Master of science thesis

Quantifying commodity basis risk by simulating the price dynamics of futures

Robert Sirviö

May 10, 2016

University of Helsinki
Faculty of Science
Department of Mathematics and Statistics

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department	
Faculty of Science	Department of Mathematics and Statistics	
Tekijä — Författare — Author		
Robert Sirviö		
Työn nimi — Arbetets titel — Title		
Quantifying commodity basis risk by simulating futures price dynamics		
Oppiaine — Läroämne — Subject		
Applied Mathematics		
Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	May 2016	100 p.

Tiivistelmä — Referat — Abstract

Measuring risk is mandatory in every form of responsible asset management; be it mitigating losses or maximizing performance, the level of risk dictates the magnitude of the effect of the strategy the asset manager has chosen to execute. Many common risk measures rely on simple statistics computed from historic data. In this thesis, we present a more dynamic risk measure explicitly aimed at the commodity futures market.

The basis of our risk measure is built on a stochastic model of the commodity spot price, namely the Schwartz two-factor model. The model is essentially determined by a system of stochastic differential equations, where the spot price and the convenience yield of the commodity are modelled separately. The spot price is modelled as a Geometric Brownian Motion with a correction factor (the convenience yield) applied to the drift of the process, whereas the convenience yield is modelled as an Ornstein-Uhlenbeck process. Within this framework, we show that the price of a commodity futures contract has a closed form solution. The pricing of futures contracts works as a coupling between the unobservable spot price and the observable futures contract price, rendering model fitting and filtering techniques applicable to our theoretic model.

The parameter fitting of the system parameters of our model is done by utilizing the prediction error decomposition algorithm. The core of the algorithm is actually obtained from a by-product of a filtering algorithm called Kalman filter; the Kalman filter enables the extraction of the likelihood of a single parameter set. By subjecting the likelihood extraction process to numerical optimization, the optimal parameter set is acquired, given that the process converges.

Once we have attained the optimal parameter sets for all of the commodity futures included in the portfolio, we are ready to perform the risk measurement procedure. The first phase of the process is to generate multiple future trajectories of the commodity spot prices and convenience yields. The trajectories are then subjected to the trading algorithm, generating a distribution of the returns for every commodity. Finally, the distributions are aggregated, resulting in a returns distribution on a portfolio level for a given target time frame. We show that the properties of this distribution can be used as an indicator for possible anomalies in the returns within the given time frame.

Avainsanat — Nyckelord — Keywords

Commodity futures, Schwartz two-factor model, Kalman filter

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpula Campus Library

Muita tietoja — Övriga uppgifter — Additional information

Acknowledgements

The main research and implementations for this thesis were conducted at Estlander & Partners during the summer of 2015. I would like to express my gratitude for the unique opportunity I received, and also thank the whole staff at Estlander & Partners for the warm welcome. I would especially like to thank my supervisor and superior Gillis Danielsen for all the guidance and help during the whole writing process, and for all the inspiring discussions we shared.

I would also like to thank my supervisor from Helsinki University, Samuli Siltanen, for all his efforts in guiding and helping me finish this thesis. I would also like to thank Samuli Siltanen for his exciting lectures, and in general for his input on the program of computer assisted mathematics.

I would like to thank my loving parents for their endless support, especially during my studies. This accomplishment would not have been possible without their continuous encouragement.

Finally, thank you Tipi for your support and patience, and for reminding me of the other necessities in life.

Contents

Chapter 1

Introduction

For every financial investment strategy, regardless of the traded financial instruments or the choice of strategy, it is imperative to be able to assess the the level of risk involved in the strategy. A standard procedure for evaluating the level of risk is to analyse the historic volatility of the returns of the trading strategy; the higher the volatility of the returns, the more exposed the investment portfolio is for extreme losses from the strategy.

An example measure is the value at risk. Given the distribution of the returns, the value at risk (at a level α) is defined as the negative α-quantile of the distribution [1], denoted $\mathrm{VaR}_\alpha(\cdot)$. More explicitly, assume a trading strategy over n trading days for which $\mathbf{\Lambda} \in \mathbb{R}^{n\times 1}$ contains the allotted capital for each trading day. Furthermore, assume that $\boldsymbol{R} \in \mathbb{R}^{n\times 1}$ contains the returns for each trading day, and that the distribution of the returns of the trading strategy follow a normal distribution. The standard deviation σ of the distribution is then simply acquired from

$$\sigma = \sqrt{\mathbf{\Lambda}\mathrm{Cov}\left(\boldsymbol{R}\right)\mathbf{\Lambda}^T},$$

and

$$\mathrm{VaR}_\alpha(\boldsymbol{R}) = \sigma\Phi^{-1}\left(\alpha\right). \tag{1.1}$$

In this thesis, we aim our focus on an investment portfolio consisting of 26 different commodity futures[1]. The risk measure described above performs well as a rule of thumb for the portfolio in question, but fails to predict relatively frequently actualized outliers. A few typical observations are tabulated in table 1.1.

[1]See table 5.4 for details on the content of the portfolio.

Date	$\text{VaR}_{5\%}(R)$ for 5 day return	5 day return
01-Jun-1999	-0.78%	-2.35%
17-Mar-1998	-0.58%	-1.98%
13-Nov-2001	-0.59%	-2.27%
11-Mar-2003	-0.76%	-2.28%
14-Oct-2008	-0.47%	-1.57%
27-Jan-2015	-0.76%	-2.74%

Table 1.1: A comparison of a 5% value at risk measure and the actualized returns. The estimated value at risk implies that the actualized 5 day returns are beyond the (practical) support of the distribution, indicating that our choice of distribution for the returns is incorrect.

The inaccurateness of the procedure is, in fact, something to be expected; the return distributions appear to be leptokurtic rather than Gaussian [6].

We propose a different conduct for risk measurement. In essence, our risk measure relies in simulating the dynamics of the commodity futures market rather than simply calculating statistics from historic data. The procedure can be roughly deconstructed into three phases.

I) Model the spot-price dynamics for each commodity and fit the model to historic data.

II) Simulate multiple futures trajectories.

III) Run the trading algorithm and construct the distribution of the returns for the simulated trajectories.

We show that we are able to accurately model the commodity spot price dynamics with a so called Schwartz two factor model through numeric optimization. Furthermore, we show that our procedure can potentially predict the actualization of outlier returns of the trading strategy by analysing the distributions of the simulated returns.

Chapter 2

Preliminaries

This is an introductory chapter containing a collection of useful and fundamental results from stochastic calculus, as well as a short review on terminology and principles of futures and futures trading.

In sections 2.1, 2.2 and 2.3 we present central theorems and results of stochastic calculus, with which we will establish the fundamental basis of this thesis. The primary aim of this section is to make this thesis, and other related publications, comprehensible for readers unacquainted with financial mathematics and pricing of derivatives. The results in this chapter are generally considered as intrinsic tools in stochastic calculus and financial mathematics, and are therefore seldom presented in scientific literature within the respective fields. We will, however, assume that the reader has some prior basic knowledge of stochastic calculus and measure theory. A reader unfamiliar with either one of these concepts is directed to [40] or [25].

We will begin with the definition and properties of Brownian motion.

2.1 Brownian motion

In 1827, botanist Robert Brown experimented with pollen particles in a fluid suspension, and noticed highly irregular movement of the pollen particles. The observations were reported in [5], which is the ultimate origin of the modern concept of Brownian motion. Later, in the beginning of the twentieth century, the first mathematical formulations of the reported irregular movements were built. Albert Einstein proposed in [12] that irregular movement of the pollen particles is caused by bombardment of miniscule liquid molecules, and derived a mathematical formulation for the irregular motion. It wasn't until 1920, almost a hundred years after the first observations of the phenomenon, that a mathematically complete and rigorous model for Brownian motion was built. This was done by Norbert Wiener. Brownian motion has since been used in several applications, including modelling increments in stock prices.

Definition 2.1. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space, where the the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is a natural filtration, namely $\mathcal{F}_t = \sigma(W_s \mid 0 \leq s \leq t)$, of a stochastic process $\{W_t\}_{t\geq 0}$. The process $\{W_t\}_{t\geq 0}$ is a Brownian motion (or a Wiener process) if it meets the following conditions (where $\omega \in \Omega$).*

I) The increments are independent of the past, or more precisely, $\sigma(W_t(\omega) - W_s(\omega)) \perp\!\!\!\perp \{\mathcal{F}_u\}_{0\leq u\leq s}$

II) The increments $W_t(\omega) - W_s(\omega)$ are normally distributed with $\mathbb{E}_{\mathbb{P}}(W_t(\omega) - W_s(\omega)) = 0$ and $\mathrm{Var}_{\mathbb{P}}(W_t(\omega) - W_s(\omega)) = t - s$, that is, $W_t(\omega) - W_s(\omega) \sim N(0, t-s)$

III) The sample paths (or trajectories) $t \mapsto W_t(\omega)$ are continuous.

If $W_0 \equiv 0$ additionally, then $\{W_t\}_{t\geq 0}$ is a standard Brownian motion.

Notational remark 2.2. We will henceforth omit the explicit use of the state of nature (ω), and the domain subscripts of the processes. This is purely for notational convenience. These properties should, however, not be neglected, and are constantly (although implicitly) present in our reasoning.

Note that condition II implies that $W_t \sim N(0, t)$ for a standard Brownian motion.

Condition I corresponds to the observed effects of the collisions between a particle and fluid molecules in that the collisions are seemingly independent from each other.

The normality of the increments (condition II) is a consequence of the central limit theorem. The total disposition of the particle between sequential observations is the sum of multiple collisions between the particle and the fluid molecules. The central limit theorem then states that the limit of such a sum is approximately normal. The mean is assumed to be zero as the particle has no preferred direction, and the variance is assumed to be proportional to the length of the time that the movement has been observed.

Note that we implicitly assume that a process such as Brownian motion exists, and indeed, it is not trivial such processes do exist. The process of proving the existence is, however, a very meticulous one, and beyond the scope of this thesis; therefore we will simply state the existence as a fact. Readers bothered by this ambiguity are directed to [31].

We will next show that Brownian motion possesses two properties that are helpful when sample paths of Brownian motion are simulated, namely space homogeneous property and the (weak) Markov property.

Definition 2.3. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $\{X_t\}_{t\geq 0}$ *be a stochastic process. The process is space homogeneous if*

$$\mathbb{P}\left(X_t \leq x_0 | X_0 = 0\right) = \mathbb{P}\left(X_t \leq x_0 + x | X_0 = x\right) \tag{2.4}$$

for every $t \geq 0$.

The definition above implies that a stochastic process is space homogeneous if the transition probabilities of two distinct states depends only on the difference of the process at these states.

Theorem 2.5. *A Brownian motion is space homogenous.*

Proof. Let W_t be a Brownian motion. Now, since

$$(W_t - W_0 | W_0 = x) \sim N(0,t), \quad \text{for every } x \in \mathbb{R},$$

we see that

$$\begin{aligned}
\mathbb{P}\left(W_t \leq x_0 | W_0 = 0\right) &= \mathbb{P}\left(W_t - 0 \leq x_0 | W_0 = 0\right) = \mathbb{P}\left(W_t - W_0 \leq x_0 | W_0 = 0\right) \\
&= \mathbb{P}\left(W_t - W_0 \leq x_0 | W_0 = x\right) = \mathbb{P}\left(W_t - x \leq x_0 | W_0 = x\right) \\
&= \mathbb{P}\left(W_t \leq x_0 + x | W_0 = x\right)
\end{aligned}$$

□

Definition 2.6. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $\{X_t\}_{t\geq 0}$ *be a stochastic process. If*

$$\mathbb{P}\left(X_{t+x} \leq y \mid \mathcal{F}_t\right) = \mathbb{P}\left(X_{t+x} \leq y \mid X_t\right) \tag{2.7}$$

for any $t, x \geq 0$, *then* X_t *is a Markov process.*

Alternatively, we say that X_t possesses the Markov property.

Theorem 2.8. *Brownian motion is a Markov process.*

Proof. Let W_t be a Brownian motion. We will show that the moment generating functions of $W_{t+x} \mid \mathcal{F}_t$ and $W_{t+x} \mid W_t$ equal for all $t, x \geq 0$, which is equivalent with the statement in 2.6[25]. Recall that the moment generating function of a random variable X is defined as $M_X(t) = \mathbb{E}\left(e^{tX}\right)$. Now

$$\begin{aligned} M_{W_{t+x}|\mathcal{F}_t}(z) &= \mathbb{E}_{\mathbb{P}}\left(e^{zW_{t+x}}|\mathcal{F}_t\right) = e^{z(W_t-W_t)}\mathbb{E}_{\mathbb{P}}\left(e^{zW_{t+x}}|\mathcal{F}_t\right) \\ &= e^{zW_t}\mathbb{E}_{\mathbb{P}}\left(e^{z(W_{t+x}-W_t)}|\mathcal{F}_t\right) \overset{\text{(I)}}{=} e^{zW_t}\mathbb{E}_{\mathbb{P}}\left(e^{z(W_{t+x}-W_t)}\right) \\ &= e^{zW_t}\mathbb{E}_{\mathbb{P}}\left(e^{z(W_{t+x}-W_t)}|W_t\right) = \mathbb{E}_{\mathbb{P}}\left(e^{zW_{t+x}}|W_t\right) \\ &= M_{W_{t+x}|W_t}(z) \end{aligned}$$

for all $t, x \geq 0$. □

Combining properties (2.5) and (2.8) it is now easy to simulate sample paths of Brownian motions. Consider a time interval $[0, T]$ as the domain in which the sample paths are simulated and define a partition

$$P_n = \{t_i \in [0,T] \mid i \in \{0, \dots, n\},\ t_0 = 0 < t_1 < \dots t_{n-1} < t_n = T\}$$

in which the state of the Brownian motion is evaluated. Now, clearly

$$\begin{aligned} W_T - W_0 &= \left(W_T - W_{t_{n-1}}\right) + \left(W_{t_{n-1}} - W_{t_{n-1}}\right) + \dots + \left(W_{t_1} - W_0\right) \\ &= \Delta W_T + \Delta W_{t_{n-i}} + \dots + \Delta W_{t_1}, \\ \Delta x_i &= x_i - x_{i-1} \\ \Delta W_{t_i} &\sim N(0, \Delta t_i). \end{aligned}$$

Without any loss of generality, we will assume that $\Delta t_i \equiv \Delta t$ for all $t_i \in P_n$. Assume a normally distributed random variable $Z_{\Delta t}$ such that $Z_{\Delta t} \sim N(0, \Delta t)$. Sample paths can now be generated from the sum of increments of Brownian motion with the following difference equation

$$W_{t_i} = \begin{cases} W_{t_{i-1}} + Z_{\Delta t}, & i \in \{1, \dots n\} \\ W_0, & i = 0. \end{cases}$$

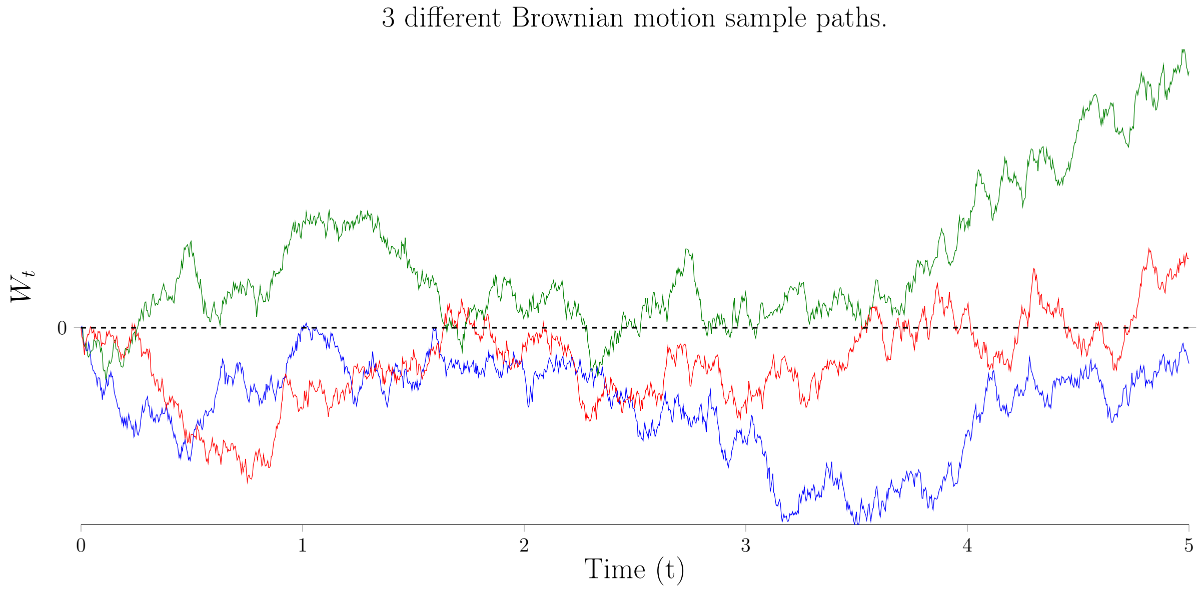


Figure 2.1: Sample paths for standard Brownian motion with $\Delta t = 1/260$ within the time domain $[0, 5]$

Figure 2.2 illustrates the, perhaps intuitive, effect of the increase in variance of a Brownian motion W_t relative to the initial deterministic boundary condition W_0; the longer we observe a Brownian motion, the more likely it has strayed further away from the initial state.

