b(t-s)} - \frac{a}{b}\left(1 - e^{b(t-s)}\right),$$

$$\mathrm{Var}[\theta_t \mid \mathcal{F}_s] = -\frac{\sigma_\theta^2}{2b}\left(1 - e^{2b(t-s)}\right).$$

Since we adopt the minimal martingale measure approach, the unused permit at $2T$ is evaluated under this measure. We can now give a precise definition of the EUA0 price under complete information.

**Definition 4.2** *Under the minimal martingale measure, the EUA0 price in Phase II contracts at $T$ is defined as*

$$P_T = \begin{cases} S_T + K & \text{if } \theta_T < \kappa, \\ \hat{\mathbb{E}}\big[1_{\{\theta_{2T} < \kappa\}}(S_{2T} + K)\big| S_T, \theta_T \geq \kappa\big] & \text{if } \theta_T \geq \kappa, \end{cases} \qquad (4.6)$$

*where $\hat{\mathbb{E}}[\cdot]$ is the expectation operator under $\hat{\mathbb{P}}$.*

With the minimal martingale measure, derivative prices are determined from Föllmer and Schweizer (1991) mentioned in Theorem 3.1. In our problem the payoff, $C$, is defined from equation (4.6) as

$$C = 1_{\{\theta_T < \kappa\}}(S_T + K) + 1_{\{\theta_T \geq \kappa\}}\hat{\mathbb{E}}\big[1_{\{\theta_{2T} < \kappa\}}(S_{2T} + K) \mid S_T, \theta_T \geq \kappa\big].$$

The requirement of $\mathcal{F}_T$- and $\mathcal{F}_{2T}$-measurability and square integrability for the payoff $C$ is met in our setting.

The EUA0 price under the complete information setting can be written as follows.

**Theorem 4.1** *The EUA0 contract price at $t < T$, (4.6), under complete information is given by*

$$P_t = (S_t + K)\left(1 - (1 - G)\hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t)\right), \qquad (4.7)$$

*where $G = \hat{\mathbb{P}}(\theta_{2T} < \kappa \mid \theta_T \geq \kappa)$.*

**Proof** With the minimal martingale measure, the EUA0 prices can be written as

$$P_t = \hat{\mathbb{E}}\big[(S_T + K)1_{\{\theta_T < \kappa\}} + \hat{\mathbb{E}}\big[1_{\{\theta_{2T} < \kappa\}}(S_{2T} + K)\big| S_T, \theta_T \geq \kappa\big]1_{\{\theta_T \geq \kappa\}}\big| \mathcal{F}_t\big].$$

Since $\theta_{2T}$ and $S_{2T}$ are independent, $1_{\{\theta_{2T}\}}$ and $S_{2T}$ are independent as well. The EUA0 price is thus given by

$$
\begin{aligned}
P_t &= \hat{\mathbb{E}}\big[(S_T + K)1_{\{\theta_T < \kappa\}} + \hat{\mathbb{E}}[1_{\{\theta_{2T} < \kappa\}} \mid \theta_T \geq \kappa]\hat{\mathbb{E}}[(S_{2T} + K) \mid S_T]1_{\{\theta_T \geq \kappa\}}\big| \mathcal{F}_t\big] \\
&= \hat{\mathbb{E}}\big[(S_T + K)1_{\{\theta_T < \kappa\}} + (S_T + K)G1_{\{\theta_T \geq \kappa\}}\big| \mathcal{F}_t\big],
\end{aligned}
$$

where $G = \hat{\mathbb{P}}(\theta_{2T} < \kappa \mid \theta_T \geq \kappa)$. Finally,

$$
\begin{aligned}
P_t &= \hat{\mathbb{E}}[(S_T + K)(1 - 1_{\{\theta_T \geq \kappa\}}) + (S_T + K)G1_{\{\theta_T \geq \kappa\}} \mid \mathcal{F}_t] \\
&= (S_t + K)(1 - \hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t) + (S_t + K)G\hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t) \\
&= (S_t + K)\left(1 - (1 - G)\hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t)\right).
\end{aligned}
$$

□

Since the EUA0 price is expressed using the relation,

$$
\frac{\hat{\mathbb{P}}(\theta_T \geq \kappa, \theta_{2T} < \kappa \mid \mathcal{F}_t)}{\hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t)} + \frac{\hat{\mathbb{P}}(\theta_T \geq \kappa, \theta_{2T} \geq \kappa \mid \mathcal{F}_t)}{\hat{\mathbb{P}}(\theta_T \geq \kappa \mid \mathcal{F}_t)} = 1,
$$

as

$$
P_t = (S_t + K)\left(1 - \hat{\mathbb{P}}(\theta_T \geq \kappa, \theta_{2T} \geq \kappa \mid \mathcal{F}_t)\right),
$$

the extension to a multi-period setting is straightforward. The details of several possible extensions are discussed in Section 4.5.

### 4.2.3 Pricing under Incomplete Information

In this subsection, we consider the pricing problem in a more realistic setting. The process $\theta_t$ is, in fact, not always observable to market participants. The information available to market participants is that generated by $S$ and only at $T$ and $2T$, which are announcement times, by $\theta$. By setting up a new filtration in such a way that information from $\theta_t$ is included only at $T$ and $2T$ and pricing is done with the new filtration, this aspect is included in the pricing. We define the new filtration, $\mathcal{G}_t$, expressing this situation where $\mathcal{G}_t$ is defined as

$$
\mathcal{G}_t = \begin{cases} \mathcal{F}_t^S & \text{for } t < T, \\ \mathcal{F}_t^S \vee \sigma(\theta_T) & \text{for } T \leq t < 2T, \\ \mathcal{F}_t^S \vee \sigma(\theta_T) \vee \sigma(\theta_{2T}) & \text{for } t = 2T. \end{cases}
$$

The new filtration $\mathcal{G}$ represents the following:

- At $t < T$, only $S_t$ is observed.
- At $t = T$, both $S_t$ and $\theta_t$ are observed.
- At $T < t < 2T$, only $S_t$ is observed.
- At $t = 2T$, both $S_t$ and $\theta_t$ are observed.

Due to the definition of filtration (see, for example, Section 1.1 in Protter (2004)), it has an increasing family of $\sigma$-algebras, meaning that the filtration $\mathcal{G}_t$ includes all information before $t$. In other words, once observed, the values of $\theta_T$ and $\theta_{2T}$ are held in the filtration.

$\theta$ is no longer an adapted process with respect to $\mathcal{G}$. It is however partially observed through the fluctuations of $S$, from equation (4.1), thereby allowing use of the projection of $\theta$ onto $\mathcal{G}$. We first consider the distribution of $\theta_t$ on $(\mathcal{G}_t, \mathbb{P})$.

**Theorem 4.2** *(Theorem 11.1 in Liptser and Shiryaev (2001)) $\theta_t$ on $(\mathcal{F}_t^S, \mathbb{P})$ follows a normal distribution.*

See, for the proof, Liptser and Shiryaev (2001).

The expectation and variance fully characterize normal distributions so that we define the conditional expectation and variance of $\theta_t$ on $(\mathcal{G}_t, \mathbb{P})$ respectively, for $0 \leq t \leq 2T$, by

$$\bar{\theta}_t := \mathbb{E}[\theta_t \mid \mathcal{G}_t],$$
$$\gamma_t := \mathbb{E}[(\theta_t - \bar{\theta}_t)^2 \mid \mathcal{G}_t]$$
$$= \mathbb{E}[(\theta_t - \bar{\theta}_t)^2].$$

Note that since $\theta_t - \bar{\theta}_t$ is independent of $\mathcal{G}_t$, $\gamma_t$ is a deterministic function and these details are given later. We consider the spot and forward relationship under the new filtration $\mathcal{G}$ with these variables. We start by deriving the dynamics of $\bar{\theta}$ and $\gamma$ under $\mathcal{G}$.

**Theorem 4.3** *Define the innovation process $\overline{W}$ by*

$$\overline{W}_t := \int_0^t \frac{dS_s - (\mu + \alpha\bar{\theta}_s)S_s \, ds}{\sigma_S S_s}. \tag{4.8}$$

*Then $\overline{W}$ is a $\mathcal{G}$-Brownian motion.*

**Proof** From standard filtering theory, see for example Liptser and Shiryaev (2001, Theorem 8.1), $\overline{W}_t$ is a $\mathcal{G}$-Brownian motion for $0 \leq t < T$ and for $T < t < 2T$. $\overline{W}_t$ stays a martingale at $T$ since for $t < T$,

$$\mathbb{E}[\overline{W}_T \mid \mathcal{G}_t] = \overline{W}_t.$$

By Lévy's characterisation theorem of Brownian motion, (see for example, (Protter, 2004, Theorem 39)), $\overline{W}$ is a $\mathcal{G}$-Brownian motion at $T$. An analogous relation holds for $\overline{W}$ at $2T$. $\qquad\square$

Exploring the processes for $\bar{\theta}_t$ and $\gamma_t$ under the setting that $S_t$ and $\theta_t$ follow linear SDEs, as in equations (4.1) and (4.2), has been carried out by several authors, see, for example, Liptser and Shiryaev (2001). This system is referred to as the Kalman-Bucy filter. For $0 \leq t < T$, the SDE for $\bar{\theta}_t$ is derived with $\overline{W}$,

$$d\bar{\theta}_t = (a + b\bar{\theta}_t)\,dt + \frac{\alpha}{\sigma_S}\gamma_t\,d\overline{W}_t, \quad \bar{\theta}_0 = \mathbb{E}[\theta_0 \mid \mathcal{G}_0], \tag{4.9}$$

and for $T \leq t < 2T$, since the value of $\theta_T$ is revealed at $T$, it is given by

$$d\bar{\theta}_t = (a + b\bar{\theta}_t)\,dt + \frac{\alpha}{\sigma_S}\gamma_t\,d\overline{W}_t, \quad \bar{\theta}_T = \theta_T. \tag{4.10}$$