Probability distributions for a standard Brownian motion.

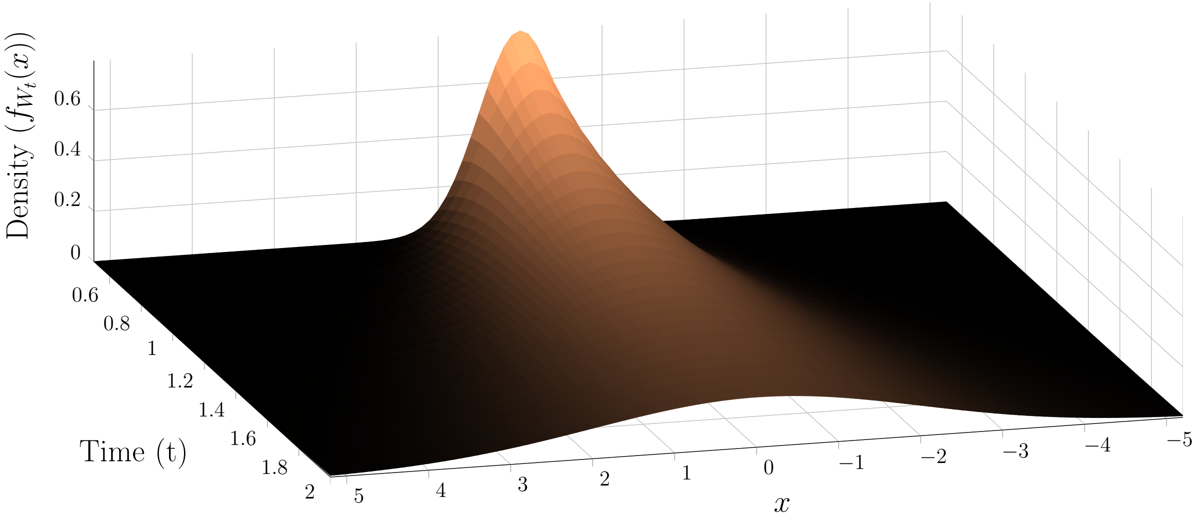


Figure 2.2: Probability densities ($f_{W_t}(x)$) for standard Brownian motion W_t where $t \in \left[\frac{1}{2}, 2\right]$. Note that $f_{W_t}(x) = \Phi'(x/\sqrt{t})$ as $W_t \sim N(0,t)$ for all $t \in \mathbb{R}_+$.

Recall the definition of a martingale.

Definition 2.9. *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *be a filtered probability space, where the the filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ *is a natural filtration of a stochastic process* $\{X_t\}_{t\geq 0}$. *If*

$$\mathbb{E}_{\mathbb{P}}\left(|X_t|\right) < \infty, \quad \textit{for every } t \geq 0 \tag{2.10}$$

$$\mathbb{E}_{\mathbb{P}}\left(X_{t+s}|\mathcal{F}_t\right) = X_t, \quad \textit{for every } t, s \geq 0, \tag{2.11}$$

then $\{X_t\}_{t\geq 0}$ *is a martingale.*

Theorem 2.12. *Brownian motion* W_t *is a martingale.*

Proof. As W_t is normally distributed, we know that W_t is integrable for all $t \geq 0$. Therefore it is sufficient to prove the martingale property of W_t, which is achieved with simple calculations and utilizing the properties of the (conditional) expectation.

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left(W_{t+s}|\mathcal{F}_t\right) &= \mathbb{E}_{\mathbb{P}}\left(W_t + (W_{t+s} - W_t)|\mathcal{F}_t\right) = \mathbb{E}_{\mathbb{P}}\left(W_t|\mathcal{F}_t\right) + \mathbb{E}_{\mathbb{P}}\left(W_{t+s} - W_t|\mathcal{F}_t\right)\\ &= W_t + \mathbb{E}_{\mathbb{P}}\left(W_{t+s} - W_t\right) = W_t,\end{aligned}$$

where the last identitt follows from the adaptedness of W_t, which implies that $\mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_t) = W_t$, and the independence condition I of Brownian motions for $\mathbb{E}_{\mathbb{P}}\left(W_{t+s} - W_t|\mathcal{F}_t\right)$. $\square$

A highly significant property of Brownian motion is that the sample paths $t \mapsto W_t$ are nowhere differentiable. This can be seen by constructing the difference quotient

$$D_{W_t} := \frac{W_{t+\Delta t} - W_t}{\Delta t} = \frac{1}{\Delta t}\left(W_{t+\Delta t} - W_t\right) \sim N\left(0, \frac{1}{\Delta t}\right),$$

as $W_{t+\Delta t} - W_t \sim N\left(0, \Delta t\right)$. By taking the usual limit $\Delta t \to 0$, and by allowing some informal notation, we see that

$$\left[\lim_{\Delta t \to 0} D_{W_t}\right] \sim N(0, \infty).$$

The expression above is, of course, not well-defined, but serves well as an illustration of the fact that by taking the limit of the quotient, we never end up with a construction with bounded variation.

More precisely, let $M > 0$ be arbitrary and $Z \sim N(0,1)$. Now

$$\begin{aligned}\mathbb{P}\left(|D_{W_t}| > M\right) &= \mathbb{P}\left(\left|\frac{Z}{\sqrt{\Delta t}}\right| > M\right) = \mathbb{P}\left(|Z| > \sqrt{\Delta t} M\right) \underset{\Delta t \to 0}{\longrightarrow} \mathbb{P}\left(|Z| > 0\right) \\ &= 1,\end{aligned}$$

which again implies unbounded rate of change and the fact that W_t is not differentiable for any $t \in R_+$.

This has given rise to an alternative form of differential calculus for Brownian motion, namely Itô calculus, which will be briefly covered in section 2.2.

2.2 Itô calculus

As already mentioned, Brownian motion is incorporated in many real-world applications, especially within the field of mathematical modelling. The typical role of Brownian motion in these models is to introduce a probabilistic property to the system, or in other words, transforming a deterministic model into a probabilistic one. For example, in most stock market modelling frameworks, if one were to remove Brownian motion from the model, the model would be with deterministic boundary conditions completely deterministic.

These models are usually formulated based on assumptions on the transitional properties of the modelled phenomena. For example, stochastic models of financial derivatives are, almost exclusively, based on assumptions on the price movement on the underlying assets. Hence, it would certainly seem tempting to characterize the assumptions through differentials. This is, as demonstrated in the previous chapter, problematic as Brownian motion is nowhere differentiable. Perhaps a bit surprisingly, the troublesome property of Brownian motion has not, however, been a cause great enough to abandon these models. Instead of reformulating the models, new forms of differential calculus have been developed to tackle the non-differentiability of Brownian motion. The most used and renowned theory was developed by Kiyoshi Itô, a Japanese mathematician, in the 1950's, and is nowadays known as Itô calculus.

2.2.1 The Itô integral

The Itô integral is constructed as a transition from a discrete setting through limits of simple functions, similar to for example the Riemann integral. We will begin by defining the discrete stochastic integral, for which we will need the definition of simple functions.

Definition 2.13. *Let $P^n_{[a,b]}$ be a partition of the interval $[a,b]$ such that*

$$P^n_{[a,b]} = \{t_i \in [a,b] \mid i = \{0,\ldots,n\},\ t_0 = a < t_1 < \ldots < t_{n-1} < t_n = b\}\,. \tag{2.14}$$

A function $f : [a,b] \to \mathbb{R}$ is called simple if

$$f = \sum_{k=0}^{n-1} g(t_k) 1_{[t_k,t_{k+1}[} \tag{2.15}$$

where $t_k \in P^n_{[a,b]}$ for all $k \in \{0,\ldots,n\}$, and for some $g : [a,b] \to \mathbb{R}$.

A central, and perhaps elementary, result in measure theory is that every measurable function can be approximated with simple functions. A less known, and stronger, result, stating that a given general probabilistic function can be approximated by random simple functions, holds.

Theorem 2.16. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $f : [a,b]\times\Omega \to \mathbb{R}$ and $f_n : [a,b]\times\Omega \to \mathbb{R}$ such that*

$$f_n = \sum_{k=0}^{n-1} f(t_k,\omega) 1_{[t_k,t_{k+1})}$$

where $t_k \in P^n_{[a,b]}$ *for all* $k \in \{0, \dots, n\}$ *and* $\omega \in \Omega$. *Then*

$$\lim_{n\to\infty} f_n = f$$

holds.

Proof. See for example [25]. $\square$

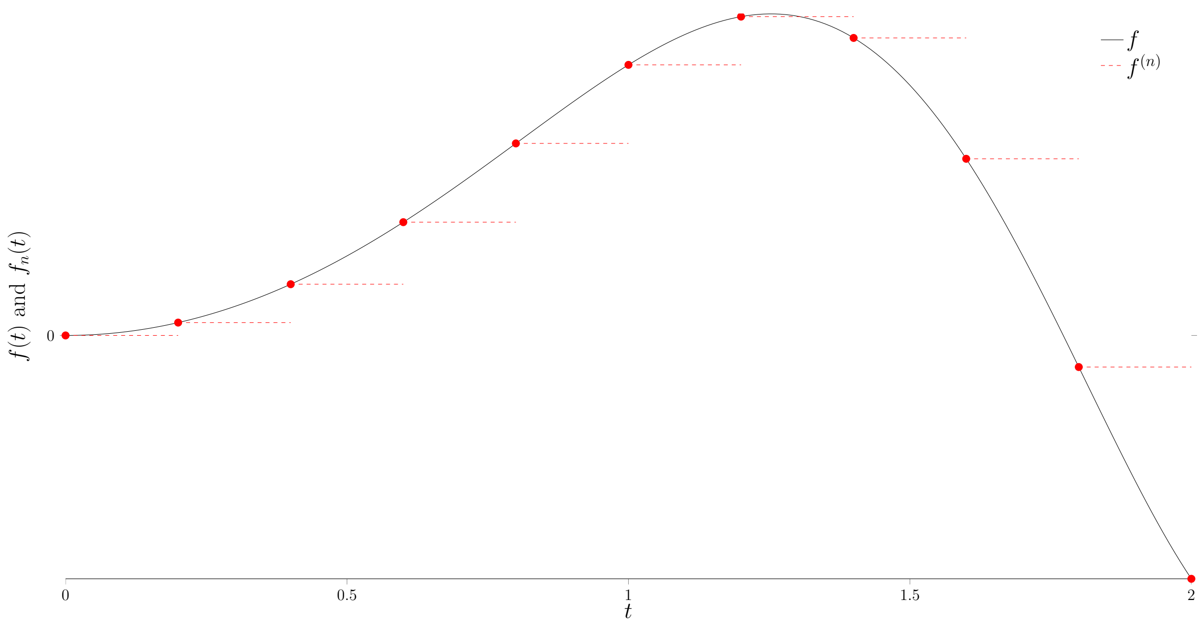


Figure 2.3: An example of the simple function approximation of a deterministic function f over the time domain $[0, 2]$ with $\Delta t = 0.2$ and $n = 10$

Next we will define the discrete Itô integral. The discrete Itô integral can be seen as a stochastic resemblance of the Riemann sums, both in the sense of the similar expression and the fact that they both establish the foundation of the respective integrals.

Definition 2.17. *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *be a filtered probability space and* $f(t) \equiv f(t, \omega)$ *such that* $f : [0, T] \times \Omega \to \mathbb{R}$ *is adapted to the filtration* $\mathcal{F}_t$. *Assume additionally that* W_t *is a standard Brownian motion. Then the discrete Itô integral* $I_{P^n_{[0,T]}}(f)$, *over the partition* $P^n_{[0,T]}$, *is defined as*

$$I_{P^n_{[0,T]}}(f) = \sum_{k=0}^{n-1} f(t_k)\Delta W_{t_{k+1}}. \tag{2.18}$$

Above we defined the discrete Itô integral for probabilistic functions, but the definition is well defined also for deterministic functions. In fact, if f should be deterministic, then $I_{P^n_{[0,T]}}(f)$ is normally distributed, as $W_{t_{k+1}} - W_{t_k} \sim N(0, \Delta t_{k+1})$ and non-overlapping increments are independent. Since

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[I_{P^n_{[0,T]}}(f)\right] &\overset{\perp\!\!\!\perp}{=} \sum_{k=0}^{n-1} f(t_k)\mathbb{E}_{\mathbb{P}}\left[\Delta W_{t_{k+1}}\right] = 0 \\
\mathrm{Var}_{\mathbb{P}}\left[I_{P^n_{[0,T]}}(f)\right] &\overset{\perp\!\!\!\perp}{=} \sum_{k=0}^{n-1} f(t_k)^2\mathrm{Var}_{\mathbb{P}}\left[\Delta W_{t_{k+1}}\right] = \sum_{k=0}^{n-1} f(t_k)^2 \Delta t_{k+1},
\end{aligned} \tag{2.19}$$

then

$$I^n_{P_{[0,T]}}(f) \sim N\left(0, \sum_{k=0}^{n-1} f(t_k)^2 \Delta t_{k+1}\right).$$

This is not the case when f is probabilistic as we shall later see.

Next we will give the definition of the Itô integral. The assumptions, and the rationale, of the definition will be discussed afterwards.