The variance satisfies the following ordinary differential equation, for $0 \leq t < T$,

$$d\gamma_t = 2b\gamma_t + \sigma_\theta^2 - \frac{\alpha^2}{\sigma_S^2}\gamma_t^2, \quad \gamma_0 = \mathbb{E}[(\theta_0 - \bar{\theta}_0)^2 \mid \mathcal{G}_0],$$

and for $T \leq t < 2T$,

$$d\gamma_t = 2b\gamma_t + \sigma_\theta^2 - \frac{\alpha^2}{\sigma_S^2}\gamma_t^2, \quad \gamma_T = 0.$$

For the derivation of these processes, see Theorem 10.1 in Liptser and Shiryaev (2001). In our setting, it has an analytical solution, for $0 \leq t < T$,

$$\gamma_t = \frac{2p\gamma_0 + ((b+p)\gamma_0 + \sigma_\theta^2)(e^{2pt} - 1)}{2p - (b - p - \frac{\alpha^2}{\sigma_S^2}\gamma_0)(e^{2pt} - 1)}, \tag{4.11}$$

and for $T \leq t < 2T$,

$$\gamma_t = \frac{\sigma_\theta^2(e^{2p(t-T)} - 1)}{2p - (b - p)(e^{2p(t-T)} - 1)}, \tag{4.12}$$

where

$$p = 2\sqrt{b^2 + \frac{\sigma_\theta^2}{\sigma_S^2}\alpha^2}.$$

Note that since the actual value of the net position of the zone is revealed at time $T$, $\bar{\theta}_t$ will jump to $\theta_t$ at $t = T$. Similarly at time $2T$, $\bar{\theta}_{2T}$ will jump to $\theta_{2T}$. Using the $\mathcal{G}$-Brownian motion $\overline{W}_t$, the EUA1 price process is written as

$$dS_t = (\mu + \alpha\theta_t)S_t\,dt + \sigma_S S_t\left(d\overline{W}_t - \frac{\alpha(\theta_t - \bar{\theta}_t)}{\sigma_S}\,dt\right)$$

$$= (\mu + \alpha\bar{\theta}_t)S_t\,dt + \sigma_S S_t\,d\overline{W}_t. \qquad (4.13)$$

Now we consider pricing under $\mathcal{G}$. Since $\theta$ is not tradable, this pricing is again considered as being in an incomplete market and the local risk-minimizing approach is capable of dealing with this setting. The minimal martingale measure $\hat{\mathbb{P}}$ used in the complete information setting would not coincide with the martingale measure with respect to the new filtration $\mathcal{G}$. However, there is still a unique minimal martingale measure $\mathbb{P}^*$, which is to be distinguished from $\hat{\mathbb{P}}$. The Radon-Nikodým derivative is given by

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = Z_{2T}^*,$$

where

$$Z_t^* = \exp\left(-\int_0^t \frac{\mu + \alpha\bar{\theta}_s}{\sigma_S}\,d\overline{W}_s - \frac{1}{2}\int_0^t \left(\frac{\mu + \alpha\bar{\theta}_s}{\sigma_S}\right)^2\,dt\right), \quad 0 \le t \le 2T.$$

From the Girsanov-Meyer theorem, under $\mathbb{P}^*$ the process $W_t^*$ is a $\mathcal{G}$-Brownian motion where

$$W_t^* = \overline{W}_t + \int_0^t \frac{\mu + \alpha\bar{\theta}_s}{\sigma_S}\,ds. \qquad (4.14)$$

The underlying processes under $\mathbb{P}^*$ are written, along with $W^*$, as

$$dS_t = \sigma_S S_t\,dW_t^*, \qquad (4.15)$$

$$d\bar{\theta}_t = \left(a - \frac{\alpha\mu}{\sigma_S^2}\gamma_t + \left(b - \frac{\alpha^2}{\sigma_S^2}\gamma_t\right)\bar{\theta}_t\right)dt + \frac{\alpha}{\sigma_S}\gamma_t\,dW_t^*, \qquad (4.16)$$

$$\bar{\theta}_0 = \mathbb{E}^*[\theta_0 \mid \mathcal{G}_0],$$

$$\bar{\theta}_T = \theta_T,$$

$$\bar{\theta}_{2T} = \theta_{2T}.$$

The EUA0 price is defined under the filtration $\mathcal{G}$ as follows:

**Proposition 4.1** *The price of the EUA0 contract at time $t < T$ under incomplete information is written as*

$$P_t = \mathbb{E}^*\big[(S_T + K)1_{\{\theta_T < \kappa\}} + \mathbb{E}^*\big[1_{\{\theta_{2T} < \kappa\}}(S_{2T} + K)\big|\, S_T,\, \theta_T \geq \kappa\big]1_{\{\theta_T \geq \kappa\}}\big|\, \mathcal{G}_t\big],$$
$$(4.17)$$

*where $\mathbb{E}^*[\cdot]$ is the expectation operator under $\mathbb{P}^*$.*

Since $\bar{\theta}_t$ will jump to $\theta_t$ at $T$ and $2T$ and these values, along with the process (4.15), are used for calculating equation (4.17), the joint distributions of $(S_T, \theta_T)$ and $(S_{2T}, \theta_{2T})$ restricted to $(\mathcal{G}_t, \mathbb{P}^*)$ and $(\mathcal{G}_T, \mathbb{P}^*)$, respectively are exactly what we need. The following result shows that these distributions follow normals.

**Theorem 4.4** *The joint distribution of $(\theta_T,\, \log S_T)$ on $(\mathcal{G}_t,\, \mathbb{P}^*)$ is normal.*

**Proof** The joint characteristic function of $(\theta_T,\, \log S_T)$ on $(\mathcal{G}_t,\, \mathbb{P}^*)$ is defined as

$$\psi(\theta_T,\, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} := \mathbb{E}^*\Big[\exp\big(i(\lambda_1 \theta_T + \lambda_2 \log S_T)\big)\, |\, \mathcal{G}_t\Big].$$

Because the minimal martingale measure, $\mathbb{P}^*$, is defined by means of the Radon-Nikodým derivative, the Bayes rule (see Karatzas and Shreve (1991, p. 193)) yields the characteristic function under the original measure, $\mathbb{P}$, as

$$\psi(\theta_T,\, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} = \frac{\mathbb{E}\Big[\exp\big(i(\lambda_1 \theta_T + \lambda_2 \log S_T)\big)Z_T^*\big|\, \mathcal{G}_t\Big]}{Z_t^*}.$$

Using the definition of $\mathcal{G}_t$, $t < T$ and the tower property, it becomes

$$\psi(\theta_T,\, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} = \frac{\mathbb{E}\Big[\mathbb{E}\Big[\exp\big(i(\lambda_1 \theta_T + \lambda_2 \log S_T)\big)Z_T^*\big|\, \mathcal{F}_T^S\Big]\big|\, \mathcal{G}_t\Big]}{Z_t^*}.$$

$\theta_T$ follows a normal distribution on $(\mathcal{F}_T^S, \mathbb{P})$, as shown in Theorem 4.2 and we use the following notation that $\bar{\theta}_{T-} = \lim_{t \to T} \bar{\theta}_t$ and $\gamma_{T-} = \lim_{t \to T} \gamma_t$,

$$\mathbb{E}\big[\exp(i\lambda\theta_T)\,|\, \mathcal{F}_T^S\big] = \exp\Big(i\lambda\bar{\theta}_{T-} - \frac{1}{2}\lambda^2\gamma_{T-}\Big),$$

and so the characteristic function is written as

$$\psi(\theta_T,\, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} = \exp\Big(-\frac{1}{2}\lambda_1^2\gamma_{T-}\Big)\frac{\mathbb{E}\Big[\exp(i(\lambda_1\bar{\theta}_{T-} + \lambda_2 \log S_T))Z_T^*\big|\, \mathcal{G}_t\Big]}{Z_t^*}.$$

We go back to the minimal martingale measure and under this measure the function is written as

$$\psi(\theta_T, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} = \exp\left(-\frac{1}{2}\lambda^2\gamma_{T-}\right)\mathbb{E}^*\left[\exp\left(i(\lambda_1\bar{\theta}_{T-} + \lambda_2\log S_T)\right)\big|\,\mathcal{G}_t\right].$$

The process $\log S_t$ under the minimal martingale measure is derived from equation (4.15) as

$$d\log S_t = -\frac{\sigma_S^2}{2}\,dt + \sigma_S\,dW_t^*, \tag{4.18}$$

and the process $\bar{\theta}_t$ is given from equation (4.16). For notational simplicity, define $\xi_{t,u}$, $X_t$ and $Y_t$ by

$$\xi_{t,u} := \mathbb{E}^*\left[\exp\left(i(\lambda_1 X_u + \lambda_2 Y_u)\right)\big|\,\mathcal{G}_t\right],$$
$$dX_t := (u_1 + u_2 X_t)\,dt + \sigma_X\,dW_t^*,$$
$$dY_t := v_1\,dt + \sigma_Y\,dW_t^*.$$

Since both processes are Markov and satisfy the Lipschitz condition, the Feynman-Kac formula (see Karatzas and Shreve (1991, p. 268) for example) yields the following partial differential equation with respect to $\xi_{t,u}$,

$$\frac{\partial\xi}{\partial t} + (u_1 + u_2 X_t)\frac{\partial\xi}{\partial X_t} + v_1\frac{\partial\xi}{\partial Y_t} + \frac{1}{2}\left(\sigma_X^2\frac{\partial^2\xi}{\partial X_t^2} + 2\sigma_X\sigma_Y\frac{\partial^2\xi}{\partial Y_t\partial X_t} + \sigma_Y^2\frac{\partial^2\xi}{\partial Y_t^2}\right) = 0, \tag{4.19}$$

with boundary condition,

$$\xi_{u,u} = \exp\left(i(\lambda_1 X_u + \lambda_2 Y_u)\right).$$

Since $X_t$ and $Y_t$ are linear SDEs, we make the following ansatz:

$$\xi_{t,u} = \exp\left(A_{t,u} + B_{t,u}X_t + C_{t,u}Y_t\right),$$

where $A_{t,u}$, $B_{t,u}$ and $C_{t,u}$ are deterministic. Each term in the partial differential

equation is calculated as

$$\frac{\partial \xi}{\partial t} = \xi \left( \frac{dA_{t,u}}{dt} + \frac{dB_{t,u}}{dt} X_t + \frac{dC_{t,u}}{dt} Y_t \right),$$

$$\frac{\partial \xi}{\partial X_t} = \xi B_{t,u},$$

$$\frac{\partial \xi}{\partial Y_t} = \xi C_{t,u},$$

$$\frac{\partial^2 \xi}{\partial X_t^2} = \xi B_{t,u}^2,$$

$$\frac{\partial^2 \xi}{\partial Y_t^2} = \xi C_{t,u}^2,$$

$$\frac{\partial^2 \xi}{\partial X_t \partial Y_t} = \xi B_{t,u} C_{t,u}.$$

(4.20)

By substituting equation (4.20) into equation (4.19) and remembering that this equation is to be satisfied by any $X_t$ and $Y_t$, the partial differential equation is reduced to the system of ordinary differential equations for $A_{t,u}$, $B_{t,u}$ and $C_{t,u}$,

$$\frac{dA_{t,u}}{dt} + u_1 B_{t,u} + v_1 C_{t,u} + \frac{\sigma_X^2}{2} B_{t,u}^2 + \frac{\sigma_Y^2}{2} C_{t,u}^2 + \sigma_X \sigma_Y B_{t,u} C_{t,u} = 0,$$

$$\frac{dB_{t,u}}{dt} + u_2 B_{t,u} = 0,$$

$$\frac{dC_{t,u}}{dt} = 0,$$

and boundary conditions with respect to each variable are computed from

$$\xi_{u,u} = \exp\big(i(\lambda_1 X_u + \lambda_2 Y_u)\big),$$

so that we have

$$A_{u,u} = 0,$$

$$B_{u,u} = i\lambda_1,$$

$$C_{u,u} = i\lambda_2.$$

These ordinary differential equations are solved as

$$A_{t,u} = i\lambda_1 \int_t^u \exp\left(\int_s^u u_2\,du\right) u_1\,ds + i\lambda_2 \int_t^u v_1\,ds$$

$$- \frac{1}{2}\lambda_1^2 \int_t^u \exp\left(\int_s^u 2u_2\,du\right) \sigma_X^2\,ds - \frac{1}{2}\lambda_2^2 \int_t^u \sigma_Y^2\,ds$$

$$- \lambda_1\lambda_2 \int_t^u \exp\left(\int_s^u u_2\,du\right) \sigma_X\sigma_Y\,ds,$$

$$B_{t,u} = i\lambda_1 \exp\left(\int_t^u u_2\,du\right),$$

$$C_{t,u} = i\lambda_2.$$

By computing the partial derivatives of $\xi_{t,u}$ and substituting them into the equation (4.19), the partial differential equation is separable in $X_t$ and $Y_t$ and the ordinary differential equations are obtained by identifying coefficients.

As a result, the characteristic function is written as

$$\psi(\theta_T, \log S_T)_{(\mathcal{G}_t, \mathbb{P}^*)} = \exp\left(i\lambda_1\left(\exp(K_{t,T})\bar{\theta}_t + \int_t^T \exp(K_{s,T})\left(a - \frac{\alpha\mu\gamma_s}{\sigma_S^2}\right) ds\right)\right.$$

$$+ i\lambda_2\left(\log S_t - \frac{1}{2}\sigma_S^2(T-t)\right) - \frac{1}{2}\lambda_1^2\left(\gamma_{T-} + \int_t^T \exp(2K_{s,T})\frac{\alpha^2}{\sigma_S^2}\gamma_s^2\,ds\right)$$

$$- \frac{1}{2}\lambda_2^2\sigma_S^2(T-t) - \lambda_1\lambda_2 \int_t^T \exp(K_{s,T})\alpha\gamma_s\,ds\bigg), \tag{4.21}$$

where $K_{t,T} = \int_t^T \left(b - \frac{\alpha^2}{\sigma_S^2}\gamma_u\right) du$. It verifies that $(\theta_T, \log S_T)$ on $(\mathcal{G}_t, \mathbb{P}^*)$ follows a joint normal distribution. $\qquad\square$

In the same way, we can prove $(\log S_{2T}, \theta_{2T})$ on $(\mathcal{G}_T, \mathbb{P}^*)$ also follows a joint normal distribution.

Theorem 4.4 shows all characteristics of the normal distribution, that is the means and variances of $\log S_T$ and $\theta_T$ and covariance between them on $(\mathcal{G}_t, \mathbb{P}^*)$. The mean of $\log S_T$ is specified in the $i\lambda_2$ term in equation (4.21) as

$$\log S_t - \frac{1}{2}\sigma_S^2(T-t).$$

The variance of $\log S_T$ is obtained from the $-1/2\lambda_2^2$ term in equation (4.21) as

$$\sigma_S^2(T-t).$$

From Theorem 4.4, the mean and variance of $\theta_T$ is specified using the $i\lambda_1$ and $-1/2\lambda_1^2$ terms respectively in equation (4.21) as

$$\exp(K_{t,T})\bar{\theta}_t + \int_t^T \exp(K_{s,T})\left(a - \frac{\alpha\mu\gamma_t}{\sigma_S^2}\right) ds,$$

and
$$\exp(K_{t,T})\bar{\theta}_t + \int_t^T \exp(K_{s,T})\left(a - \frac{\alpha\mu\gamma_t}{\sigma_S^2}\right) ds,$$
$$\gamma_{T-} + \int_t^- \exp(2K_{s,T})\frac{\alpha^-}{\sigma_S^2}\gamma_s^2 ds.$$

Their covariance arises from the $-\lambda_1\lambda_2$ term in equation (4.21) as

$$\int_t^T \exp(K_{s,T})\alpha\gamma_s ds.$$

By computing the means, variances and covariance numerically and applying Monte Carlo simulation, we can price the EUA0 contracts. To compute these variables, we use the trapezium rule with equi-distant steps. The Monte Carlo algorithm is summarized below:

**Step 1:** Compute numerically means, variances and covariances in $[t, T]$.

**Step 2:** Simulate $S_t$ and $\theta_t$ at $T$ by generating a two-dimensional normal random number computed in Step 1.

**Step 3:** If $\theta_T < \kappa$, go to Step 4, otherwise simulate $S_t$ and $\theta_t$ at $2T$ by generating two-dimensional normal random number with $S_T$ and $\theta_T$ already generated in Step 2.

**Step 4:** Calculate the EUA0 price, $P_i$, with equation (4.17).

**Step 5:** Repeat Step 2 to Step 4 independently $M$ times and obtain independent samples $P_i$, $i = 1, 2, \ldots, M$. The EUA0 price is then calculated by the experimental average,

$$EUA0^{MC} = \frac{1}{M} \sum_{i=1}^{M} P_i.$$

Extension to a multi-period setting is straightforward by repeating Step 3. The details of several possible extensions are discussed in Section 4.5.

## 4.3   Parameter Estimation

In this section, we estimate the parameters of the proposed model. There are two means of estimating parameters: one is with historical data, the other is with contemporaneous day-to-day market data. In the former approach, parameters are estimated from time-series of associated data. An alternative is to estimate parameters with the data of a contemporaneous day, or implied data. In this approach, parameters are estimated with derivative prices of the day. Derivative prices are functions of variables so that the parameters are implied from their prices. Since there are not enough liquid derivatives markets for estimating all parameters, we estimate all parameters from historical data and use them for pricing.

Since an investor can observe only the EUA1 prices, we are in the situation of estimating parameters with the Kalman filter in which the observable process and unobservable process are $S_t$ and $\theta_t$ respectively. Given the setting that the unobservable variable follows a linear diffusion and the log increment of the observable variable, conditioned on the unobservable variable, follows a linear diffusion, it is possible to estimate parameters by maximizing the likelihood function. We adopt the EM algorithm for maximizing the likelihood function. After reviewing the algorithm for finding maximum likelihood estimators, we estimate parameters.

The Kalman filter consists of two equations: state and measurement. The state equation expresses the dynamics of the unobservable variable and the measurement equation expresses the dynamics of the observable process. To apply the EM algorithm, we derive the associated discrete-time version of the SDEs. Partition the interval $[0, T]$ at the points $0 = t_0 < t_1 < \cdots < t_N = T$ where $t_k = k\Delta$, for $k = 0, 1, \ldots, N$ and $\Delta = T/N$, the length of the discrete time interval. The state equation is obtained by discretizing equation (4.2) with the Euler–Maruyama scheme as

$$\theta_{k+1} = \theta_k + (a + b\theta_k)\Delta + \sigma_\theta(B_{k+1} - B_k)$$
$$:= A\theta_k + B + w_k, \tag{4.22}$$

where $A := b\Delta + 1$ and $B := a\Delta$. $w_k$ follows a normal distribution whose mean

and variance are zero and $Q := \sigma_\theta^2 \Delta$, respectively. As for the measurement equation, equation (4.1) is discretized as follows:

$$S_{k+1} = S_k \exp\left(\left(\mu + \alpha\theta_k - \frac{1}{2}\sigma_S^2\right)\Delta + \sigma_S(W_{k+1} - W_k)\right).$$

The log increment of the observable process is thus calculated as

$$\log S_{k+1} - \log S_k = \left(\mu + \alpha\theta_k - \frac{1}{2}\sigma_S^2\right)\Delta + \sigma_S(W_{k+1} - W_k).$$

For notational simplicity, the log increment of the observable process, $y_t$, is defined as follows

$$y_k := \log S_{k+1} - \log S_k = C\theta_k + D + v_k, \tag{4.23}$$

where $C := \alpha\Delta$ and $D := (\mu - 1/2\sigma_S^2)\Delta$. $v_k$ follows a normal distribution whose mean and variance are zero and $R := \sigma_S^2\Delta$, respectively. From the assumption in Section 4.2, $w_k$ and $v_k$ are independent. This system generates sequences of observations in the following way: the initial state of the unobservable variable is determined probabilistically. At each time step, the system produces an unobservable state $\theta_k$ from $\theta_{k-1}$ one step before, according to the state equation. Once $\theta_k$ is determined, the observation is produced according to the measurement equation.