Definition 2.20. *Assume a setting similar in the definition 2.17. Furthermore, if*

$$\int_0^T \mathbb{E}_{\mathbb{P}}\left[f(t,\omega)^2\right] dt < \infty, \tag{2.21}$$

then the Itô integral $\int_0^T f(t,\omega)dW_t$ is defined as

$$\int_0^T f(t,\omega)dW_t = \lim_{n\to\infty} I_{P^n_{[0,T]}}(f_n), \tag{2.22}$$

where f_n is a simple approximation of f in the sense that

$$\lim_{n\to\infty} \int_0^T \mathbb{E}_{\mathbb{P}}\left[(f(t,\omega) - f_n(t,\omega))^2\right] = 0. \tag{2.23}$$

The important question that requires an answer is whether the conditions (2.21) and (2.23) guarantee the existence of the limit in (2.22). We shall consider the convergence in the mean square sense, which is equivalent with the usual sense of convergence [39], where the condition of convergence is

$$\lim_{n\to\infty} \left\| I_{P^n_{[0,T]}}(f_n) - I_{P^n_{[0,T]}}(f_{2n}) \right\| = 0.$$

We immediately notice that

$$\begin{aligned}
\left\| I_{P^n_{[0,T]}}(f_n) - I_{P^n_{[0,T]}}(f_{2n}) \right\| &= \left\| \sum_{k=0}^{n-1} f_n(t_k,\omega)\Delta W_{t_{k+1}} - \sum_{k=0}^{n-1} f_{2n}(t_k,\omega)\Delta W_{t_{k+1}} \right\| \\
&= \left\| \sum_{k=0}^{n-1} (f_n - f_{2n})\,(t_k,\omega)\Delta W_{t_{k+1}} \right\| = \left\| I_{P^n_{[0,T]}}(f_n - f_{2n}) \right\|
\end{aligned}$$

Let us take a closer look at the squared L^2 norm of the discrete Itô integral. By recalling the tower property of the expectation, we see that

$$\begin{aligned}
&\left\| I_{P^n_{[0,T]}}(f_n) \right\|_2^2 \\
&= \mathbb{E}_{\mathbb{P}}\left[I_{P^n_{[0,T]}}(f_n)^2 \right] = \mathbb{E}_{\mathbb{P}}\left[\left(\sum_{k=0}^{n-1} f_n(t_k,\omega)\Delta W_{t_{k+1}} \right)^2 \right] \\
&= \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \Delta W_{t_{k+1}}^2 \right] \\
&\quad + 2\sum_{k<m}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega) f_n(t_m,\omega)\Delta W_{t_{k+1}}\Delta W_{t_{m+1}} \right] \\
&= \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \Delta W_{t_{k+1}}^2 \right] \\
&\quad + 2\sum_{k<m}^{n-1} \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega) f_n(t_m,\omega)\Delta W_{t_{k+1}}\Delta W_{t_{m+1}} \middle|\, \mathcal{F}_m \right] \right] \\
&\overset{\perp\!\!\!\perp}{=} \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \Delta W_{t_{k+1}}^2 \right] \\
&\quad + 2\sum_{k<m}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega) f_n(t_m,\omega)\Delta W_{t_{k+1}} \underbrace{\mathbb{E}_{\mathbb{P}}\left[\Delta W_{t_{m+1}} \middle|\, \mathcal{F}_m \right]}_{=0} \right] \\
&= \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \Delta W_{t_{k+1}}^2 \middle|\, \mathcal{F}_k \right] \right] \\
&\overset{\perp\!\!\!\perp}{=} \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \mathbb{E}_{\mathbb{P}}\left[\Delta W_{t_{k+1}}^2 \middle|\, \mathcal{F}_k \right] \right] = \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \right] \Delta t_{k+1},
\end{aligned}$$

that is,

$$\left\| I_{P^n_{[0,T]}}(f_n) \right\|_2^2 = \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}} \left[f_n(t_k, \omega)^2 \right] \Delta t_{k+1}. \tag{2.24}$$

Now, consider the following definition of a squared norm on f_n

$$\|f_n\|^2 = \int_0^T \mathbb{E}_{\mathbb{P}} \left(f_n^2 \right) dt. \tag{2.25}$$

The fact that (2.25) actually defines a norm is not trivial. This can, however, be quite easily verified[39]. The norm in (2.25) can be interpreted as an average of the L^2 norm. By expanding the definition, we see that

$$\begin{aligned} \|f_n\|^2 &= \int_0^T \mathbb{E}_{\mathbb{P}} \left(f_n^2 \right) dt = \int_0^T \sum_{k=0}^{n-1} \underbrace{\mathbb{E}_{\mathbb{P}} \left[f_n(t_k, \omega)^2 1_{[t_k, t_{k+1}[} \right]}_{=0, \text{ when } t \notin [t_k, t_{k+1}[} dt \\ &= \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E}_{\mathbb{P}} \left[f_n(t_k, \omega)^2 \right] dt = \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}} \left[f_n(t_k, \omega)^2 \right] \Delta t_{k+1} \end{aligned} \tag{2.26}$$

By combining (2.24) and (2.26), we acquire an important result, namely

$$\left\| I_{P^n_{[0,T]}}(f_n) \right\|_2 = \|f_n\|. \tag{2.27}$$

In fact, this is all we need to conclude that the assumptions in 2.20 guarantee the existence of the limit; by utilizing (2.27), we claim that

$$\lim_{n \to \infty} \left\| I_{P^n_{[0,T]}}(f_n - f_{2n}) \right\|_2 = \lim_{n \to \infty} \|f_n - f_{2n}\| = 0$$

holds, since theorem 2.16 guarantees that there exists a simple function f_n that approximates f. Of course, since it should approximate f in the sense of the norm $\|\cdot\|$ defined in (2.25), we require that $\|f\|$ exists. However,

$$\|f\|^2 = \int_0^T \mathbb{E}_{\mathbb{P}} \left[f(t, \omega)^2 \right] dt < \infty$$

is assumed in (2.21), which implies that

$$\|f\| < \infty$$

and thus the norm exists.

Remark 2.28. If a function f satisfies the conditions stated in the definition 2.20 of the Itô integral, we say that f is *Itô integrable*.

The expression in (2.24) is actually important in an additional sense, as it practically proves another important result, namely the *Itô isometry*.

Theorem 2.29. *If f is Itô integrable, then*

$$\mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t,\omega)dW_t\right)^2\right] = \mathbb{E}_{\mathbb{P}}\left[\int_0^T f(t,\omega)^2 dt\right]. \tag{2.30}$$

Proof. The theorem is proved with the help of simple functions.

From (2.24) we know that

$$\mathbb{E}_{\mathbb{P}}\left[I_{P^n_{[0,T]}}(f_n)^2\right] = \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2\right]\Delta t_{k+1} = \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[f_n(t_k,\omega)^2 \Delta t_{k+1}\right]$$

By taking the limit of $n \to \infty$ (that is, shrinking the partition), we arrive at the result stated in the theorem. □

With the Itô isometry, we are able to prove a significant property of the Itô integrals where the integrand is deterministic. This is of great use in the upcoming chapters when securities are evaluated.

Theorem 2.31. *Let f be a Itô integrable function. If f is also deterministic, then*

$$\int_0^T f(t)dW_t \sim N\left(0, \int_0^T f(t)^2 dt\right) \tag{2.32}$$

Proof. We will again prove the claim through simple functions.

From (2.19) we know that

$$\mathbb{E}_{\mathbb{P}}\left[I_{P^n_{[0,T]}}(f_n)\right] = \sum_{k=0}^{n-1} f(t_k) \underbrace{\mathbb{E}_{\mathbb{P}}\left[\Delta W_{t_{k+1}}\right]}_{\substack{=0 \text{ for all } t_k \in P^n_{[0,T]} \\ \text{for all } n\in\mathbb{N}}} \xrightarrow[n\to\infty]{} 0, \tag{2.33}$$

and thus

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T f(t)dW_t\right] = 0. \tag{2.34}$$

For the variance, we can use the Itô isometry.

$$\begin{aligned}\mathrm{Var}_{\mathbb{P}}\left[\int_0^T f(t)dW_t\right] &= \mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t)dW_t\right)^2\right] - \left(\underbrace{\mathbb{E}_{\mathbb{P}}\left[\int_0^T f(t)dW_t\right]}_{=0}\right)^2 \\ &= \mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t)dW_t\right)^2\right] = \int_0^T \mathbb{E}_{\mathbb{P}}\left[f(t)^2\right] dt = \int_0^T f(t)^2 dt.\end{aligned} \tag{2.35}$$

As $I_{P^n_{[0,T]}}(f_n)$ can be seen as a sequence of normally distributed random variables, one would be inclined to take the limit and simply claim the normality of the limit as a triviality. The claim holds, but is in no way trivial. The claim can be proven e.g. through characteristic functions of moment-generating functions [39].

Thus we conclude that

$$\int_0^T f(t)dW_t \sim N\left(0, \int_0^T f(t)^2 dt\right)$$

□

Notational remark 2.36. Due to the fact that models are usually formulated through transitional properties, a simplifying notation, similar to the one in ordinary differential calculus, has been accepted as a standard within fields that frequently deal with stochastic calculus. Namely, a expression

$$F(T) = \int_0^T f(t)dW_t$$

is usually written as

$$dF(T) = f(T)dW_T,$$

by informally differentiating the integral. We will adapt to this notation in this thesis.

2.2.2 Itôs lemma and stochastic differential equations

The calculations on most of the results in the previous section have relied solely on the definition of the Itô integral. In order to actually apply the theory of stochastic integration, one might hope for more sophisticated tools to work with, as having simple functions as a starting point might get quite tedious in the long run, especially in more complex settings. In this section we will present a very powerful tool for evaluating Itô integrals, namely, Itô's lemma. We will first present a special case of Itô's lemma with proofs, and then present the general case. We will also present some examples and applications of the presented results.

Theorem 2.37. *(Itô's lemma, special case)*

Let f be Itô integrable. Furthermore, if $f \in C^2(\mathbb{R})$, then, for any $t \in [0,T]$,

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt, \tag{2.38}$$

or more explicitly,

$$f(W_t) = f(0) + \int_0^t f'(W_s)dW_s + \frac{1}{2}\int_0^t f''(W_s)ds, \tag{2.39}$$

Proof. We will begin by formulating a useful expression for $f(W_t)$.

$$\begin{aligned} \sum_{k=0}^{n-1} f(W_{t_{k+1}}) - f(W_{t_k}) &= f(W_t) - f(0) \\ \Leftrightarrow \quad f(W_t) &= f(0) + \sum_{k=0}^{n-1} f(W_{t_{k+1}}) - f(W_{t_k}) \end{aligned} \tag{2.40}$$

where $t_k \in P^n_{[0,t]}$ and $t \in [0,T]$.

The trick in the proof of the lemma is to use the Taylor's expansion to the above expression (note that we assume that $f \in C^2(\mathbb{R})$). Utilizing Taylor's expansion on (2.40), we assert that

$$\begin{aligned} f(W_{t_{k+1}}) &= f(W_{t_k}) + f'(W_{t_k})\Delta W_{t_{k+1}} + \frac{1}{2}f''(W_{t_k})\left(\Delta W_{t_{k+1}}\right)^2 \\ \Leftrightarrow \quad f(W_{t_{k+1}}) - f(W_{t_k}) &= f'(W_{t_k})\Delta W_{t_{k+1}} + \frac{1}{2}f''(W_{t_k})\left(\Delta W_{t_{k+1}}\right)^2 \end{aligned} \tag{2.41}$$

Inserting (2.41) into (2.40), we see that

$$\begin{aligned} f(W_t) = f(0) & \\ &+ \sum_{k=0}^{n-1} f'(W_{t_k})\Delta W_{t_{k+1}} && (2.42) \\ &+ \frac{1}{2}\sum_{k=0}^{n-1} f''(W_{t_k})\left(\Delta W_{t_{k+1}}\right)^2. && (2.43) \end{aligned}$$

Straight from the definition, we acquire

$$\lim_{n\to\infty}\sum_{k=0}^{n-1} f'(W_{t_k})\left(\Delta W_{t_{k+1}}\right) = \int_0^t f'(W_s)dW_s \tag{2.44}$$

The second sum (2.43) requires a bit more work. First off, consider the Riemann-approach in determining the second integral in (2.39).

$$\lim_{n\to\infty}\sum_{k=0}^{n-1} f''(W_{t_k})\Delta t_{k+1} = \int_0^t f''(W_s)ds \tag{2.45}$$

Define

$$\delta_n = \max\left\{\Delta t_k \mid t_k \in P^n_{[0,t]}\right\}, \tag{2.46}$$

from which $\delta_n \underset{n\to\infty}{\longrightarrow} 0$ follows.

We will again consider the limit in the square mean sense, and show, that the Riemann sums converge to the same limit as (2.43). Now, by first recognizing that[1]

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right] &= \mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{k+1}}\right)^2\right] - \Delta t_{k+1} = \Delta t_{k+1} - \Delta t_{k+1} = 0\\
\mathbb{E}_{\mathbb{P}}\left[2\left(W_{t_{k+1}}\right)^2 \Delta t_{k+1}\right] &= 2\Delta t_{k+1}\mathbb{E}_{\mathbb{P}}\left[\left(W_{t_{k+1}}\right)^2\right] = 2\left(\Delta t_{k+1}\right)^2\\
\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{k+1}}\right)^4\right] &= 3\left(\Delta t_{k+1}\right)^2
\end{aligned}$$

we see that

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left[\left(\sum_{k=0}^{n-1} f''(W_{t_k})\left(\Delta W_{t_{k+1}}\right)^2 - \sum_{k=0}^{n-1} f''(W_{t_k})\Delta t_{k+1}\right)^2\right] \\
&\quad = \mathbb{E}_{\mathbb{P}}\left[\left(\sum_{k=0}^{n-1} f''(W_{t_k})\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)\right)^2\right] \\
&\quad = \mathbb{E}_{\mathbb{P}}\left[\sum_{k=0}^{n-1} f''(W_{t_k})^2\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)^2\right] + \\
&\qquad\quad 2\mathbb{E}_{\mathbb{P}}\left[\sum_{k<m} f''(W_{t_k})f''(W_{t_m})\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)\left(\left(\Delta W_{t_{m+1}}\right)^2 - \Delta t_{m+1}\right)\right] \\
&\quad \overset{\perp\!\!\!\perp}{=} \mathbb{E}_{\mathbb{P}}\left[\sum_{k=0}^{n-1} f''(W_{t_k})^2\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)^2\right] + \\
&\qquad\quad 2\sum_{k<m} f''(W_{t_k})f''(W_{t_m})\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{m+1}}\right)^2 - \Delta t_{m+1} \middle| \mathcal{F}_{t_{m+1}}\right] \\
&\quad \overset{\perp\!\!\!\perp}{=} \sum_{k=0}^{n-1} f''(W_{t_k})^2\mathbb{E}_{\mathbb{P}}\left[\left(\left(\Delta W_{t_{k+1}}\right)^2 - \Delta t_{k+1}\right)^2 \;\middle|\; \mathcal{F}_{t_k}\right] \\
&\quad = \sum_{k=0}^{n-1} f''(W_{t_k})^2\left(\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{k+1}}\right)^4 \;\middle|\; \mathcal{F}_{t_k}\right] - \mathbb{E}_{\mathbb{P}}\left[2\left(W_{t_{k+1}}\right)^2\Delta t_{k+1} \;\middle|\; \mathcal{F}_{t_k}\right] + \left(\Delta t_{k+1}\right)^2\right) \\
&\quad = 2\sum_{k=0}^{n-1} f''(W_{t_k})^2\left(\Delta t_{k+1}\right)^2 \leq 2\delta_n \sum_{k=0}^{n-1} f''(W_{t_k})^2\Delta t_{k+1} \underset{n\to\infty}{\longrightarrow} 0
\end{aligned}
$$

and thus (in L^2)

$$
\lim_{n\to\infty}\sum_{k=0}^{n-1} f''(W_{t_k})\left(\Delta W_{t_{k+1}}\right)^2 = \lim_{n\to\infty}\sum_{k=0}^{n-1} f''(W_{t_k})\Delta t_{k+1} = \int_0^t f''(W_s)ds. \tag{2.47}
$$

Inserting (2.44) into (2.42) and (2.47) into (2.43) yields the claim in the theorem. □

[1]The equality in the last equation stems from the formula of central moments of normally distributed random variables. More precisely, if $X \sim N(\mu, \sigma^2)$, then

$$
\mathbb{E}_{\mathbb{P}}\left[(X-\mu)^k\right] = \sigma^k (k-1)!!
$$

for all $k \in \{m \in \mathbb{N} \mid \mod(m,2) = 0\}$[41].

Next we will show an example application of the above theorem. The example is quite a classic one, and is showcased in many stochastic calculus textbooks.

Example 2.48. The objective is to evaluate the following integral

$$\int_0^T W_t dW_t, \tag{2.49}$$

where W_t is, as usual, a standard Brownian motion.

Based on ordinary calculus, we guess that the solution is of quadratic form. We begin by applying the theorem 2.37 on the function $f(x) = x^2$ (for which $f \in C^\infty(\mathbb{R})$), and we assert that

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt = 2W_t dW_t + \left(\frac{1}{2} \cdot 2\right) dt = 2W_t dW_t + dt,$$

or equivalently, by noting that $f(W_0) = f(0) = 0$,

$$\begin{aligned} W_T^2 &= 2\int_0^T W_t dW_t + \int_0^T dt = 2\int_0^T W_t dW_t + T \\ \Leftrightarrow \int_0^T W_t dW_t &= \frac{1}{2}\left(W_T^2 + T\right) \end{aligned} \tag{2.50}$$

This example will also be of good use later in this thesis.

Prior to introducing the general form of Itô's lemma, familiarity with the concept of stochastic differential equations is imperative. Stochastic differential equations are a concrete way to formalize the transitional properties of a stochastic process. This is the fundamental starting point also in our attempt to model the commodity futures price dynamics.

Definition 2.51. *A stochastic differential equation, that describes the dynamics of a stochastic process X_t, is of the form*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{2.52}$$

where

$\mu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$

$\sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$

W_t is a standard Brownian motion.

The expression (2.52) is not well-defined as is, but again a shorthand notation of (over a predefined time period $[0, T]$)

$$X_T = X_0 + \int_0^T \mu(t, X_t)dt + \int_0^T \sigma(t, X_t)dW_t \tag{2.53}$$

where the first integral is an ordinary Riemann integral and the second integral is a stochastic Itô integral.