## 4.3.1  EM algorithm

We explain the EM algorithm used to estimate the parameters of the models. For models with incomplete data in general, the EM algorithm is more popular than gradient ascent approaches. Originally developed by Dempster, Laird and Rubin (1977), it is a numerical method for calculating maximum likelihood parameter estimates of partially observed models or with missing data. It computes the incomplete data log-likelihood via the iterative maximization of the expected complete data log-likelihood, conditional on the observed data. It has the appealing property that an iteration does not decrease the value of the likelihood function. Convergence results for the EM algorithm under general conditions are investigated in Wu (1983).

Suppose that we observe data $y_0, \ldots, y_N$ (or in short, $y_{0:N}$), which are generated from the unobservable sequence $\theta_{0:N}$. Let $\Lambda$ be a family of parameters.

121

The algorithm consists of an E step and an M step at each iteration summarized below:

**Step 1:** Set initial value $\lambda_0 \in \Lambda$.

**Step 2: (E step)** Evaluate the expected complete data log-likelihood function.

$$L(\lambda, \lambda_n) = \mathbb{E}\left[\log \mathbb{P}\left(y_{0:N}, \theta_{0:N} \mid \lambda\right) \mid y_{0:N}, \lambda_n\right]. \qquad (4.24)$$

**Step 3: (M step)** Derive $\lambda_{n+1}$ from $\lambda_n$ which takes the following form:

$$\lambda_{n+1} = \arg\max_{\lambda \in \Lambda} L(\lambda, \lambda_n).$$

**Step 4** Repeat E and M steps until a stopping criterion is satisfied.

Given initialisations of the parameters and the observed data, in the expectation part of the algorithm, the EM algorithm starts by calculating the smoothers of a particular state. With the smoothers, the algorithm constructs the expected complete data log-likelihood function. The expected complete data log-likelihood function is then maximized in the maximization part of the algorithm in order to obtain new estimates of the parameters. This is done by equating the first-order derivatives, with respect to the parameters, to zero. Using these new estimates, the algorithm returns to the expectation part again until new estimates satisfy some stopping rule. In most cases, however, the EM algorithm is not guaranteed to converge to the global optimum. Instead, it stops at a local optimum. The extent to which the algorithm correctly estimates parameters is reflected by the initial set of parameters. Initialization of the algorithm is therefore an important consideration. We run multiple procedures of the EM algorithm with different initial values and compare the likelihood of different convergences.

In order to apply the EM algorithm outlined above to our setting, we derive the expected complete data log-likelihood function. Based on equations (4.22) and (4.23), $\theta_k$ follows a normally-distributed random variable with mean $A\theta_{k-1} + B$ and variance $Q$, conditioned on $\theta_{k-1}$, and similarly conditioned on $\theta_k$, $y_k$ follows a normally-distributed random variable with mean $C\theta_k + D$ and variance

$R$,

$$\mathbb{P}(\theta_k \mid \theta_{k-1}, \lambda) = \frac{1}{\sqrt{2\pi Q}} \exp\left(-\frac{(\theta_t - A\theta_{k-1} - B)^2}{2Q}\right),$$

$$\mathbb{P}(y_k \mid \theta_k, \lambda) = \frac{1}{\sqrt{2\pi R}} \exp\left(-\frac{(y_k - C\theta_k - D)^2}{2R}\right).$$

Using the Markov property, the joint probability is

$$\mathbb{P}(y_{0:N}, \theta_{0:N} \mid \lambda) = \mathbb{P}(\theta_0 \mid \lambda) \prod_{k=1}^{N} \mathbb{P}(\theta_k \mid \theta_{k-1}, \lambda)\, \mathbb{P}(y_k \mid \theta_k, \lambda).$$

By assuming that the unconditional distribution for the initial state $\theta_0$ follows a normal distribution with mean $\tilde{\theta}_0$ and variance $Q_0$,

$$\mathbb{P}(\theta_0 \mid \lambda) = \frac{1}{\sqrt{2\pi Q_0}} \exp\left(-\frac{(\theta_0 - \tilde{\theta}_0)^2}{2Q_0}\right),$$

the log of the joint probability is written as,

$$\begin{aligned}
\log \mathbb{P}(\theta_{0:N}, y_{0:N} \mid \lambda) = &-\sum_{k=0}^{N} \frac{(y_k - C\theta_k - D)^2}{2R} - \frac{N+1}{2} \log |R| \\
&- \sum_{k=1}^{N} \frac{(\theta_k - A\theta_{k-1} - B)^2}{2Q} - \frac{N}{2} \log |Q| \\
&- \frac{(\theta_0 - \tilde{\theta}_0)^2}{2Q_0} - \frac{1}{2} \log |Q_0| - (N+1) \log 2\pi.
\end{aligned}$$

The set of parameters is given by $\Lambda = (A, B, Q, C, D, R, \tilde{\theta}_0, Q_0)$. In the E-step, the expected log-likelihood function, (4.24), can be written as

$$\begin{aligned}
L(\lambda, \lambda_n) = &\ \mathbb{E}\left[\log \mathbb{P}\left(\theta_{0:N}, y_{0:N} \mid \lambda\right) \mid y_{0:N}, \lambda_n\right] \\
= &-\frac{1}{2} \mathbb{E}\left[\sum_{k=0}^{N} \frac{(y_k - C\theta_k - D)^2}{R} \,\Bigg|\, y_{0:N}, \lambda_n\right] - \frac{N+1}{2} \log |R| \\
&- \frac{1}{2} \mathbb{E}\left[\sum_{k=1}^{N} \frac{(\theta_k - A\theta_{k-1} - B)^2}{Q} \,\Bigg|\, y_{0:N}, \lambda_n\right] - \frac{N}{2} \log |Q| \\
&- \frac{1}{2} \mathbb{E}\left[\frac{(\theta_0 - \tilde{\theta}_0)^2}{Q_0} \,\Bigg|\, y_{0:N}, \lambda_n\right] - \frac{1}{2} \log |Q_0| - (N+1) \log 2\pi. \quad (4.25)
\end{aligned}$$

In the M-step, the new set of parameters is given by differentiating $L$ with respect to each parameter, setting to zero, and solving. To do so, we need three

smoothers, $\theta_{k|N}$, $P_{k|N}$ and $P_{k,k-1|N}$, denoted respectively by the symbols:

$$\theta_{i|j} := \mathbb{E}[\theta_i \,|\, y_{0:j}, \lambda_n], \tag{4.26}$$

$$P_{i|j} := \mathbb{V}\mathrm{ar}[\theta_i \,|\, y_{0:j}, \lambda_n], \tag{4.27}$$

$$P_{i+1,i|j} := \mathbb{C}\mathrm{ov}[\theta_{i+1}\theta_i \,|\, y_{0:j}, \lambda_n], \quad i \le j. \tag{4.28}$$

These quantities are called smoother estimates in the sense that they are estimates of unobservable variables given the future observation and are obtained through the Kalman smoother procedure. To obtain smoother estimates, it requires the filter estimates $\theta_{k|k}$ and $P_{k|k}$. They are called filter estimates in the sense that they are estimates of unobservable variables at time $k$ given observations up to time $k$. The difference between the Kalman filter and smoother lies in the fact that the recursion in the filter moves forward and in the smoother moves backward. In other words, the smoother estimates differ from the ones obtained in the filter in that they depend on past and future observations. Consequently, in the E-step, we first implement the Kalman filter to obtain filter estimates $\theta_{k|k}$ and $P_{k|k}$, for $k = 1, \ldots, N$, and then implement the Kalman smoother, along with filtered estimates, to obtain smoother estimates $\theta_{k|N}, P_{k|N}$ and $P_{k+1,k|N}$, for $k = 0, \ldots, N-1$ with the current set of parameters. The full procedures are explained later.