The function μ is usually referred as the *drift coefficient.* The drift coefficient can be interpreted as a long-term mean of the change in the stochastic process.

The σ function is called the *diffusion coefficient* of the process. It can be interpreted as a term that calibrates the "randomness" of the process; the bigger $\|\sigma(t, X_t)\|$, the more the process is likely to fluctuate over time. Especially, note that if $\sigma(t, X_t) \equiv 0$, the process becomes deterministic and fully defined by an ordinary differential equation.

Theorem 2.54. *(Itô's lemma)*

Assume a stochastic process X_t such that its' dynamics are described by

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t.$$

Furthermore, assume that f is a function of the form $f : [0, T] \times \mathbb{R} \to \mathbb{R}$. If $f \in C^{1,2}([0, T] \times \mathbb{R})$, then

$$df = \left(f_t + \mu f_{X_t} + \frac{1}{2}\sigma^2 f_{X_t X_t} \right) dt + \sigma f_{X_t} dW_t \tag{2.55}$$

where $f_x := \frac{\partial f}{\partial x}$, and df is the usual shorthand notation similar to (2.38).

Proof. The proof relies on the same methodology as the proof of the special case of Itô's lemma and can be found in e.g. [27] □

Example 2.56. (The special case of Itô's lemma)

Note that we acquire the special case of Itô's lemma with the following procedure. Let X_t be a stochastic process such that

$$dX_t = dW_t,$$

that is, a process where $\mu(t, X_t) \equiv 0$ and $\sigma(t, X_t) \equiv 1$. Now, by applying Itô's lemma on any function that satisfies the conditions in 2.54 we see that

$$df = f_{X_t} dWt + \frac{1}{2} f_{X_t X_t} dt,$$

or

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt.$$

2.3 Variations

Another important tool and concept distinctively used in stochastic calculus is quadratic variation and covariation. Quadratic variation and covariation are especially useful when evaluating stochastic integrals and handling correlated processes. In this section we will present results of quadratic variation and covariation regarding Brownian motions and Itô integrals.

Definition 2.57. *The covariation of processes X_t and Y_t over $[0,T]$ is*

$$\langle X, Y\rangle_t = \lim_{n\to\infty} \sum_{k=0}^{n-1} \Delta X_{t_{k+1}} \Delta Y_{t_{k+1}} \tag{2.58}$$

The quadratic variation of a process X_t is the covariation of X_t with itself,

$$\langle X, X\rangle_t = \lim_{n\to\infty} \sum_{k=0}^{n-1} \left(\Delta X_{t_{k+1}}\right)^2. \tag{2.59}$$

In both cases $t_k \in P^n_{[0,T]}$ for all k.

By simply expanding the expressions below, we see that the following identities hold.

$$\begin{aligned} \langle X, Y\rangle &= \langle Y, X\rangle && \text{(symmetry)} \\ \langle aX + bY, Z\rangle &= a\langle X, Z\rangle + b\langle Y, Z\rangle && \text{(bilinearity)} \end{aligned}$$

In a general setting, nothing guarantees the existence of (2.58) nor (2.58). However, in the case of Brownian motions, we see that variations always exist.

Theorem 2.60. *Let W_t be a Brownian motion. Then*

$$\langle W, W\rangle_t = t \tag{2.61}$$

Proof. Define

$$S_n = \mathbb{E}_{\mathbb{P}}\left[\sum_{k=0}^{n-1} \left(\Delta W_{t_{k+1}}\right)^2\right]. \tag{2.62}$$

Begin by noticing that

$$\mathbb{E}_{\mathbb{P}}\left(S_n\right)=\sum_{k=0}^{n-1}\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_{k+1}}\right)^2\right]=\sum_{k=0}^{n-1}\Delta t_{k+1}=t \tag{2.63}$$

Thus

$$\mathbb{E}_{\mathbb{P}}\left[\left(S_n-t\right)^2\right]=\mathrm{Var}_{\mathbb{P}}\left(S_n\right). \tag{2.64}$$

The equation (2.64) shows that proving the proposition is equivalent (in the L^2 sense) to showing that the variance of S_n vanishes as n tends towards infinity.

Let $Z\sim N(0,1)$. Now

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left(Z^4\right)&=3\\ \mathbb{E}_{\mathbb{P}}\left(Z^2\right)&=\mathrm{Var}_{\mathbb{P}}\left(Z\right)=1,\end{aligned}$$

from which we conclude that

$$\mathrm{Var}_{\mathbb{P}}\left(Z^2\right)=\mathbb{E}_{\mathbb{P}}\left(Z^4\right)-\mathbb{E}_{\mathbb{P}}\left(Z^2\right)^2=3-1=2,$$

and

$$\mathrm{Var}_{\mathbb{P}}\left[\left(\Delta W_{t_k}\right)^2\right]=\mathrm{Var}_{\mathbb{P}}\left[\left(\sqrt{\Delta t_k}Z\right)^2\right]=\left(\Delta t_k\right)^2\mathrm{Var}_{\mathbb{P}}\left(Z^2\right)=2\left(\Delta t_k\right)^2.$$

Combining all above implies

$$\begin{aligned}\mathrm{Var}_{\mathbb{P}}\left(S_n\right)&=\sum_{k=0}^{n-1}\mathrm{Var}_{\mathbb{P}}\left[\left(\sqrt{\Delta t_k}Z\right)^2\right]=2\sum_{k=0}^{n-1}\left(\Delta t_{k+1}\right)^2\\ &\leq 2\delta_n\sum_{k=0}^{n-1}\Delta t_{k+1}\underset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

where δ_n is defined as in (2.46), which concludes our proof. $\square$

Recall the definition of correlation.

Definition 2.65. *The correlation rho between two random variables X and Y is*

$$\rho=\mathrm{Corr}_{\mathbb{P}}(X,Y)=\frac{\mathrm{Cov}_{\mathbb{P}}(X,Y)}{\sqrt{\mathrm{Var}_{\mathbb{P}}(X)\mathrm{Var}_{\mathbb{P}}(Y)}}=\frac{\mathbb{E}_{\mathbb{P}}\left[\left(X-\mathbb{E}_{\mathbb{P}}(X)\right)\left(Y-\mathbb{E}_{\mathbb{P}}(Y)\right)\right]}{\sqrt{\mathrm{Var}_{\mathbb{P}}(X)\mathrm{Var}_{\mathbb{P}}(Y)}}. \tag{2.66}$$

If $\rho=0$, we say that X and Y are uncorrelated, otherwise we say that X and Y are correlated.

In the case of standard Brownian motions, say W_t and W_t^*, the definition of correlation translates to

$$\operatorname{Corr}_{\mathbb{P}}\left(W_t, W_t^*\right) = \frac{\mathbb{E}_{\mathbb{P}}\left(W_t W_t^*\right)}{t} \tag{2.67}$$

Theorem 2.68. *Assume that W_t and W_t^* are correlated Brownian motions with $\operatorname{Corr}_{\mathbb{P}}\left(W_t, W_t^*\right) = \rho$. Then*

$$\langle W_t, W_t^* \rangle = \rho t \tag{2.69}$$

Proof. We will begin by defining

$$\begin{aligned} X_t &= C\left(W_t + W_t^*\right) \\ C &= \frac{1}{\sqrt{2(1+\rho)}}. \end{aligned}$$

As linear combinations of normally distributed random variables are also normally distributed, it follows that X_t is normal. With similar argumentation we see that, over some partition $P^n_{[0,T]}$

$$\begin{aligned} \Delta X_{t_k} &= C\left[\Delta\left(W_{t_k} + W_{t_k}^*\right)\right] = C\left[W_{t_k} + W_{t_k}^* - \left(W_{t_{k-1}} + W_{t_{k-1}}^*\right)\right] = \\ &= C\left(\Delta W_{t_k} + \Delta W_{t_k}^*\right) \end{aligned}$$

is also normally distributed and, as a linear combination, also preserves the property of independence to past states $\{X_{t_s}\}_{s \leq k-1}$. Additionally,

$$\mathbb{E}_{\mathbb{P}}\left(\Delta X_{t_k}\right) = C\left[\mathbb{E}_{\mathbb{P}}\left(\Delta W_{t_k}\right) + \mathbb{E}_{\mathbb{P}}\left(\Delta W_{t_k}^*\right)\right] = 0, \tag{2.70}$$

and so, by first noting that $\mathbb{E}_{\mathbb{P}}\left(\Delta W_{t_k} \Delta W_{t_k}^*\right) = \rho \Delta t_k$ according to the definition of the correlation, we argue that

$$\begin{aligned} \operatorname{Var}_{\mathbb{P}}\left(\Delta X_{t_k}\right) &= \mathbb{E}_{\mathbb{P}}\left[\left(\Delta X_{t_k}\right)^2\right] \\ &= C^2\left(\mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_k}\right)^2\right] + \mathbb{E}_{\mathbb{P}}\left[\left(\Delta W_{t_k}^*\right)^2\right] + 2\mathbb{E}_{\mathbb{P}}\left[\Delta W_{t_k} \Delta W_{t_k}^*\right]\right) \\ &= C^2\left(\Delta t_k + \Delta t_k + 2\rho \Delta t_k\right) = \frac{1}{2(1+\rho)} 2(1+\rho) \Delta t_k \\ &= \Delta t_k. \end{aligned} \tag{2.71}$$

And now, by combining (2.70) and (2.71), we have proven that $\Delta X_{t_k} \sim N\left(0, \Delta t_k\right)$. Furthermore, by taking all the above listed properties of X_t into account, we can deduce that X_t is, in fact, a Brownian motion. Hence, by utilizing the results from theorem 2.60, we finally conclude that

$$\begin{aligned} & \langle X, X\rangle_t && = C^2 \langle W + W^*, W + W^*\rangle_t \\ & && = C^2 \left(\langle W, W\rangle_t + \langle W^*, W^*\rangle_t + 2\langle W, W^*\rangle_t\right) = 2C^2\left(t + \langle W, W^*\rangle_t\right) = t \\ \Leftrightarrow \quad & t + \langle W, W^*\rangle_t && = t(1+\rho) \\ \Leftrightarrow \quad & \langle W, W^*\rangle_t && = \rho t \end{aligned}$$

□

Theorem 2.72. *Let W_t be a Brownian motion. Then, the following identities hold.*

$$\langle t, t\rangle_t = 0 \tag{2.73}$$

$$\langle W, t\rangle_t = 0 \tag{2.74}$$

Proof. We will begin with (2.73), which follows directly from the definition.

$$\sum_{k_0}^{n-1} (\Delta t_k)^2 \leq \delta_n \sum_{k_0}^{n-1} \Delta t_k \underset{n\to\infty}{\longrightarrow} 0, \tag{2.75}$$

where δ_n is defined as in (2.46).

The convergence of the second expression (2.74) will yet again be regarded in the L^2 sense. By utilizing the results from the previous section, namely the expression (2.24), we see that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left[\left(\sum_{k=0}^{n-1} \Delta t_{k+1}\Delta W_{t_{k+1}}\right)^2\right] &= \sum_{k=0}^{n-1} \mathbb{E}_{\mathbb{P}}\left[(\Delta t_{k+1})^2\right]\Delta t_{k+1} = \sum_{k=0}^{n-1} (\Delta t_{k+1})^3 \\ &\leq \delta_n \sum_{k=0}^{n-1} (\Delta t_{k+1})^2 \underset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{2.76}$$

The results in (2.75) and (2.76) complete the proof. □

The theorems 2.60, 2.68 and 2.72 together form a so called 'box rule' for the covariation calculus for Brownian motions. The box rule illustrates the results from the above theorem (see table 2.1).

$\langle \cdot, \cdot \rangle_t$	t	W	W^*
t	0	0	0
W	0	t	ρt
W^*	0	ρt	t

Table 2.1: The box rule.

The concept of variations might seem a bit separate from the other parts of stochastic calculus covered so far. We will therefore conclude this section with two important theorems, that bind together the variation of Brownian motion with stochastic Itô integrals and their properties. The first one can be considered analogous to the *integration by parts* - formula in ordinary calculus.

Theorem 2.77. *Assume that W_t and W_t^* are standard Brownian motions. Then, in the differential form,*

$$d(W_t W_t^*) = W_t dW_t^* + W_t^* dW_t + d\langle W, W^* \rangle_t, \tag{2.78}$$

or, in the explicit form

$$W_T W_T^* = \int_0^T W_t dW_t^* + \int_0^T W_t^* dW_t + \langle W, W^* \rangle_T \tag{2.79}$$

Proof. The expression in the claim is easily achievable by simply expanding the sum in the definition of $\langle W, W^* \rangle_T$, namely

$$\begin{aligned}
\sum_{k=0}^{n-1} \Delta W_{t_{k+1}} \Delta W_{t_{k+1}} &= \sum_{k=0}^{n-1} W_{t_{k+1}} W^*_{t_{k+1}} + W_{t_k} W^*_{t_k} - W_{t_{k+1}} W^*_{t_k} - W_{t_k} W^*_{t_{k+1}} \\
&= \sum_{k=0}^{n-1} W_{t_{k+1}} W^*_{t_{k+1}} - W_{t_k} W^*_{t_k} \\
&\quad + \sum_{k=0}^{n-1} 2W_{t_k} W^*_{t_k} - W_{t_{k+1}} W^*_{t_k} - W_{t_k} W^*_{t_{k+1}} \\
&= W_T W_T^* - \sum_{k=0}^{n-1} W_{t_k} \Delta W^*_{t_{k+1}} - \sum_{k=0}^{n-1} W^*_{t_k} \Delta W_{t_{k+1}} \\
&\underset{n\to\infty}{\longrightarrow} W_T W_T^* - \int_0^T W_t dW_t^* - \int_0^T W_t^* dW_t,
\end{aligned}$$

that is,

$$\begin{aligned}
\langle W, W^* \rangle_T &= W_T W_T^* - \int_0^T W_t dW_t^* - \int_0^T W_t^* dW_t \\
\Leftrightarrow \quad W_T W_T^* &= \int_0^T W_t dW_t^* + \int_0^T W_t^* dW_t + \langle W, W^* \rangle_T
\end{aligned}$$

□

The term $d\langle W, W^* \rangle_t$ in the differential form of the integration by parts formula can be alternatively interpreted as ρdt, where $\rho = \mathrm{Corr}_{\mathbb{P}}(W_t, W_t^*)$. This does, indeed agree with both the definition of covariation, and the explicit form of the integration by parts formula, as can be seen from

$$\int_0^T \rho dt = \langle W, W^* \rangle_T .$$

Notational remark 2.80. The differential notation $d\langle W, W^* \rangle_t$ is often written as a multiplication of the increments of Brownian motion, namely

$$d\langle W, W^* \rangle_t = dW_t dW_t^* = \rho dt.$$

This will be the preferred notation also in this thesis.

Finally, we will present a formula that can be seen as a more general version of the Itô isometry.

Theorem 2.81. *Let W_t and W_t^* be Brownian motions and f and g adapted functions. If f and g are further Itô integrable, then*

$$\mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t,\omega) dW_t\right)\left(\int_0^T g(t,\omega) dW_t^*\right)\right] = \mathbb{E}_{\mathbb{P}}\left(\int_0^T f(t,\omega) g(t,\omega) dW_t dW_t^*\right) \quad (2.82)$$

Proof. The proof of this theorem requires advanced stochastic calculus methodology and results, for example the Kunita-Watanabe inequality, and is beyond the scope of this thesis. Interested readers are directed to [31]. □

Note that we acquire the Itô isometry as a limiting case of the above result.

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t,\omega)dW_t\right)^2\right] &= \mathbb{E}_{\mathbb{P}}\left[\left(\int_0^T f(t,\omega)dW_t\right)\left(\int_0^T f(t,\omega)dW_t\right)\right] = \\
&= \mathbb{E}_{\mathbb{P}}\left[\int_0^T f(t,\omega)^2 \left(dW_t\right)^2\right] \overset{(2.61)}{=} \mathbb{E}_{\mathbb{P}}\left[\int_0^T f(t,\omega)^2 dt\right] \\
&= \int_0^T \mathbb{E}_{\mathbb{P}}\left[f(t,\omega)^2\right] dt
\end{aligned}$$

2.4 A brief introduction into futures

In this section we give a very brief introduction of futures, futures trading and term structure. We give a review of the basic principles of the aforementioned topics, as well as describe in more detail our problem setting in this thesis.