By differentiating equation (4.25) with respect to each parameter, setting it to zero and solving, new parameters which appear with smoother estimates are obtained. For the state equation, this results in the following:

$$\hat{A} = \frac{\sum_{k=1}^{N} \theta_{k|N}\theta_{k-1|N} + P_{k,k-1|N} - \hat{B}\theta_{k-1|N}}{\sum_{k=1}^{N} \theta_{k-1|N}^2 + P_{k-1|N}}, \tag{4.29}$$

$$\hat{B} = \frac{\sum_{k=1}^{N} \theta_{k|N} - \dfrac{\sum_{k=1}^{N} \theta_{k|N}\theta_{k-1|N} + P_{k,k-1|N}\sum_{k=1}^{N} \theta_{k-1|N}}{\sum_{k=1}^{N} \theta_{k-1|N}^2 + P_{k-1|N}}}{N - \dfrac{\sum_{k=1}^{N} \theta_{k-1|N}\sum_{k=1}^{N} \theta_{k-1|N}}{\sum_{k=1}^{N} \theta_{k-1|N}^2 + P_{k-1|N}}}, \tag{4.30}$$

$$\hat{Q} = \frac{1}{N}\sum_{k=1}^{N} \theta_{k|N}^2 + P_{k|N} - \hat{A}(\theta_{k|N}\theta_{k-1|N} + P_{k,k-1|N}) - \hat{B}\theta_{k|N}, \tag{4.31}$$

$$\hat{\tilde{\theta}}_0 = \theta_{0|N}, \tag{4.32}$$

$$\hat{Q}_0 = P_{0|N}. \tag{4.33}$$

124

New parameters of the measurement equation are given as follows.

$$\hat{C} = \frac{\sum_{k=0}^{N} y_k \theta_{k|N} - \hat{D}\theta_{k|N}}{\sum_{k=0}^{N} \theta_{k|N}^2 + P_{k|N}}, \tag{4.34}$$

$$\hat{D} = \frac{\sum_{k=0}^{N} y_k - \dfrac{\sum_{k=0}^{N} y_k \theta_{k|N} \sum_{k=0}^{N} \theta_{k|N}}{\sum_{k=0}^{N} \theta_{k|N}^2 + P_{k|N}}}{N + 1 - \dfrac{\sum_{k=0}^{N} \theta_{k|N} \sum_{k=0}^{N} \theta_{k|N}}{\sum_{k=0}^{N} \theta_{k|N}^2 + P_{k|N}}}, \tag{4.35}$$

$$\hat{R} = \frac{1}{N+1} \sum_{k=0}^{N} y_k^2 - \hat{C} y_k \theta_{k|N} - \hat{D} y_k. \tag{4.36}$$

New parameters are set as $\lambda_{n+1} = (\hat{A}, \hat{B}, \hat{Q}, \hat{C}, \hat{D}, \hat{R}, \hat{\tilde{\theta}}_0, \hat{Q}_0)$. Acquiring these updated parameters from $\lambda_n$ requires three quantities, $\theta_{k|N}$, $P_{k|N}$ and $P_{k,k-1|N}$, which are included in equations (4.29)-(4.36) and are obtained with the Kalman smoother algorithm. With these parameters, we return to the E-step. See the derivations of the EM algorithm for the Kalman filter in Shumway and Stoffer (2006, Chapter 4).

In our setting, the stopping criterion used to stop the algorithm is the absolute difference of the maximized log-likelihood functions from one step before being less than $10^{-4}$. That is, the algorithm is stopped if

$$\left| l(y_{0:N} \mid \lambda_{n+1}) - l(y_{0:N} \mid \lambda_n) \right| < 10^{-4}.$$

From the assumption, the density function of $y_k$ conditional on $y_{k-1}$, $k = 1, \ldots, N$, follows a normal distribution with mean and variance

$$y_{k|k-1} := C\theta_{k|k-1} + D,$$
$$F_{k|k-1} := C^2 P_{k|k-1} + R,$$

respectively, and that of $y_0$ follows a normal distribution with mean and variance

$$y_{0|-1} := C\tilde{\theta}_0 + D,$$
$$F_{0|-1} := C^2 Q_0 + R,$$

respectively. As a result, the log-likelihood function, $l(y_{0:N} \mid \lambda)$, is computed

with the Markov property of $y_k$ as

$$l(y_{0:N} \mid \lambda) = \log \mathbb{P}(y_0 \mid \lambda) \prod_{k=1}^{N} \mathbb{P}(y_k \mid y_{0:k-1}, \lambda)$$

$$= -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{k=1}^{N} \log |F_{k|k-1}| - \frac{1}{2} \sum_{k=1}^{N} \frac{(y_k - y_{k|k-1})^2}{F_{k|k-1}}$$

$$- \frac{1}{2} \log 2\pi - \frac{1}{2} \log |F_{0|-1}| - \frac{1}{2} \frac{(y_0 - y_{0|-1})^2}{F_{0|-1}}. \tag{4.37}$$

## Kalman Filter and Smoother

In the Kalman filter, filter estimates are given recursively from $k = 1$ to $k = N$. The prediction equations are given below with initial values $\theta_{0|0} = \tilde{\theta}_0$ and $P_{0|0} = Q_0$. From $k = 1$ to $k = T$,

$$\theta_{k|k-1} = A\theta_{k-1|k-1} + B,$$

$$P_{k|k-1} = A^2 P_{k-1|k-1} + Q.$$

The updating equations are given by

$$\theta_{k|k} = \theta_{k|k-1} + K_k \left( y_k - (C\theta_{k|k-1} - D) \right),$$

$$P_{k|k} = P_{k|k-1} - K_k P_{k|k-1} C,$$

where $K_k$ is called the Kalman gain and is defined as

$$K_k = \frac{P_{k|k-1}C}{C^2 P_{k|k-1} + R}.$$

In the Kalman smoother, smoother estimates are given backwards from $k = N$ to $k = 0$. Note that at $k = N$, initial values of the smoother are the same as the filter, $\theta_{N|N}$ and $P_{N|N}$ and define the initial covariance, $P_{N,N-1|N} = (1 - K_N C)AP_{N-1|N-1}$. From $k = N - 1$ to $k = 0$,

$$\theta_{k|N} = \theta_{k|k} + J_k(\theta_{k+1|N} - \theta_{k+1|k}),$$

$$P_{k|N} = P_{k|k} + J_k^2(P_{k+1|N} - P_{k+1|k}),$$

$$P_{k,k-1|N} = J_{k-1}P_{k|k} + J_k J_{k-1}(P_{k+1,k|N} - AP_{k|k}),$$

$$J_k = \frac{P_{k|k}A}{P_{k+1|k}}.$$

Using these smoother estimates, equations (4.29)-(4.36) are computed.

It is worth mentioning that maximizing the likelihood function directly also works for estimating parameters. A direct approach is to find parameters, $\lambda = \{A, B, Q, C, D, V, \tilde{\theta}_0, Q_0\}$, so as to maximize the log-likelihood function with respect to the parameters we want to estimate. This is done by differentiating the log-likelihood function numerically, and performing the gradient ascent. In this maximization, only the Kalman filter is used and the Kalman smoother need not be implemented. Because of the recursive nature of the Kalman filter, the dependence of the log-likelihood on parameters is complicated and maximization is usually done numerically. See Harvey (1991, Chapter 3) for deriving the direct maximization.

## 4.3.2 Data and Estimation Results

The spot, futures and options contracts are traded both over-the-counter (OTC) and on exchanges. Parameters are estimated from the same data as in Chapter 3. Parameter estimation based on the EM algorithm described in the previous subsection is performed. We set 240 trading days per year, $\Delta = 1/240$. Estimates are presented in Table 4.1. The maximum likelihood estimates appear in the first row and the corresponding asymptotic standard deviations calculated from the outer product and Hessian matrix appear in the second row.

Table 4.1: Estimation results of equations (4.22) and (4.23)

| $A$ | $B$ | $Q$ |
|---|---|---|
| 0.941 | $-0.014$ | 0.267 |
| 0.0053 | 0.0133 | 0.6796 |

| $C$ | $D$ | $R$ |
|---|---|---|
| $-0.00023$ | $-0.00027$ | 0.00059 |
| $1.219 \times 10^{-4}$ | $7.842 \times 10^{-5}$ | $1.502 \times 10^{-5}$ |

The log-likelihood = 986.9.

An obvious shortcoming of adopting a continuous-valued process for expressing the net position of the zone is that identifying the extent of being long or

short is difficult from data, although a linear diffusion approach renders express-
ing the extent of being long or short of the net position of the zone feasible.
The EM algorithm leaves $\kappa$ unspecified. After estimating all parameters except
for $\kappa$ with the EM algorithm, we estimate $\kappa$ from historical spot prices so as to
minimize the pricing error defined as

$$\min \sum_t \left( P_t^{\text{model}}(\kappa) - P_t^{\text{market}} \right)^2,$$

where $P_t^{\text{market}}$ is the spot price traded in the market. Since data period is in
Phase I, $P_t^{\text{model}}$ is the spot price defined as

$$P_t^{\text{model}} = \mathbb{E}^* \left[ (S_T + K) 1_{\{\theta_T < \kappa\}} | \mathcal{G}_t \right].$$

Corresponding parameters in continuous-time are given in Table 4.2.

Table 4.2: Parameters in continuous-time

| $\alpha$ | $\mu$ | $\sigma_s$ | $a$ | $b$ | $\sigma_\theta$ | $\kappa$ |
|---|---|---|---|---|---|---|
| $-0.056$ | $0.005$ | $0.376$ | $-3.37$ | $-11.89$ | $7.99$ | $-2.26$ |

## 4.4 Numerical Studies for EUA Prices

The purpose of this section is to examine price characteristics between complete
information and incomplete information. Emphasis will be placed on experi-
ments which highlight differences in key quantities such as initial prices and
drift coefficients. The EUA0 price under the complete information setting is
given in equation (4.7) and that under the incomplete information setting in
equation (4.17). For prices under complete information, we compute $G$ and
$\hat{\mathbb{P}}\{\theta_T \geq \kappa \mid \mathcal{F}_t\}$ in equation (4.7) with Monte Carlo simulation. For prices un-
der incomplete information, we generate the two-dimensional random numbers
following normal in equation (4.21) after computing the means, variances and
covariance. The results indicate some differences between the two settings might
be of interest. We use the following set of parameters shown in Table 4.3. Monte
Carlo prices are obtained with 100,000 simulations.

Table 4.3: Parameters in numerical studies

| $s_0$ | $\alpha$ | $\mu$ | $\sigma_s$ | $a$ | $b$ | $\sigma_\theta$ | $K$ | $\theta_0$ | $\gamma_0$ | $\kappa$ | $T$ |
|-------|----------|-------|------------|------|-------|-----------------|-----|------------|------------|----------|-----|
| 30 | $-0.06$ | 0.005 | 0.4 | $-3.0$ | $-11.4$ | 8.2 | 100 | 0.07 | 2.9 | $-2.0$ | 1.0 |

Figures 4.1, 4.2, 4.3, 4.4 and 4.5 show the prices under complete information and incomplete information with respect to $s_0$, $a$, $b$, $\sigma_\theta$ and $\kappa$, respectively. The overall impression is that their sensitivities have similar patterns.