Definition 2.83. *(Futures contracts)*
A futures contract is an agreement between two parties (a buyer and a seller) to exchange an asset at a future time for a specific price.

The definition 2.83 is, of course, an atomic and crudely simplified definition of the futures contracts, as the trading of futures contracts contains many details and rules.

Futures contracts are traded in many international exchanges, such as the Chicago Board of Trade, Tokyo Financial Exhange and NASDAQ OMX Commodities Europe in Oslo. This means that the two parties entering the futures contract never make the agreement directly between each other, but rather through brokers of the exchange.

Within this market framework, there are obvious payment default risks; as the time to maturity (the time to the delivery date) of the contract grow's longer, perhaps spanning over a couple of years, the more uncertain it is that the parties of the agreement are able to provide sufficient financial resources in order to fulfil the contract. In order to counter these risks, a procedure of *daily settlements* on *margin accounts* on the price of the futures contract is being followed. In practice, both parties of the contract deposit a fixed amount on their margin accounts (usually a substantially lower amount than the actual contract price) into which price fluctuations of the underlying asset are aggregated on a daily basis.

To give an example, assume that an investor takes a long position (that is, enter's a futures contract in the role of a buyer) on a NYMEX crude oil futures contract on 1000 U.S. barrels, and is obliged to deposit \$5000 on a margin account. The price of a barrel of crude oil is \$50, and the futures contract is valued at \$50000. Further, assume that the price of a barrel of crude oil drop's \$2 in one day. This leads to a withdrawal of $\$2 \cdot 1000 = \$2000$ from the investor's margin account.

The example illustrates the huge leverage of futures contracts; without remarkable initial wealth, one is able to enter contracts of substantial value, making futures contracts a compelling investment option for investors that are able to withstand high risk. This fact leads actually to perhaps the most important observation of the futures market: *futures contracts are mostly used as instruments to speculate on the price of the underlying asset.* It is rare for actors on the market to actually hold the contract until the delivery; the vast majority of the actors on the futures market will close out their position prior to the delivery. In essence, producers and consumers of the underlying commodity transfer the risk of price fluctuations to speculators, who are willing to undertake this risk in the hope of a large positive return [15].

Commodity futures

Commodity futures are futures contracts where the underlying asset is an investment commodity, such as precious metals, or consumption commodities, such as crops, cattle and coffee. As the investment portfolio in our focus contains mostly of consumption commodities, we will be concentrating on these in this chapter.

Commodity futures differ from other futures on financial assets, such as stocks and indices, in a very significant aspect. Namely, the underlying assets in commodity futures are concrete objects that require some physical production and storage. Therefore merely holding on to a commodity asset is likely to generate expenses, for example storage costs, to the owner of the asset. On the other hand, the owner of the physical asset is able to take advantage of sudden peaks in the demand of the asset, for example, in the event of a natural disaster; an extreme period of drought that effects the quality and quantity of different crops is likely to increase the demand of the effected crops quite suddenly. The aggregated effect of these properties for the owner of the asset is called *convenience yield.*

Definition 2.84. *(Convenience yield)*

The flow of services accruing to the holder of the commodity asset but not to the owner of a futures contract is called convenience yield [36].

The convenience yield plays a significant role in the pricing of commodity futures.

The pricing of futures contracts of investment assets (that do not generate any income for the holder of the asset) is quite straightforward; in the absence of arbitrage opportunities, it is easy to show [20] that the futures price F of a futures contract with time to maturity T is

$$F = Se^{rT},$$

where S is the spot price of the commodity and r is the risk-free interest rate. For commodities, and a constant convenience yield c, it can be shown that

$$F = Se^{(r-c)T}. \tag{2.85}$$

From (2.85) we see that the notion of 'negative interest' is well defined for commodity futures; if $r < c$, then the futures curve will be (as a function of time) decreasing. As previously mentioned, the convenience yield reflects the market's expectations concerning the future availability of the commodity.

The greater the possibility that shortages will occur, the higher the convenience yield. In this situation consumers of the underlying asset are willing to pay a premium today in order to secure the availability of the commodity. Hence, the futures curve will be declining. In this situation, we say that the futures contracts are in *backwardation.*

On the other hand, if users of the commodity have high inventories, there is very little chance of shortages in the near future and the convenience yield tends to be low. In this situation, the availability of the commodity is trivial, which will push the spot price lower, and we will observe an increasing futures curve. In this situation, we say that the futures contracts are in *contango*. If inventories are low, shortages are more likely and the convenience yield is usually higher, and the futures contracts will be in backwardation [20].

These different situations are illustrated in figure 2.4.

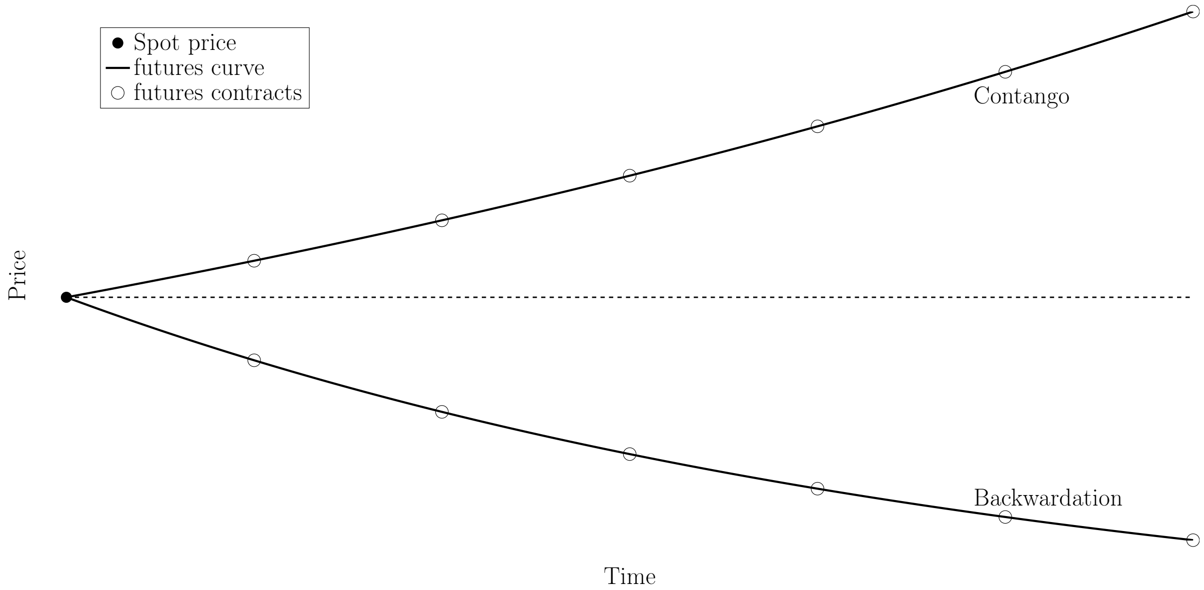


Figure 2.4: An illustration of different kinds of commodity futures curves.

Many commodity futures trading strategies utilize the coupling of the convenience yield and the term structure. For example, in [15] quite a few of the presented strategies fundamentally rely on going long on futures that are in contango and short on futures that are in backwardation. This is however not a riskless feat, and it is not uncommon for strategies to allocate capital to be invested according to some measure of risk. This is our aim in this thesis; to develop a measure of risk that could accurately predict possible anomalies in the returns of the trading strategy. A strategy that also outperforms the naïve risk measurement strategy described in the introduction, that is, we expect our risk measure to predict anomalies in the returns where outliers were observed.

Chapter 3

Model description

In this chapter we will describe and derive central formulae related to the model used in the simulations of the spot price and convenience yield of commodities. The model we have used is commonly referred to as the Schwartz two-factor model. The model is used, as such or as a basis, in several publications, including [36, 9, 3, 32], and has been originally derived in [16].

Recall that one has to make a clear distinction between pricing contingent claims on commodities and other common contingent claims, for example on currencies, by including the convenience yield to the model. As many factors influence the rate of convenience yield, for example seasonalities and sudden, unexpected changes in the local market, it is rarely sufficient to assume a constant convenience yield. Therefore also the convenience yield is modelled as a stochastic process in our model.

3.1 Analytic model

Both the spot price of the commodity and the convenience yield are assumed to be stochastic in our model. They are described with the following stochastic differential equations.

$$dS_t = (\mu - \delta_t)\, S_t dt + \sigma_S S_t dW_t^S \tag{3.1}$$

$$d\delta_t = \kappa\,(\alpha - \delta_t)\, dt + \sigma_\delta dW_t^\delta \tag{3.2}$$

$$dW_t^S dW_t^\delta = \rho dt \tag{3.3}$$

with

$$S_0 \equiv S$$
$$\delta_0 \equiv \delta$$

S_t denotes the spot price of the commodity at time t, and it is modelled as a geometric Brownian motion with a stochastic correction term (the convenience yield) applied to the drift factor (μ).

The convenience yield associated to the commodity at time t is denoted by δ_t,and is modelled as an Ornstein-Uhlenbeck process [7]. The mean reverting properties of the Ornstein-Uhlenbeck process is the central motivation behind the choice of the process.

Note that (3.3) implies that the increments of the standard Brownian motions in the stochastic differential equations are correlated. The motivation behind the correlation between the increments can be motivated through changes of inventory; when inventories of the commodity decrease, the spot price should increase since the commodity is scarce and the convenience yield should also increase since futures prices will not increase as much as the spot price, and vice versa when inventories increase [36].

The biggest challenge is obviously to derive an explicit closed form solution to the spot price of the commodity. This will be done in the following subsection 3.1.1.

3.1.1 Derivation and properties of the futures price

We will begin with the derivation of the solution of the separate processes of the spot price and the convenience yield.

Lemma 3.4. *Assume that a stochastic process* $\{S_t\}_{t\in\mathbb{R}_+}$ *follows a stochastic differential equation of the form* (3.1) *where* δ_t *is also assumed to be stochastic. Then the solution to* (3.1) *is*

$$S_t = S\exp\left[\left(\mu - \frac{\sigma_S^2}{2}\right)t - \int_0^t \delta_s ds + \sigma_S W_t^S\right]. \tag{3.5}$$

Proof. Let $X = \log(S_t)$. Now, by applying Itô's lemma (on the function $f = \log$), we see that

$$\begin{aligned}
dX_t &= \left(\frac{(\mu - \delta_t) S_t}{S_t} - \frac{\sigma_S^2 S_t^2}{2S_t^2} \right) dt + \frac{\sigma_S S_t}{S_t} dW_t^S \\
&= \left(\mu - \delta_t - \frac{\sigma_S^2}{2} \right) dt + \sigma_S dW_t^S \\
\Leftrightarrow d(\log S_t) &= \left(\mu - \delta_t - \frac{\sigma_S^2}{2} \right) dt + \sigma_S dW_t^S,
\end{aligned}$$

or, in the explicit form

$$\begin{aligned}
\log S_t &= \log S + \left(\mu - \frac{\sigma_S^2}{2} \right) t - \int_0^t \delta_s ds + \sigma_S \int_0^t dW_s^S \\
\Leftrightarrow S_t &= S \exp \left[\left(\mu - \frac{\sigma_S^2}{2} \right) t - \int_0^t \delta_s ds + \sigma_S W_t^S \right].
\end{aligned}$$

□

Next, we will derive the explicit solution to an Ornstein-Uhlenbeck process.

Lemma 3.6. *Assume that* $\{\delta_t\}_{t \in \mathbb{R}_+}$ *is a stochastic process that follows a stochastic differential equation of the form* (3.2)*, or equivalently, is an Ornstein-Uhlenbeck process. Then the solution to* (3.2) *is*

$$\delta_t = \alpha + e^{-\kappa t} (\delta - \alpha) + \sigma_\delta \int_0^t e^{\kappa(s-t)} dW_s^\delta. \tag{3.7}$$

Proof. Let $Y_t = \delta_t - \alpha$. Now, clearly

$$d\delta_t = dY_t = -\kappa Y_t dt + \sigma_\delta dW_t^\delta$$

We will do another change of variables. Let $Z_t = e^{\kappa t} Y_t$. Now, by applying the product rule [25] and Itô's lemma on $f(t,x) = xe^{\kappa t}$, we see that

$$\begin{aligned}
dZ_t &= \kappa e^{\kappa t} Y_t dt + e^{\kappa t} dY_t \\
&= \kappa e^{\kappa t} Y_t dt + e^{\kappa t} \left(-\kappa Y_t dt + \sigma_\delta dW_t^\delta \right) \\
&= \kappa e^{\kappa t} Y_t dt - \kappa e^{\kappa t} Y_t dt + e^{\kappa t} \sigma_\delta dW_t^\delta \\
&= e^{\kappa t} \sigma_\delta dW_t^\delta
\end{aligned}$$

The above is equivalent with

$$
\begin{aligned}
Z_t &= Z_0 + \sigma_\delta \int_0^t e^{\kappa s} dW_s^\delta \\
\Leftrightarrow Y_t &= e^{-\kappa t} Y_0 + e^{-\kappa t} \sigma_\delta \int_0^t e^{\kappa s} dW_s^\delta \\
&= e^{-\kappa t} Y_0 + \sigma_\delta \int_0^t e^{\kappa(s-t)} dW_s^\delta
\end{aligned}
$$

And finally, as $\delta_t = Y_t + \alpha$, we obtain the solution to the Ornstein-Uhlenbeck process, namely

$$
\delta_t = \alpha + e^{-\kappa t} (\delta - \alpha) + \sigma_\delta \int_0^t e^{\kappa(s-t)} dW_s^\delta .
$$

□

Our derivation of the fair price of the commodity is relying on results from the equivalent martingale measure theory. Hence we are required to perform a transformation of our model in order to take the risk premium into account. Luckily, this is a quite straightforward task as the transformation is almost a direct consequence of the Girsanov's theorem [7].

Lemma 3.8. *Let r denote the risk free interest rate, and λ denote the the market price of convenience yield risk. Both are assumed to be constant.*

The transformation of the model from the real-world probability measure $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$ can be executed using the following relations.

$$
\begin{aligned}
W_t^{S^*} &= W_t^S + \frac{\mu - r}{\sigma_S} t \\
W_t^{\delta^*} &= W_t^\delta + \lambda t,
\end{aligned}
\tag{3.9}
$$

or equivalently

$$
\begin{aligned}
dW_t^{S^*} &= dW_t^S + \frac{\mu - r}{\sigma_S} dt \\
dW_t^{\delta^*} &= dW_t^\delta + \lambda dt.
\end{aligned}
\tag{3.10}
$$

Proof. Note that as the convenience yield can not be hedged [16], this will have to be taken into account in the risk-free transformation.

The fact that W_t^{S*} ans $W_t^{\delta*}$ are standard Brownian motions is a direct consequence of the Girsanov's theorem[7]. Our risk-neutral model complies to the one in [2], and for a detailed deduction, the reader is directed to [16]. □

Under the previously defined risk-neutral measure, the system of processes is transformed to

$$\begin{aligned} dS_t &= (r - \delta_t)\, S_t dt + \sigma_S S_t dW_t^{S*} \\ d\delta_t &= (\kappa(\alpha - \delta_t) - \sigma_\delta \lambda)\, dt + \sigma_\delta W_t^{\delta*}, \end{aligned} \tag{3.11}$$

as can be seen by expanding the increments of the Brownian motions under the risk-neutral measure, namely

$$\begin{aligned} dS_t &= (r - \delta_t)\, S_t dt + \sigma_S S_t dW_t^{S*} = (r - \delta_t)\, S_t dt + \sigma_S S_t \left(dW_t^S + \frac{\mu - r}{\sigma_S} dt \right) \\ &= (r - \delta + \mu - r)\, S_t dt + \sigma_S S_t dW_t^S \\ &= (\mu - \delta_t)\, S_t dt + \sigma_S S_t dW_t^S. \end{aligned}$$

and

$$\begin{aligned} d\delta_t &= (\kappa(\alpha - \delta_t) - \sigma_\delta \lambda)\, dt + \sigma_\delta W_t^{\delta*} = (\kappa(\alpha - \delta_t) - \sigma_\delta \lambda)\, dt + \sigma_\delta \left(dW_t^\delta + \lambda dt \right) \\ &= \kappa(\alpha - \delta_t)\, dt + \sigma_\delta dW_t^\delta. \end{aligned}$$

We will now proceed to the main purpose of this chapter, the derivation of the solution to the spot price. As already mentioned in the introduction of this chapter, a version of the following theorem has been previously derived in [16].

Theorem 3.12. *In the economy outlined within this chapter, the current value $V_0(S_T)$ of a claim on a delivery of the commodity at a future time T is*

$$V_0(S_T) = S \exp\left[-\frac{\delta\left(1 - e^{-\kappa T}\right)}{\kappa} + C(T) \right], \tag{3.13}$$

where

$$
\begin{aligned}
C(T) = {} & \left[\frac{\sigma_\delta}{\kappa}\left(\frac{\sigma_\delta}{2\kappa} + \lambda + \sigma_S \rho\right) - \alpha\right] T \\
& + \frac{1 - e^{-\kappa T}}{\kappa}\left[\alpha + \frac{\sigma_\delta}{\kappa}\left(\sigma_S\rho - \lambda - \frac{\sigma_\delta}{\kappa}\right)\right] \\
& + \frac{\sigma_\delta^2\left(1 - e^{-2\kappa T}\right)}{4\kappa^3}.
\end{aligned} \tag{3.14}
$$

Proof. Results from the equivalent martingale theory [7] implies that the current value of a future contingent claim is equal to the expectation (under the risk-neutral measure) of the discounted value of the underlying asset at the maturity of the contract. In our case, the results from the equivalent martingale measure theory translates to

$$
V_0\left(S_T\right) = \mathbb{E}_{\mathbb{Q}}\left(e^{-rT} S_T\right) \tag{3.15}
$$

Lemma 3.4 allows us to rewrite the expression inside the expectation to

$$
e^{-rT} S_T = S \exp\left[\left(\mu - \frac{\sigma_S^2}{2} - r\right) T - \int_0^T \delta_s ds + \sigma_S W_T^S\right]. \tag{3.16}
$$

We will need evaluate the integral $\int_0^t \delta_s ds$ in the exponent. Utilizing lemma 3.6 and inserting (3.7) to (3.16) we assert that

$$
\int_0^T \delta_s ds = \alpha T + (\delta - \alpha)\int_0^t e^{-\kappa s} ds + \sigma_\delta \int_0^T \int_0^s e^{\kappa(u-s)} dW_u^\delta ds \tag{3.17}
$$

We will next evaluate the the double integral in (3.17). By applying Fubini's theorem we can change the order of integration and evaluate the Riemann integral, yielding

$$
\begin{aligned}
\int_0^T \int_0^s e^{\kappa(u-s)} dW_u^\delta ds &= \int_0^T e^{\kappa u} \int_u^T e^{-\kappa s} ds dW_u^\delta = -\int_0^T \frac{e^{\kappa u}}{\kappa}\left(e^{-\kappa T} - e^{-\kappa u}\right) dW_u^\delta \\
&= -\frac{1}{\kappa}\int_0^T e^{-\kappa(T-u)} - 1 dW_u^\delta = \frac{W_T^\delta}{\kappa} - \frac{1}{\kappa}\int_0^T e^{-\kappa(T-u)} dW_u^\delta
\end{aligned} \tag{3.18}
$$

By inserting (3.18) into (3.17) we obtain

$$
\int_0^T \delta_s ds = \alpha T + \frac{\delta - \alpha}{\kappa}\left(e^{-\kappa T} - 1\right) + \frac{\sigma_\delta W_T^\delta}{\kappa} - \frac{\sigma_\delta}{\kappa}\int_0^T e^{-\kappa(T-u)} dW_u^\delta.
$$

Inserting the evaluated integral into (3.16), we assert that

$$
\begin{aligned}
e^{-rT}S_T = S\exp\Bigg[\left(\mu - \frac{\sigma_S^2}{2} - r - \alpha\right)T - \frac{\delta-\alpha}{\kappa}\left(e^{-\kappa T}-1\right) \\
- \frac{\sigma_\delta W_T^\delta}{\kappa} + \frac{\sigma_\delta}{\kappa}\int_0^T e^{-\kappa(T-u)}dW_u^\delta + \sigma_S W_T^S\Bigg]
\end{aligned}
$$

In order the evaluate the expectation taken under the risk-neutral measure, we are required to perform a change of measure regarding the standard Brownian motions. We can deduce from (3.10) that

$$
\begin{aligned}
dW_t^{\delta^*} &= dW_t^\delta + \lambda dt \\
\Leftrightarrow \int_0^T e^{-\kappa(T-u)}dW_u^\delta &= \int_0^T e^{-\kappa(T-u)}dW_u^{\delta^*} - \lambda e^{-\kappa T}\int_0^T e^{\kappa u}du \\
\Leftrightarrow \int_0^T e^{-\kappa(T-u)}dW_u^\delta &= \int_0^T e^{-\kappa(T-u)}dW^{\delta^*} - \lambda\left(1-e^{-\kappa T}\right).
\end{aligned}
$$

The above combined with (3.9) and some rearranging leads to

$$
\begin{aligned}
e^{-rT}S_T &= S\exp\Bigg[\left(\frac{\sigma_\delta\lambda}{\kappa} - \frac{\sigma_S^2}{2} - \alpha\right)T + \left(\alpha - \delta - \frac{\sigma_\delta\lambda}{\kappa}\right)\frac{1-e^{-\kappa T}}{\kappa} \\
&\qquad - \frac{\sigma_\delta W_T^{\delta^*}}{\kappa} + \frac{\sigma_\delta}{\kappa}\int_0^T e^{-\kappa(T-u)}dW_u^{\delta^*} + \sigma_S W_T^{S^*}\Bigg] \\
&=: Se^{X_T}
\end{aligned}
$$

Here we denote the exponent as X_T for readability. Let us take a closer look at the exponent. Taking the expectation (under the risk-neutral measure) of the expression inside the exponent, yields

$$
\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left(X_T\right) &= \left(\frac{\sigma_\delta\lambda}{\kappa} - \frac{\sigma_S^2}{2} - \alpha\right)T + \left(\alpha - \delta - \frac{\sigma_\delta\lambda}{\kappa}\right)\frac{1-e^{-\kappa T}}{\kappa} \\
&\quad - \frac{\sigma_\delta}{\kappa}\underbrace{\mathbb{E}_{\mathbb{Q}}\left(W_T^{\delta^*}\right)}_{=0} + \frac{\sigma_\delta}{\kappa}\underbrace{\mathbb{E}_{\mathbb{Q}}\left(\int_0^T e^{-\kappa(T-u)}dW_u^{\delta^*}\right)}_{=0} + \sigma_S\underbrace{\mathbb{E}_{\mathbb{Q}}\left(W_T^{S^*}\right)}_{=0} \\
&= \left(\frac{\sigma_\delta\lambda}{\kappa} - \frac{\sigma_S^2}{2} - \alpha\right)T + \left(\alpha - \delta - \frac{\sigma_\delta\lambda}{\kappa}\right)\frac{1-e^{-\kappa T}}{\kappa}
\end{aligned}
$$

With similar argumentation we see that

$$\begin{aligned}
\operatorname{Var}_{\mathbb{Q}}(X_T) &= \mathbb{E}_{\mathbb{Q}}\left(X_T^2\right) - \mathbb{E}_{\mathbb{Q}}(X_T)^2 \\
&= \mathbb{E}_{\mathbb{Q}}\left[\left(\frac{\sigma_\delta W_T^{\delta^*}}{\kappa}\right)^2 + \left(\frac{\sigma_\delta e^{-\kappa T}}{\kappa}\int_0^T e^{\kappa u} dW_u^{\delta^*}\right)^2 + \left(\sigma_S W_T^{S^*}\right)^2\right. \\
&\quad - \frac{\sigma_\delta^2 W^{\delta^*} e^{-\kappa T}}{\kappa}\int_0^T e^{\kappa u} dW_u^{\delta^*} - \frac{\sigma_\delta \sigma_S W_T^{\delta^*} W_T^{S^*}}{\kappa} \\
&\quad \left. + \frac{\sigma_\delta \sigma_S e^{-\kappa T} W_T^{S^*}}{\kappa}\int_0^T e^{\kappa u} dW_u^{\delta^*}\right].
\end{aligned}$$

Utilizing the results of the Itô isometry (see theorems 2.29 and 2.81) we can separately evaluate the terms in the expression inside the expectation above. By recalling that $\left(dW_t^{\delta^*}\right)^2 = dW_t^{\delta^*} dW_t^{\delta^*} = dt$, we see that

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{\sigma_\delta W_T^{\delta^*}}{\kappa}\right)^2\right] = \frac{\sigma_\delta^2}{\kappa}\mathbb{E}_{\mathbb{Q}}\left(W_T^{\delta^{*2}}\right) = \frac{\sigma_\delta^2}{\kappa}T, \tag{3.19}$$

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{\sigma_\delta e^{-\kappa T}}{\kappa}\int_0^T e^{\kappa u} dW_u^{\delta^*}\right)^2\right] &= \frac{\sigma_\delta^2 e^{-2\kappa T}}{\kappa^2}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_0^T e^{\kappa u} dW_u^{\delta^*}\right)^2\right] \\
&= \frac{\sigma_\delta^2 e^{-2\kappa T}}{\kappa^2}\int_0^T e^{2\kappa u} du \\
&= \frac{\sigma_\delta^2\left(1 - e^{-2\kappa T}\right)}{2\kappa^3},
\end{aligned} \tag{3.20}$$

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\sigma_S W_T^{S^*}\right)^2\right] = \sigma_S^2 \mathbb{E}_{\mathbb{Q}}\left[\left(W_T^{S^*}\right)^2\right] = \sigma_S^2 T \tag{3.21}$$

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left[\frac{\sigma_\delta^2 W_T^{\delta^*} e^{-\kappa T}}{\kappa^2}\int_0^T e^{\kappa u} dW_u^{\delta^*}\right] &= \frac{\sigma_\delta^2 e^{-\kappa T}}{\kappa^2}\mathbb{E}_{\mathbb{Q}}\left[\int_0^T e^{\kappa u}\left(dW_u^{\delta^*}\right)^2\right] \\
&= \frac{\sigma_\delta^2 e^{-\kappa T}}{\kappa^2}\int_0^T e^{\kappa u} du \\
&= \frac{\sigma_\delta^2\left(1 - e^{-\kappa T}\right)}{\kappa^3},
\end{aligned} \tag{3.22}$$

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{\sigma_\delta\sigma_S W_T^{\delta^*}W_T^{S^*}}{\kappa}\right]=\frac{\sigma_\delta\sigma_S}{\kappa}\mathbb{E}_{\mathbb{Q}}\left(W_T^{\delta^*}W_T^{S^*}\right)=\frac{\sigma_\delta\sigma_S}{\kappa}\int_0^T\rho dt \\ =\frac{\sigma_\delta\sigma_S\rho}{\kappa}T, \tag{3.23}$$

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{\sigma_\delta\sigma_S e^{-\kappa T}W_T^{S^*}}{\kappa}\int_0^T e^{\kappa u}dW_T^{\delta^*}\right]=\frac{\sigma_\delta\sigma_S e^{-\kappa T}}{\kappa}\rho\int_0^T e^{ku}du \\ =\frac{\sigma_\delta\sigma_S\rho\left(1-e^{-\kappa T}\right)}{\kappa^2} \tag{3.24}$$

Now, by combining equations (3.19) to (3.24) and rearranging the terms, we can conclude that

$$\begin{aligned}\mathrm{Var}_{\mathbb{Q}}\left(X_T\right)=&\left(\frac{\sigma_\delta^2}{\kappa^2}+\sigma_S^2-\frac{2\sigma_\delta\sigma_S\rho}{\kappa}\right)T\\&+\frac{2\left(1-e^{-\kappa T}\right)}{\kappa^2}\left(\sigma_\delta\sigma_S\rho-\frac{\sigma_\delta^2}{\kappa}\right)+\frac{\sigma_\delta^2\left(1-e^{-2\kappa T}\right)}{2\kappa^3}\end{aligned} \tag{3.25}$$

As a linear combination of normally distributed random variables, X_T is also normally distributed, and we can hence utilize the properties of the normal distribution and evaluate (3.15). More explicitly, the property that states that if $Z\sim N(\xi,\nu)$, then $e^Z\sim \ln\left(N(\xi,\nu)\right)$. We finally conclude that

$$\begin{aligned}V_0\left(S_T\right)=&\mathbb{E}_{\mathbb{Q}}\left(Se^{X_T}\right)=S\exp\left(\mathbb{E}_{\mathbb{Q}}(X_T)+\frac{1}{2}\mathrm{Var}_{\mathbb{Q}}(X_T)\right)\\=&S\exp\left\{\left[\frac{\sigma_\delta}{\kappa}\left(\frac{\sigma\delta}{2\kappa}+\lambda+\sigma_S\rho\right)-\alpha\right]T\right.\\&+\frac{1-e^{-\kappa T}}{\kappa}\left[\alpha-\delta+\frac{\sigma_\delta}{\kappa}\left(\sigma_S\rho-\lambda-\frac{\sigma_\delta}{\kappa}\right)\right]\\&\left.+\frac{\sigma_\delta^2\left(1-e^{-2\kappa T}\right)}{4\kappa^3}\right\}\\=&S\exp\left[-\frac{\delta\left(1-e^{-\kappa T}\right)}{\kappa}+C(T)\right],\end{aligned} \tag{3.26}$$

where

$$\begin{aligned}C(T)=&\left[\frac{\sigma_\delta}{\kappa}\left(\frac{\sigma_\delta}{2\kappa}+\lambda+\sigma_S\rho\right)-\alpha\right]T\quad+\frac{1-e^{-\kappa T}}{\kappa}\left[\alpha-\delta+\frac{\sigma_\delta}{\kappa}\left(\sigma_S\rho-\lambda-\frac{\sigma_\delta}{\kappa}\right)\right]\\&+\frac{\sigma_\delta^2\left(1-e^{-2\kappa T}\right)}{4\kappa^3}\end{aligned}$$

$\square$

Let $F \equiv F(S,\delta,T)$ denote the futures price (with the commodity being the underlying asset). By invoking the assumption of the absence of arbitrage opportunities, the relation between the futures (with maturity T) price and the value of a claim on a delivery of the commodity at a future time T becomes

$$V_0\left(F(S,\delta,T) - S_T\right) = 0 \Leftrightarrow F(S,\delta,T) = e^{rT}V_0(S_T).$$

We have now obtained a closed form solution for the futures price.

$$F(S,\delta,T) = S\exp\left[-\frac{\delta\left(1-e^{-\kappa T}\right)}{\kappa} + A(T)\right] \tag{3.27}$$

where

$$\begin{aligned} A(T) = &\left[r - \alpha + \frac{\sigma_\delta}{\kappa}\left(\frac{\sigma_\delta}{2\kappa} + \lambda + \sigma_S\rho\right)\right]T + \frac{\sigma_\delta^2\left(1-e^{-2\kappa T}\right)}{4\kappa^3} \\ &+ \frac{1-e^{-\kappa T}}{\kappa}\left[\alpha - \delta + \frac{\sigma_\delta}{\kappa}\left(\sigma_S\rho - \lambda - \frac{\sigma_\delta}{\kappa}\right)\right] \end{aligned} \tag{3.28}$$

Recall that κ denotes the mean-reversion speed of the convenience yield process. Thus, a great increase in the value of κ can be interpreted as an equivalent decrease on the stochastic properties of the convenience yield. More precisely, the convenience yield δ_t approaches the long-term mean α as κ increases. If the derived futures price is subjected to this phenomenon, we see that

$$\lim_{\kappa\to\infty} F(S,\delta,t) = \lim_{\kappa\to\infty} S\exp\left[-\frac{\delta\left(1-e^{-\kappa T}\right)}{\kappa} + A(T)\right] = Se^{(r-\alpha)T}, \tag{3.29}$$

as

$$\begin{aligned} \frac{\delta\left(1-e^{-\kappa T}\right)}{\kappa} &\xrightarrow[\kappa\to\infty]{} 0 \\ A(T) &\xrightarrow[\kappa\to\infty]{} (r-\alpha)T. \end{aligned}$$

Equation (3.29) corresponds to the standard pricing formula of commodity futures where the convenience yield is assumed to be constant[20]. Hence, we acquire the futures price of commodity futures, where the convenience yield is assumed to be constant, as a limiting case of our model.

We can further verify the correctness of our results with the Feynman-Kac theorem.