Prices under incomplete information are, in the sense that they depend on the coefficients of $S$, $\mu$, $\alpha$ and $\sigma_S$, different to those under complete information. Figures 4.6, 4.7 and 4.8 show the sensitivities with respect to $\mu$, $\alpha$ and $\sigma_S$, respectively. Since prices under complete information do not depend on these parameters, they are flat, while prices under incomplete information do depend on these parameters. Since equation (4.16) indicates that a higher value of $\mu$ leads to a lower value of $\bar{\theta}$ with our parameters, the higher value of $\mu$ leads to the lower EUA0 price as shown in Figure 4.6. Both $\alpha$ and $\sigma_S$ are included in equation (4.16) as a quadratic form, leading to the shape of quadratic function as in Figures 4.7 and 4.8.

Another interesting phenomenon is the jump at $T$ in the conditional expectation, which occurs when the verified emissions are announced. Consequently, it causes the discontinuous changes in the EUA0 price since the cash flow at $T$ is determined based on $\theta_T$ and the initial value to compute the expectation at $2T$ will vary. We confirm this on a sample path by sample path basis. At $T$, the true position of the zone is revealed to the market and $\bar{\theta}_{T-}$ will jump to the revealed value $\theta_T$. Simulation is done as follows. The EUA1 price and process of the conditional expectation under $\mathbb{P}$ are simulated with equations (4.13) and (4.9). At each $t_i$, with $(S_{t_i}, \bar{\theta}_{t_i})$, the EUA0 price is computed using Monte Carlo simulation. Figure 4.10 presents the sample paths for the EUA0 and EUA1 prices. Assume at $T$, the verified emissions amount is revealed and the market is long. Consequently, the process of the conditional expectation jumps as shown in Figure 4.9 and the prices for EUA0 is discontinuous as shown in Figure 4.10.

Figure 4.1: The EUA0 price with various initial values, $s_0$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
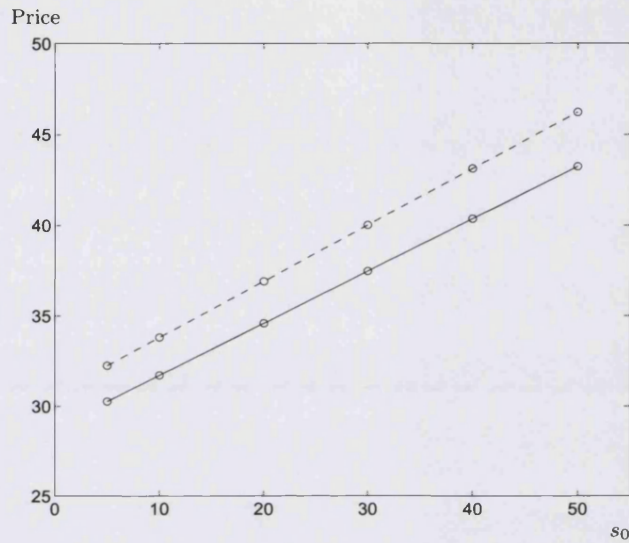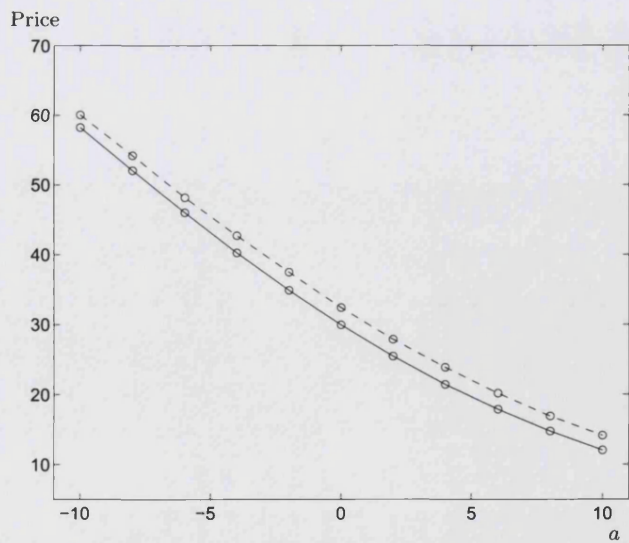


Figure 4.2: The EUA0 price with various initial values, $a$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.

Figure 4.3: The EUA0 price with various initial values, $b$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
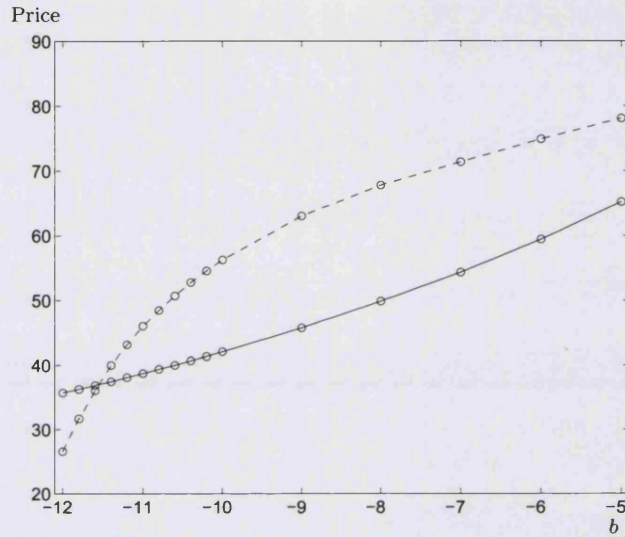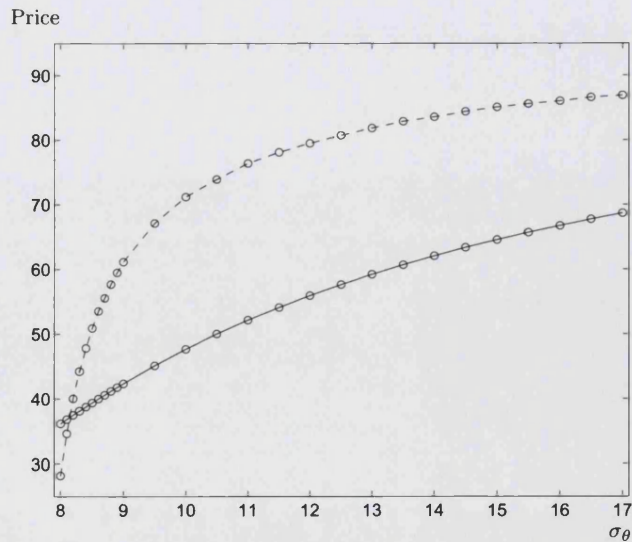


Figure 4.4: The EUA0 price with various initial values, $\sigma_\theta$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.

Figure 4.5: The EUA0 price with various initial values, $\kappa$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
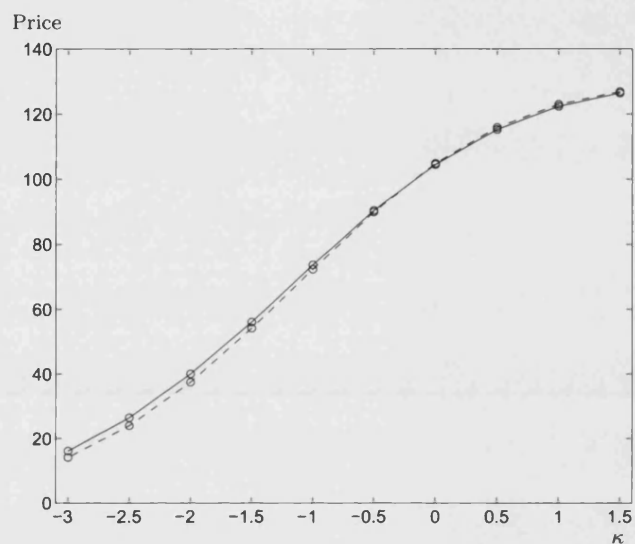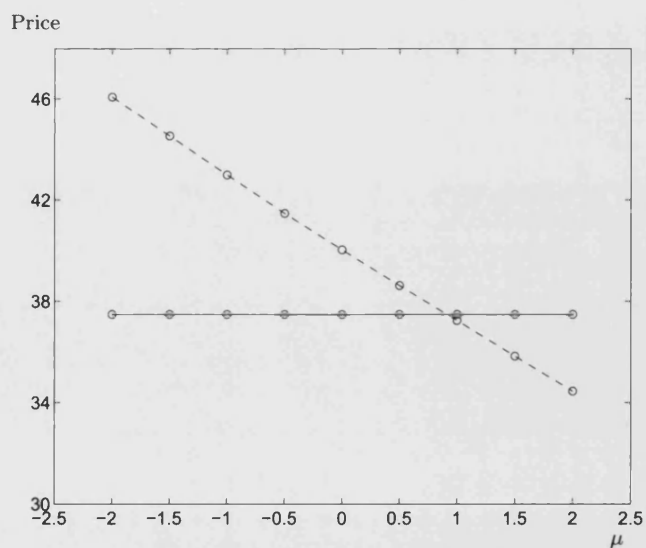


Figure 4.6: The EUA0 price with various initial values, $\mu$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.

Figure 4.7: The EUA0 price with various initial values, $\alpha$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
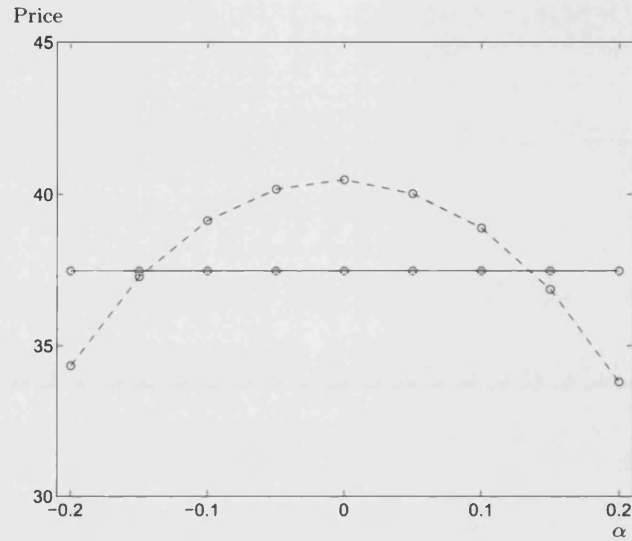


Figure 4.8: The EUA0 price with various initial values, $\sigma_S$. Solid line represents the price under the complete information setting and dashed line under the incomplete information setting.
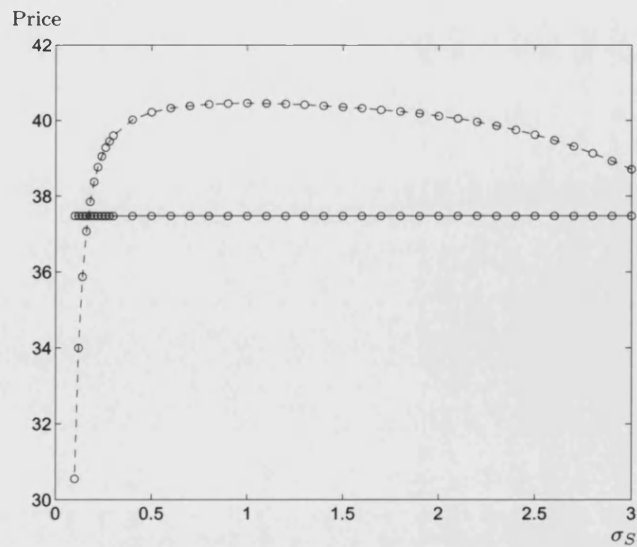
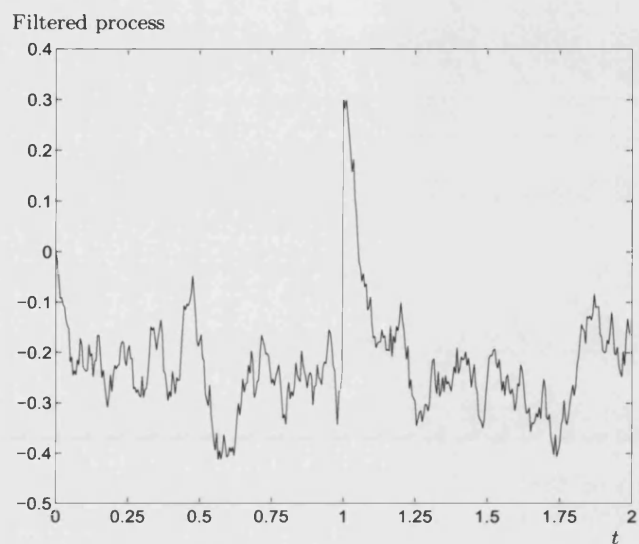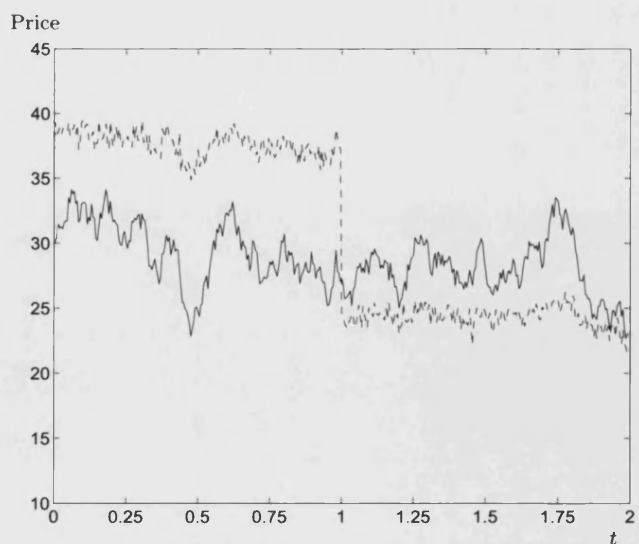Figure 4.9: Simulated process of the conditional expectation.



Figure 4.10: Simulated EUA0 and EUA1 prices. Solid line represents the EUA1 price and dashed line the EUA0 price.



134

## 4.5 Concluding Remarks

In this chapter, we consider the spot and forward relationship in the carbon emissions market based upon the no-arbitrage principle. This market differs due to the trading regulations and the market design from other financial markets: the net position of the zone is taken into consideration for explaining the relationship. We assume the net position of the zone follows a linear diffusion and the drift term of the forward price process depends on the net position of the zone. Bankability is included in the framework in the form of an expectation of future cash-flows.

Two information settings are considered: complete and incomplete. The complete information setting is defined when market participants can access full information. We adopt the locally risk-minimization technique so as to fix the equivalent martingale measure since the net position of the zone is not tradable. With the minimal martingale measure, the analytical arbitrage-free price is obtained. Under the incomplete information setting when market participants can observe the forward process and the net position of the zone only at the announcement time, a filtering technique is applied to derive the optimal projection of the net position of the zone onto the new filtration. This leads to the use of the Kalman-Bucy filter. The joint distribution of the underlying processes is derived and it follows normal under the minimal martingale measure. Consequently, the arbitrage-free price is obtained using Monte Carlo simulation.

There could be several extensions without changing the framework. The assumption that there are only two trading periods can easily be extended to multi-periods. In the multi-period setting when banking is prohibited only at the end of the phase and intra-period banking is available, the payoff function has a nested structure. We denote by $T_N$ the end of the phase when the banking is prohibited and by $T_1, \ldots, T_{N-1}$ the end of each trading year when intra-banking is available. The payoff function can be written as follows:

$$P_{T_N} = \begin{cases} S_{T_N} + K & \text{if} \quad \theta_{T_N} < \kappa, \\ 0 & \text{if} \quad \theta_{T_N} \geq \kappa. \end{cases}$$

$$P_{T_i} = \begin{cases} S_{T_i} + K & \text{if} \quad \theta_{T_i} < \kappa, \\ \mathbb{E}\big[P_{T_{i+1}} \big| S_{T_i}, \theta_{T_i} \geq \kappa\big] & \text{if} \quad \theta_{T_i} \geq \kappa, \quad i = 1, \ldots, N-1. \end{cases}$$

A discount factor, or interest rate, can be included in the framework. A deterministic or stochastic interest rate which is independent of EUA prices is easily implemented in the form of a zero-coupon bond. To implement a stochastic interest rate which is correlated with EUA prices, a change of measure technique can be used to manipulate the equations.

Since the market for the EU ETS has become considerably more active, the link between EUA contracts and other carbon contracts has deepened. For example, the EU ETS scheme not only allows companies to trade among themselves, but enables interactions with other countries not included in the EU ETS via the Clean Development Mechanism and Joint Implementation schemes. During Phase II, firms can cover their requirements by purchasing Certified Emission Reduction Credits from such projects. The amount of these contracts are growing, and therefore their pricing is important.

# Chapter 5

# Summaries and Conclusions

This dissertation consists of two topics. The first discusses approximation methods for pricing swaptions and the second proposes frameworks for the spot and forward relationship in carbon emissions markets.

In Chapter 2, the validity of two kinds of approximation methods based on moment expansions with multi-factor affine jump-diffusion processes are investigated and their precision are compared numerically. One is based upon the Gram–Charlier expansion in which normal distributions are used to approximate the density of underlying swap values and the associated approximation formula is derived. The formula is given in additive form and coefficients are written with the cumulants of the swap values, in that the gap between the density of swap values and that of normal distributions is adjusted with cumulants of underlying swap values. The other is based upon the generalized Edgeworth expansion in which any distribution is used and we adopt the one zero-coupon bond representing the final cash-flow in swap values as the approximating variable. The associated approximation formula consists of the options on the zero-coupon bond and adjustment terms. This decomposition of swaption prices indicates that swaptions can be partially hedged with options on zero-coupon bonds. The gap between the density of swap values and that of the zero-coupon bond is adjusted with cumulants of underlying swap values and that of the zero-coupon bond. The calculation of swaptions is reduced to that of higher-order moments of products of zero-coupon bonds, which is obtained by solving ordinary differential equations under a multi-factor affine jump-diffusion process.

We implement five numerical examples: the model with Gaussian-type volatility and exponentially, normally or truncated normally distributed jump size,

and that with CIR-type volatility and exponentially or truncated normally distributed jump size. Each price with the moment expansions is compared with that using Monte Carlo simulation. Based on the numerical examples, the higher-order approximation yields more accurate prices on average for both the Gram–Charlier and the generalized Edgeworth expansions in any jump setting. For some special cases, such as low swap rate and low volatility, the generalized Edgeworth expansion outperforms the Gram–Charlier expansion in the sense that it can attain the same level of pricing error but requires a lower degree of moments and less CPU time. The idea that the options on the zero-coupon bonds can be used for hedging swaptions is supported numerically. For the exponentially distributed jump size setting, the ordinary differential equations are solved analytically and moment expansion methods work efficiently, much faster than Monte Carlo methods. For the normally distributed and the truncated normally distributed jump sizes, the ordinary differential equations are solved numerically. Solving the ordinary differential equations numerically works well although it is more time-consuming than that for an exponentially distributed setting. Even so, approximation methods run much faster than Monte Carlo simulation.