Theorem 3.30. *(Feynman-Kac)*

Assume that $\boldsymbol{X}_t \in \mathbb{R}^{n\times 1}$ *is an* n *dimensional stochastic process defined by*

$$d\boldsymbol{X}_t = \boldsymbol{\mu}(t, \boldsymbol{X}_t)dt + \boldsymbol{\sigma}(t, \boldsymbol{X}_t)d\boldsymbol{W}_t$$

where

$$\begin{aligned}
\boldsymbol{X}_t &= \left[X_t^1, \dots, X_t^n\right]^T \\
\boldsymbol{\mu}(t, \boldsymbol{X}_t) &= \left[\mu_1(t, \boldsymbol{X}_t), \dots, \mu_n(t, \boldsymbol{X}_t)\right]^T \\
\boldsymbol{\sigma}(t, \boldsymbol{X}_t) &= \begin{pmatrix} \sigma_{1,1}(t, \boldsymbol{X}_t) & \cdots & \sigma_{1,m}(t, \boldsymbol{X}_t) \\ \vdots & \ddots & \vdots \\ \sigma_{n,1}(t, \boldsymbol{X}_t) & \cdots & \sigma_{n,m}(t, \boldsymbol{X}_t) \end{pmatrix} \\
\boldsymbol{W}_t &= \left[W_t^1, \dots, W_t^m\right]^T
\end{aligned}$$

and W_t^k *are Brownian motions for all* $k \in \{1, \dots, m\}$*. Define an operator*

$$\mathcal{A} = \sum_{k=1}^{n} \frac{\partial \mu_k}{\partial X_t^k}(t, \boldsymbol{X}_t) + \frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \left(\boldsymbol{\sigma}\boldsymbol{\sigma}^T\right)_{k,j}}{\partial X_t^k \partial X_t^j}(t, \boldsymbol{X}_t). \tag{3.31}$$

The theorem then states that the solution to the boundary condition problem

$$\begin{aligned}
\frac{\partial V}{\partial t}(t, \boldsymbol{X}_t) + \mathcal{A}V(t, \boldsymbol{X}_t) - rV(t, \boldsymbol{X}_t) &= 0 \\
V(T, \boldsymbol{X}_T) &= F(\boldsymbol{X}_T)
\end{aligned} \tag{3.32}$$

is of the form

$$V(t, \boldsymbol{X}_t) = \mathbb{E}_{\mathbb{Q}}\left[e^{r(T-t)}F(\boldsymbol{X}_T)\right]. \tag{3.33}$$

Proof. The proof is omitted, see [35]. □

We will begin by transforming our model to a setting similar to the one in the Feynman-Kac theorem.

The risk-free processes in our model can be written in a vectorized form as

$$\begin{pmatrix} dS_t \\ d\delta_t \end{pmatrix} = \hat{\boldsymbol{\mu}}dt + \hat{\boldsymbol{\sigma}}\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \end{pmatrix}, \tag{3.34}$$

where

$$
\begin{aligned}
\hat{\boldsymbol{\mu}} &= \begin{pmatrix} (r-\delta_t)S_t \\ \kappa(\alpha-\delta_t) - \lambda\sigma_\delta \end{pmatrix} \\
\hat{\boldsymbol{\sigma}} &= \begin{pmatrix} \sigma_S S_t & 0 \\ \sigma_\delta \rho & \sigma_\delta \sqrt{1-\rho^2} \end{pmatrix} \\
& Z_t^1, Z_t^2 \text{ are standard Brownian motions} \\
& \mathrm{Corr}(Z_t^1, Z_t^2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}
\tag{3.35}
$$

The transformation on the two uncorrelated standard Brownian motions generate correlated ones with the desired correlation ρ, and is actually a special case of the Cholesky decomposition [25]. The claim can also be easily verified by straightforward calculations on the correlations on the transformations.

With elementary linear algebra we see that

$$
\hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T = \begin{pmatrix} \sigma_S^2 S_t^2 & \sigma_S \sigma_\delta \rho S_t \\ \sigma_S \sigma_\delta \rho S_t & \sigma_\delta^2 \end{pmatrix}.
$$

The generator operator for the partial differential equation now becomes

$$
\begin{aligned}
\mathcal{A} =& (r-\delta_t)S_t \frac{\partial}{\partial S_t} + (\kappa(\alpha-\delta_t) - \lambda\sigma_\delta) \frac{\partial}{\partial \delta_t} \\
&+ \frac{1}{2}\left(\sigma_S^2 S_t^2 \frac{\partial^2}{\partial S_t^2} + \sigma_\delta^2 \frac{\partial^2}{\partial \delta_t^2}\right) + \sigma_S \sigma_\delta \rho S_t \frac{\partial}{\partial S_t \partial \delta_t}.
\end{aligned}
$$

Let $\Upsilon \equiv \Upsilon(S_t, \delta_t, t)$ denote the solution to $V_t(S_T) = \mathbb{E}_{\mathbb{Q}}\left(e^{-r(T-t)} S_T\right)$, that is, the solution to the value on a future claim of the commodity at time t. The Feynamn-Kac theorem now gives us the following condition to the solution to $V_t(S_T)$.

$$
\begin{aligned}
& \Upsilon_t + \mathcal{A}\Upsilon - r\Upsilon = 0 \\
\Leftrightarrow\ & \Upsilon_t + (r-\delta_t)S_t \Upsilon_{S_t} + (\kappa(\alpha-\delta_t) - \lambda\sigma_\delta)\Upsilon_{\delta_t} + \\
& \frac{1}{2}\left(\sigma_S^2 S_t^2 \Upsilon_{S_t S_t} + \sigma_\delta^2 \Upsilon_{\delta_t \delta_t}\right) + \sigma_S \sigma_\delta \rho S_t \Upsilon_{S_t \delta_t} - r\Upsilon = 0
\end{aligned}
\tag{3.36}
$$

In accordance to theorem 3.12, we propose that

$$
\Upsilon = V_t(S_T) = S_t \exp\left[-\frac{\delta_t\left(1 - e^{-\kappa(T-t)}\right)}{\kappa} + C(T-t)\right]
\tag{3.37}
$$

Where $C(\cdot)$ is defined as in (3.14). Note that this is just a natural extenstion of (3.26), that specifies the value of a future claim on the commodity at time t instead of the value at current time (equivalent to $t = 0$). The only correctional action we need to take is to shift the time to maturity in relation to the difference of the current time and time t.

We will denote $\phi = e^{-\kappa(T-t)}$ for notational convenience. Simple calculus and partial derivation implies that

$$\Upsilon_{S_t} = \frac{\Upsilon}{S} \tag{3.38}$$

$$\Upsilon_{S_tS_t} = 0 \tag{3.39}$$

$$\Upsilon_{\delta_t} = -\frac{1-\phi}{\kappa}\Upsilon \tag{3.40}$$

$$\Upsilon_{\delta_t\delta_t} = \left(\frac{1-\phi}{\kappa}\right)^2 \Upsilon \tag{3.41}$$

$$\Upsilon_{S_t\delta_t} = -\frac{1-\phi}{\kappa S_t}\Upsilon \tag{3.42}$$

$$\begin{aligned}\Upsilon_t = \Upsilon \Bigg(&\delta\phi - \frac{\sigma_\delta^2}{2\kappa^2} - \frac{\sigma_\delta\lambda}{\kappa} - \frac{\sigma_\delta\sigma_S\rho}{\kappa} + \alpha - \alpha\phi - \frac{\sigma_\delta\sigma_S\rho}{\kappa}\phi + \frac{\sigma_\delta\lambda}{\kappa}\phi \\ &+\frac{\sigma_\delta^2}{\kappa^2} - \frac{\sigma_\delta^2}{2\kappa^2}\phi\Bigg) =: \Upsilon\mathfrak{C}\end{aligned} \tag{3.43}$$

By plugging equations (3.38) to (3.43) into (3.36), the previously constructed partial differential equation now becomes

$$\begin{aligned}&\Upsilon_t + \mathcal{A}\Upsilon - r\Upsilon \\ =&\Upsilon\mathfrak{C} + r\Upsilon - \delta\Upsilon - \alpha\Upsilon + \delta\Upsilon + \alpha\phi\Upsilon - \delta\phi\Upsilon + \frac{\sigma_\delta\lambda}{\kappa}\Upsilon - \frac{\sigma_\delta\lambda}{\kappa}\phi\Upsilon \\ &+ \frac{\sigma_\delta^2}{\kappa^2}\Upsilon - \frac{\sigma_\delta^2}{\kappa^2}\phi\Upsilon + \frac{\sigma_\delta^2}{2\kappa^2}\phi^2\Upsilon + \frac{\sigma_\delta\sigma_S\rho}{\kappa}\Upsilon - \frac{\sigma_\delta\sigma_S\rho}{\kappa}\phi\Upsilon - r\Upsilon \\ =&\Upsilon\mathfrak{C} + \Upsilon \underbrace{\left(-\delta\phi + \frac{\sigma_\delta^2}{2\kappa^2} + \frac{\sigma_\delta\lambda}{\kappa} + \frac{\sigma_\delta\sigma_S\rho}{\kappa} - \alpha + \alpha\phi + \frac{\sigma_\delta\sigma_S\rho}{\kappa}\phi - \frac{\sigma_\delta\lambda}{\kappa}\phi - \frac{\sigma_\delta^2}{\kappa^2} + \frac{\sigma_\delta^2}{2\kappa^2}\phi\right)}_{=-\mathfrak{C}} \\ =&0,\end{aligned}$$

which is what we were set to show.

3.2 Discretized Model

We have now established a complete model of the spot price dynamics of commodities with closed form solution for the futures price with the respective commodities as the underlying assets. The model is, however, constructed within the continuous time domain. This is inconvenient as an essential part of this thesis is to simulate both spot price and convenience yield trajectories for given commodities, which naturally is performed in the discrete time domain. Hence, we need to perform a discretization on our model. This is not as trivial as one might think, and different methods, or schemes, have been developed and analysed over time.

In this section we will present the chosen scheme for our model, namely, the Milstein scheme. The Milstein scheme utilizes Itô's lemma to obtain enhanced accuracy. The scheme is frequently used, and has been subject to comparative convergence analysis with other schemes (e.g. [17], [26]). For example in [21], the Milstein scheme outperformed the Euler scheme, which is another common (and less complex) discretization scheme. We will first provide an elementary derivation of the Milstein scheme for a general process (without any convergence analysis), after which we will construct a discrete version of our model which is applicable for simulation purposes.

3.2.1 The Milstein scheme

In this section we will derive the discretization for a process of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{3.44}$$

where, as usual, W_t is a standard Brownian motion. The usual setting for the discretization is that we want to simulate a process within a time-interval, say $t \in [a, b]$. The time interval is then meshed into a partition of desired number of time-steps, say n steps. We end up with a set of time-steps in which the process is to be evaluated, namely $\{t_i \in [a, b] \quad | \quad i \in \{0, \dots, n\}, \quad a = t_0 < t_1 < \dots < t_{n-1} < t_n = b\}$.

Let us now consider the case of two consecutive time-steps t_i and t_{i+1}. The process takes now the form (within $[t_i, t_{i+1}]$)

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} \mu(X_s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s. \tag{3.45}$$

The next step is to expand the drift-, and diffusion-coefficients using Itô's lemma (note that we are implicitly assuming that the coefficients are smooth enough and fulfil the conditions of Itô's lemma). Itô's lemma yields

$$\begin{aligned} d\mu(X_t) &= \left(\mu(X_t)\mu'(X_t) + \frac{1}{2}\sigma(X_t)^2\mu''(X_t)\right)dt + \sigma(X_t)\mu'(X_t)dW_t \\ d\sigma(X_t) &= \left(\mu(X_t)\sigma'(X_t) + \frac{1}{2}\sigma(X_t)^2\sigma''(X_t)\right)dt + \sigma(X_t)\sigma'(X_t)dW_t, \end{aligned}$$

or equivalently

$$\begin{aligned} \mu(X_t) &= \mu(X) + \int_0^t \mu(X_s)\mu'(X_s) + \frac{1}{2}\sigma(X_s)^2\mu''(X_s)ds + \int_0^t \sigma(X_s)\mu'(X_s)dW_s \\ \sigma(X_t) &= \sigma(X) + \int_0^t \mu(X_s)\sigma'(X_s) + \frac{1}{2}\sigma(X_s)^2\sigma''(X_s)ds + \int_0^t \sigma(X_s)\sigma'(X_s)dW_s. \end{aligned} \tag{3.46}$$

By inserting the expanded form (3.46) of the coefficients, equation (3.45) can be expressed as

$$\begin{aligned} X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} \mu(X_{t_i}) &+ \left(\int_{t_i}^{s} \mu(X_v)\mu'(X_v) + \frac{1}{2}\sigma(X_v)^2\mu''(X_v)dv \right. \\ &\left. + \int_{t_i}^{s} \sigma(X_v)\mu'(X_v)dW_v\right) ds \\ + \int_{t_i}^{t_{i+1}} \sigma(X_{t_i}) &+ \left(\int_{t_i}^{s} \mu(X_v)\sigma'(X_v) + \frac{1}{2}\sigma(X_v)^2\sigma''(X_v)dv \right. \\ &\left. + \int_{t_i}^{s} \sigma(X_v)\sigma'(X_v)dW_v\right) dW_s \end{aligned} \tag{3.47}$$

The only term with order less or equal one is $dW_v dW_s$, which can be seen by slightly extending the notation of the quadratic variations. One can argument that

$$\begin{aligned} dsdv &= \mathcal{O}\left((dt)^2\right) \\ dW_v ds &= \mathcal{O}\left((dt)^{\frac{3}{2}}\right). \end{aligned}$$

The proofs of this argument is omitted, see for example [18].

Hence, when we desire to obtain a method that converges of order 1, the other terms are negligible, and we obtain

$$\begin{aligned} X_{t_{i+1}} \approx X_{t_i} &+ \int_{t_i}^{t_{i+1}} \mu(X_s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s \\ &+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \sigma(X_v)\sigma'(X_v)dW_v dW_s \end{aligned} \tag{3.48}$$

The integrals in (3.48) can be further approximated with the familiar left-point rule, which states that $\int_a^b f(x)dx \approx f(a)\int_a^b dx = f(a)(b-a)$. The previously mentioned Euler scheme actually relies almost entirely on the left-point rule[17].