The spot and forward relationship in carbon emissions markets is derived in Chapters 3 and 4. Two kinds of framework are proposed. An exogenous forward price process is assumed and its drift term is a function of the net position of the zone, representing the difference between total allocation and emissions. The spot price at maturity is the forward price plus penalty conditioned on the net position of the zone being short and it is an expectation of the cash-flow conditioned on the net position of the zone being long, and it is therefore considered the derivative contract on the forward price and net position of the zone. The net position of the zone is assumed to follow a Markov Chain in Chapter 3 and a linear diffusion in Chapter 4, respectively. Two filtrations used in pricing are implemented in each chapter. One of these is complete information: market participants observe both the forward price process and the net position of the zone at any time. Since the net position of the zone is not a tradable instrument, this is the pricing in an incomplete market. We apply a locally minimization approach to fix the equivalent martingale measure.

138

The no-arbitrage spot price is obtained analytically. The other is incomplete information: market participants observe the forward price process at any time and the net position of the zone only at the announcement time. The filtration is generated to express this situation. The distribution of the net position of the zone restricted to the new filtration is examined. In Chapter 3, the distribution is obtained as an application of Wonham filter. The locally risk-minimizing approach is applied and the distribution under the minimal martingale measure is calculated. In Chapter 4, the distribution is obtained as an application of Kalman-Bucy filter. The distribution under the minimal martingale measure still follows normal. The no-arbitrage price is obtained using Monte Carlo simulation.

Numerical simulations compare the prices under the complete and incomplete information settings. They have similar patterns with respect to the parameters. At announcement time, the jump in the net position of the zone occurs and trends in prices of forward and spot prices vary.

# Bibliography

Baum, L. E. and T. Petrie (1966) Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6), pp. 1554–1563.

Baum, L. E., T. Petrie, G. Soules and N. Weiss (1970) A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, 41(1), pp. 164–171.

Baz, J. and S. R. Das (1996) Analytical Approximations of the Term Structure for Jump–Diffusion Processes: A Numerical Analysis. *Journal of Fixed Income*, 6, pp. 78–86.

Benz, E. A. and S. Trück (2009) Modeling the Price Dynamics of CO2 Emission Allowances. *Energy Economics*, 31(1), pp. 4–15.

Bingham, N. H. and R. Kiesel (2004) *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer-Verlag, New York, 2nd edition.

Capoor, K. and P. Ambrosi (2008) State and Trends of the Carbon Market 2008. *The World Bank, Washington, DC*.

Carmona, R., M. Fehr, J. Hinz and A. Porchet (2009) Market Design for Emission Trading Schemes. *Siam Review*. To appear.

Carr, P. and D. B. Madan (1999) Option Valuation Using the Fast Fourier Transform. *Journal of Computational Finance*, 2(4), pp. 61–73.

Çetin, U., R. Jarrow, P. Protter and Y. Yildirim (2004) Modelling Credit Risk with Partial Information. *The Annals of Applied Probability*, 14(3), pp. 1167–1178.

Çetin, U. and M. Verschuere (2009) Pricing and Hedging in Carbon Emissions Markets. *International Journal of Theoretical and Applied Finance*, 12, pp. 949–967.

Chacko, G. and S. R. Das (2002) Pricing Interest Rate Derivatives: A General Approach. *Review of Financial Studies*, 15, pp. 195–241.

Chesney, M. and L. Taschini (2009) The Endogeneous Price Dynamics of Emission Allowances and an Application to CO2 Option Pricing. Working paper, University of Zurich.

Collin-Dufresne, P. and R. S. Goldstein (2002) Pricing Swaptions within an Affine Framework. *Journal of Derivatives*, 10(1), pp. 1–18.

Das, S. R. (2002) The Surprise Element: Jumps in Interest Rates. *Journal of Econometrics*, 106(1), pp. 27–65.

Das, S. R. and S. Foresi (1996) Exact Solutions for Bond and Option Prices with Systematic Jump Risk. *Review of Derivatives Research*, 1(1), pp. 7–24.

Daskalakis, G., D. Psychoyios and R. N. Markellos (2009) Modeling CO2 Emission Allowance Prices and Derivatives: Evidence from the European Trading Scheme. *Journal of Banking and Finance*, 33(7), pp. 1230–1241.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, pp. 1–38.

Duffie, D., A. Eckner, G. Horel and L. Saita (2009) Frailty Correlated Default. *Journal of Finance*, 64(5), pp. 2089–2123.

Duffie, D. and R. Kan (1996) A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6, pp. 379–406.

Duffie, D., J. Pan and K. J. Singleton (2000) Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(4), pp. 1343–1376.

Durham, J. B. (2006) Additional Analytical Approximations of the Term Structure and Distributional Assumptions for Jump-Diffusion Processes. *Journal of Fixed Income*, 15(4), pp. 61–73.

Ellerman, D. and B. Buchner (2008) Over-Allocation or Abatement ? A Preliminary Analysis of the EU ETS Based on the 2005-06 Emissions Data. *Environmental and Resource Economics*, 47(2), pp. 267–287.

Elliott, R. J. and J. van der Hoek (1997) An Application of Hidden Markov Models to Asset Allocation Problems. *Finance and Stochastics*, 1, pp. 229–238.

Föllmer, H. and A. Schied (2004) *Stochastic Finance, volume 27 of de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin.

Föllmer, H. and M. Schweizer (1991) Hedging of Contingent Claims under Incomplete Information. In M. H. A. Davis and R. J. Elliott, eds., *Applied stochastic analysis*, pp. 389–414. Gordon and Breach, London New York.

Föllmer, H. and M. Sondermann (1986) Hedging of Non-redundant Contingent Claims. In W. Hildenbrand and A. Mas-Colell, eds., *Contributions to Mathematical Economics*, pp. 205–223. North-Holland, Amsterdam.

Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering, volume 53 of Applications of Mathematics*. Springer-Verlag, New York.

Glasserman, P. and K.-K. Kim (2009) Saddlepoint Approximations for Affine Jump-Diffusion Models. *Journal of Economic Dynamics and Control*, 33, pp. 15–36.

Harvey, A. C. (1991) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Heston, S. L. (1993) A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6, pp. 327–343.

Jacod, J. (1975) Multivariate Point Processes: Predictable Projection, Radon-Nikodým Derivatives, Representation of Martingales. *X. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31, pp. 235–253.

Jamshidian, F. (1989) An Exact Bond Option Formula. *Journal of Finance*, 44, pp. 205–209.

Jarrow, R. and A. Rudd (1982) Approximate Option Valuation for Arbitrary Stochastic Processes. *Journal of Financial Economics*, 10(3), pp. 347–369.

Jeanblanc, M. and S. Valchev (2005) Partial Information and Hazard Process. *International Journal Theoretical and Applied Finance*, 8(6), pp. 807–838.

Johannes, M. (2004) The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *Journal of Finance*, 59(1), pp. 227–260.

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

Lakner, P. (1998) Optimal Trading Strategy for an Investor: The Case of Partial Information. *Stochastic Process and their Applications*, 76, pp. 77–97.

Landén, C. (2000) Bond Pricing in a Hidden Markov Model of the Short Rate. *Finance and Stochastics*, 4, pp. 371–389.

Lee, R. W. (2004) Option Pricing by Transform Methods: Extensions, Unification and Error Control. *Journal of Computational Finance*, 7(3), pp. 51–86.

Liptser, R. S. and A. N. Shiryaev (2001) *Statistics of Random Processes: I. General Theory.* Springer-Verlag, 2nd edition.

Munk, C. (1999) Stochastic Duration and Fast Coupon Bond Option Pricing in Multi-factor Models. *The Review of Derivatives Research*, 3, pp. 157–181.

Musiela, M. and M. Rutkowski (2005) *Martingale Methods in Financial Modeling, volume 36 of Applications of Mathematics.* Springer-Verlag, New York, 2nd edition.

Pan, J. (2002) The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics*, 63, pp. 3–50.

Paolella, M. S. and L. Taschini (2006) An Econometric Analysis of Emission Trading Allowances. Working Paper.

Piazzesi, M. (2005) Bond Yields and the Federal Reserve. *Journal of Political Economy*, 113, pp. 311–344.

Protter, P. E. (2004) *Stochastic Integration and Differential Equations.* Springer-Verlag, 2nd edition.

Sass, J. and U. G. Haussmann (2004) Optimizing the Terminal Wealth under Partial Information: The Drift Process as a Continuous Time Markov Chain. *Finance and Stochastics*, 8, pp. 553–577.

Schweizer, M. (2001) A Guided Tour through Quadratic Hedging Approaches. In E. Jouini, J. Cvitanic and M. Musiela, eds., *Option Pricing, Interest Rates and Risk Management*, pp. 538–574. Cambridge University Press, Cambridge.

Scott, L. O. (1997) Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods. *Mathematical Finance*, 7, pp. 413–426.

Seifert, J., M. Uhrig-Homburg and M. W. Wagner (2008) Dynamic Behavior of $CO_2$ Spot Prices. *Journal of Environmental Economics and management*, 56(2), pp. 180–194.

Shumway, R. H. and D. S. Stoffer (2006) *Time Series Analysis and Its Applications*. Springer-Verlag, New York, 2nd edition.

Singleton, K. J. and L. Umantsev (2002) Pricing Coupon-Bond Options and Swaptions in Affine Term Structure Models. *Mathematical Finance*, 12, pp. 427–446.

Stuart, A. and J. K. Ord, eds. (1994) *Kendall's Advanced Theory of Statistics, Volume 1 Distribution Theory*. London: Edward Arnold, 6th edition.

Tanaka, K., T. Yamada and T. Watanabe (2010) Applications of Gram-Charlier Expansion and Bond Moments for Pricing of Interest Rates and Credit Risk. *Quantitative Finance*, 10, pp. 645–662.

Tanner, M. A. (1993) *Tools for Statistical Inference, Springer Series in Statistics*. Springer-Verlag, 2nd edition.

Turnbull, S. M. and L. M. Wakeman (1991) A Quick Algorithm for Pricing European Average Options. *Journal of Financial and Quantitative Analysis*, 26, pp. 377–389.

Wu, C. F. J. (1983) On the Convergence Properties of the EM Algorithm. *Annals of Statistics*, 11, pp. 95–103.