This yields

$$X_{t_{i+1}} \approx X_{t_i} + \mu\left(X_{t_i}\right)\Delta t_i + \sigma\left(X_{t_i}\right)\Delta W_{t_i} + \sigma(X_{t_i})\sigma'(X_{t_i})\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_v dW_s, \tag{3.49}$$

where

$$\Delta x_i = x_{i+1} - x_i$$

Let us have a closer look at the last double integral in (3.49). We immediately notice that

$$\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_v dW_s = \int_{t_i}^{t_{i+1}} W_s - W_{t_i} dW_s = \int_{t_i}^{t_{i+1}} W_s dW_s - W_{t_i}W_{t_{i+1}} + W_{t_i}^2 \tag{3.50}$$

By applying Itô's lemma to the function $f(W_s) = W_s^2$ we see that

$$dW_s^2 = 2W_s dW_s + \left(\frac{1}{2}\cdot 2\right) ds = 2W_s dW_s + ds,$$

or equivalently

$$\begin{aligned} W_{t_{i+1}}^2 &= W_{t_i}^2 + \int_{t_i}^{t_{i+1}} 2W_s dW_s + \Delta t_i \\ \Leftrightarrow \int_{t_i}^{t_{i+1}} W_s dW_s &= \frac{1}{2}\left(W_{t_{i+1}}^2 - W_{t_i}^2 - \Delta t_i\right). \end{aligned} \tag{3.51}$$

Now, by plugging expression (3.51) into the double integral, we assert that

$$\begin{aligned} \int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_v dW_s &= \frac{1}{2}\left(W_{t_{i+1}}^2 - 2W_{t_i}W_{t_{i+1}} + W_{t_i}^2 - \Delta t_i\right) \\ &= \frac{1}{2}\left(\left(\Delta W_{t_i}\right)^2 - \Delta t_i\right). \end{aligned}$$

Finally, by inserting the newly acquired value of the double integral into equation (3.49), we conclude with the final form of the approximation

$$\begin{aligned} X_{t_{i+1}} &\approx X_{t_i} + \mu\left(X_{t_i}\right)\Delta t_i + \sigma\left(X_{t_i}\right)\Delta W_{t_i} + \frac{1}{2}\sigma(X_{t_i})\sigma'(X_{t_i})\left(\left(\Delta W_{t_i}\right)^2 - \Delta t_i\right) \\ &=: M(i), \end{aligned} \tag{3.52}$$

which is known as the Milstein scheme. Now simulations of the process X_t within a time domain $t \in [a, b]$ can be performed by computing iterations of the following difference equation.

$$\begin{aligned} X_{t_i} &= M(i-1), \quad i \in \{1, \ldots, n\} \\ X_{t_i} &= X_a \equiv X, \quad\; i = 0. \end{aligned} \tag{3.53}$$

3.2.2 The Milstein scheme applied to the analytic model

As our model is built on a multidimensional system of stochastic processes, we are again required to transform our model according to the Cholesky decomposition. Hence, our starting point will be

$$\begin{pmatrix} dS_t \\ d\delta_t \end{pmatrix} = \hat{\boldsymbol{\mu}} dt + \hat{\boldsymbol{\sigma}} \begin{pmatrix} dZ_t^1 \\ dZ_t^2 \end{pmatrix},$$

where

$$\begin{aligned} &\hat{\boldsymbol{\mu}} = \begin{pmatrix} (\mu - \delta_t) S_t \\ \kappa(\alpha - \delta_t) \end{pmatrix} \\ &\hat{\boldsymbol{\sigma}} = \begin{pmatrix} \sigma_S S_t & 0 \\ \sigma_\delta \rho & \sigma_\delta \sqrt{1-\rho^2} \end{pmatrix} \\ &Z_t^1, Z_t^2 \text{ are standard Brownian motions} \\ &\mathrm{Corr}(Z_t^1, Z_t^2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Note that we will only be deriving the discretization schemes to our model under the $\mathbb{P}$ measure. This is simply because the model under the risk-neutral measure is useful only for pricing purposes, and as we have already derived the pricing formula for commodity futures, we have no further interest in inspecting the price dynamics under the $\mathbb{Q}$ measure. If one should desire to perform a discretization under the $\mathbb{Q}$ measure, one could simply redefine $\tilde{\mu}_S$, $\tilde{\mu}_\delta$, $\tilde{\sigma}_S$, $\tilde{\sigma}_\delta$ to correspond with the system under the $\mathbb{Q}$ measure (3.11).

We will denote

$$\tilde{\mu}_S(S_t) = \mu - \delta_t$$
$$\tilde{\sigma}_S(S_t) = \sigma_S S_t$$
$$\tilde{\mu}_\delta(\delta_t) = \kappa(\alpha - \delta_t)$$
$$\tilde{\sigma}_\delta(\delta_t) = \sigma_\delta,$$

which allows us to express our system of process in the same form as the generalized process in the previous chapter (3.44), resulting in

$$\begin{aligned} dS_t &= \tilde{\mu}_S(S_t)dt + \tilde{\sigma}_S(S_t)dZ_t^1 \\ d\delta_t &= \tilde{\mu}_\delta(\delta_t)dt + \tilde{\sigma}_\delta(\delta_t)\left(\rho dZ_t^1 + \sqrt{1-\rho^2}dZ_t^2\right). \end{aligned} \tag{3.54}$$

Let $M_S(\cdot)$ and $M_\delta(\cdot)$ denote the Milstein discretization schemes for the two processes. The explicit form of the discretizations is now obtained by simply inserting the alternatively expressed forms (3.54) of the two processes into (3.52), yielding[1]

$$M_S(i) = S_{t_i} + (\mu - \delta_{t_i})\Delta t_i + \sigma_S S_{t_i}\Delta Z_{t_i}^1 + \frac{1}{2}\sigma_S^2 S_{t_i}\left(\left(\Delta Z_{t_i}^1\right)^2 - 1\right) \tag{3.55}$$

$$M_\delta(i) = \delta_{t_i} + \kappa(\alpha - \delta_{t_i})\Delta t_i + \sigma_\delta\left(\rho\Delta Z_{t_i}^1 + \sqrt{1-\rho^2}\Delta Z_{t_i}^2\right), \tag{3.56}$$

which leads to the difference equation system (within a time domain $[a,b]$)

$$\begin{array}{ll} S_{t_i} & = M_S(i-1), \quad i \in \{1,\ldots,n\} \\ \delta_{t_i} & = M_\delta(i-1), \quad i \in \{1,\ldots,n\} \\ S_0 & \equiv S \\ \delta_0 & \equiv \delta \end{array}, \tag{3.57}$$

where $\{t_i\}_{i\in\{0,\ldots,n\}}$ is a partition of the domain $[a,b]$, or more precisely

$$\{t_i \in [a,b] \quad | \quad i \in \{0,\ldots,n\}, \quad a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b\}.$$

Figures 3.1 and 3.2, that graph simulations of our model with different parametersets, illustrate the discretization process. Using the notation from previous chapter, both simulations were run with $n = 780$ and $\Delta t_i \equiv \Delta t = 1/260$, which resembles path evolutions over 3 years in business days.

[1]Note, that $\tilde{\sigma}'_\delta(\delta_t) = 0$.

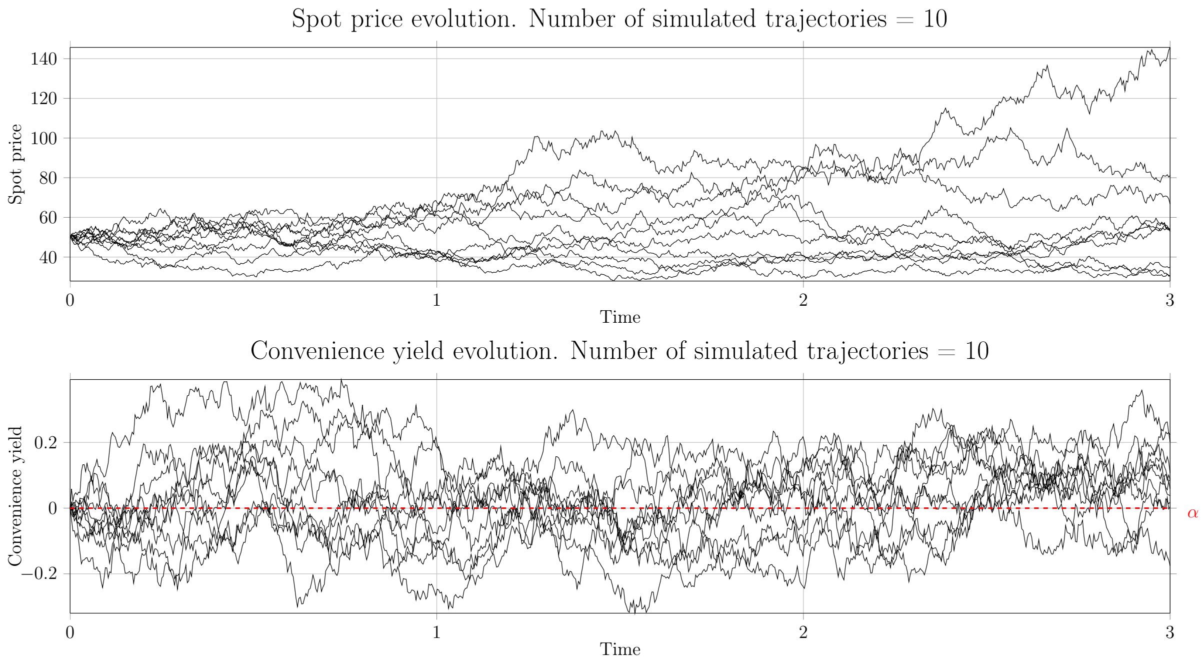


Figure 3.1: The parameters used for the simulation are listed in table 3.1.

S	δ	μ	σ_S	κ	α	σ_δ	ρ	λ
50	-0.1	0	0.3	2	-0.1	0.3	0.3	-0.1

Table 3.1: Parameters for simulation (figure 3.1)

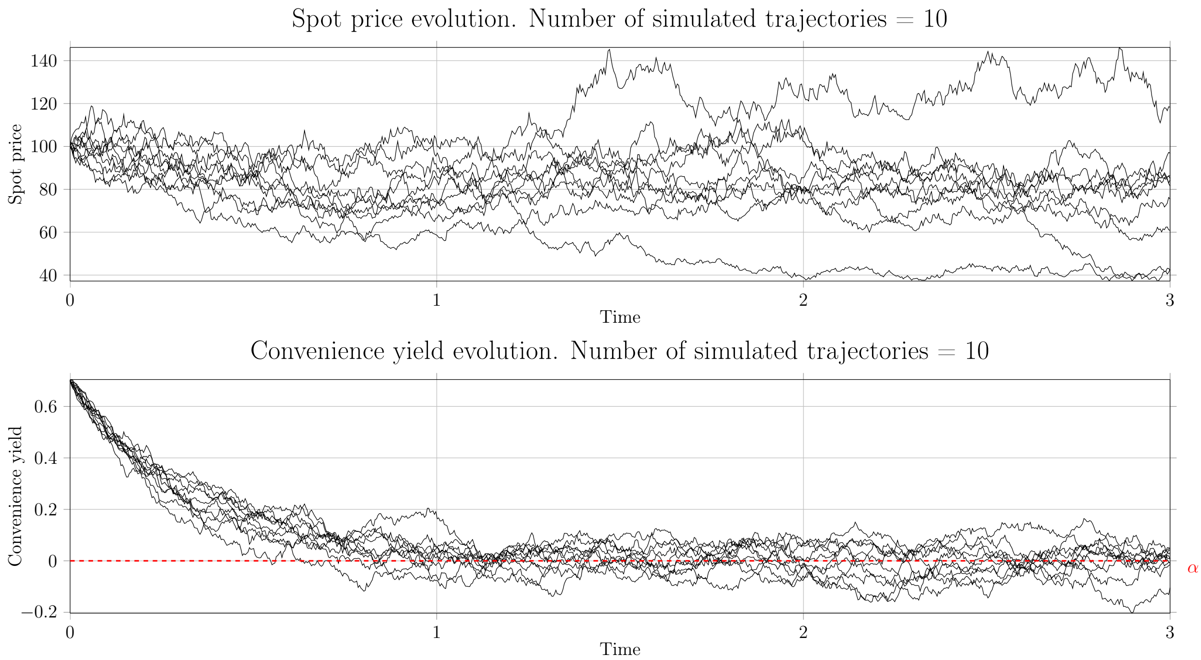


Figure 3.2: The parameters used for the simulation are listed in table 3.2. Notice the mean reverting property of the convenience yield, which becomes exaggeratedly clear when the values of δ and α are far apart.

S	δ	μ	σ_S	κ	α	σ_δ	ρ	λ
100	0.7	0	0.3	3	0	0.15	0.6	-0.5

Table 3.2: Parameters for simulation (figure 3.2)

A point that the figures above prove, is that not all the parameters in our pricing model have an intuitive translation to real world phenomena. Indeed, the long term mean α, the spot drift μ and the diffusion (or volatility) parameters have an intuitive influence on the model, and can be roughly estimated from historic data with little effort. For example, it is usually feasible to assume that the long term mean parameter α should not divert too much from the sample mean (within a reasonable time frame) of the convenience yield. The diffusion parameters σ_S and σ_δ should also roughly correspond to the historic volatility of the futures price and convenience yield, respectively.

However, the influence of κ, ρ and especially λ, on the model can be quite surprising, and even counter-intuitive. The mean-reversion coefficient κ, for example, can seem artificial and misleading for commodities with extreme seasonal fluctuations (for example crops). The process of estimating the parameters of the model will be covered in the next chapter.

Chapter 4

Parameter estimation

At this point we have been able to establish a stochastic model on the dynamics of commodity spot prices in chapter 3, and also derived the pricing formula of commodity futures within our framework in section 3.1.1. We are, however, only halfway in achieving our goal in being able to accurately simulate the price dynamics based on historic data. Indeed, as seen in section 3.2.2, we possess the required tools for the simulation process itself, *given the parameter-values in our model.* We have not yet addressed the issue of acquiring the system parameters from historic data. This issue will be addressed in this chapter.

We will begin by presenting the state space form; a specific representation of a time series model. Many sophisticated smoothing, filtering and prediction algorithms are applicable for models in the state space form, which makes the state space form a compelling option of presenting different models. The Schwartz two-factor model, as presented up until now, is not in a state space form. Fortunately, we are able to make a transformation from the previously presented form of our framework to a state space form.

We will proceed by presenting a multi-purpose recursive algorithm that can be used for smoothing, filtering, prediction and parameter fitting, namely, the Kalman filter. The Kalman filter may by applied to models in the state space form, given that the model complies with some distributional conditions. We will derive the Kalman filter algorithm, and further show that our model complies with all the requirements prior to the filtering.

We will finally describe the system parameter fitting algorithm for our model. The core of the fitting procedure relies on numerically maximizing the log-likelihood of the parameter-set utilizing the Kalman filter. The optimizing will be, perhaps not so surprisingly given the size of our parameter-set, computationally very expensive and unstable, to which we will also introduce some counteractive measures.

4.1 State space form

We will begin by presenting the state space representation, after which we will interpret and break down the different aspects of the form.

A model is in a state space form if it can be expressed as

$$\boldsymbol{Y}_t = \boldsymbol{Z}_t \boldsymbol{X}_t + \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t \tag{4.1}$$

$$\boldsymbol{X}_t = \boldsymbol{T}_t \boldsymbol{X}_{t-1} + \boldsymbol{c}_t + \boldsymbol{\tau}_t, \tag{4.2}$$

where

$$\begin{aligned}
\boldsymbol{Y}_t &\in \mathbb{R}^{n\times 1} \\
\boldsymbol{Z}_t &\in \mathbb{R}^{n\times m} \\
\boldsymbol{d}_t &\in \mathbb{R}^{n\times 1} \\
\boldsymbol{X}_t &\in \mathbb{R}^{m\times 1} \\
\boldsymbol{T}_t &\in \mathbb{R}^{m\times m} \\
\boldsymbol{c}_t &\in \mathbb{R}^{m\times 1} \\
\boldsymbol{\varepsilon}_t &\sim N(\boldsymbol{0}, \boldsymbol{H}_t), \text{ where } \boldsymbol{H}_t \in \mathbb{R}^{n\times n}, \text{ and } \mathrm{Corr}_{\mathbb{P}}\left(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}\right) = 0 \\
\boldsymbol{\tau}_t &\sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \text{ where } \boldsymbol{Q}_t \in \mathbb{R}^{n\times n}, \text{ and } \mathrm{Corr}_{\mathbb{P}}\left(\boldsymbol{Q}_t, \boldsymbol{Q}_{t-1}\right) = 0 \\
t &\in \{1, \ldots, T\}
\end{aligned} \tag{4.3}$$

The representation of the vector $\boldsymbol{Y}_t$ in the equation (4.1) is called the *measurement equation*. The elements of $\boldsymbol{Y}_t$ are observations of the modelled phenomena. We are essentially assuming, that these observations are generated from a linear combination of the vector $\boldsymbol{X}_t$. The observations are also assumed to be noisy, noise originating from the $\boldsymbol{\varepsilon}_t$ vector. That is, we assume that the observations contain some error, for example due to inaccuracies in the measurement equipment.

Equation (4.2) is called the *transition equation*. From the expression of the equation one can observe that $\boldsymbol{X}_t$ is defined as a linear combination if it's previous state, associated again with noise. Furthermore, one can also observe the implicit assumption that $\boldsymbol{X}_t$

possesses the Markov property. The interpretation of the role of $\boldsymbol{X}_t$ is that it contains the model information of the underlying, generally unobservable, model entities. It is the driving force in the model, and in common the target of primary interest.

Notice that we are also enforcing some structure on the noise associated with both the measurement and the transition equation. The properties of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\tau}_t$ in (4.3) imply that the errors are indifferent of the errors from the previous state. The non-correlation combined with the distributional assumption, i.e. that the errors follow a multivariate normal distribution, actually implies that both the model and measurement errors stem from *Gaussian white noise*[11]. Note that even though the measurement and transition noise share the same properties, they are conceptually very different. The measurement errors are can be intuitively justified, but errors in the model data might seem a bit outlandish; why introduce a faulty model? When dealing with stochastic models the errors do, in fact, not derive from model inconsistency, but from the very fundamental aspect of the problem setting. When we are accepting a stochastic model we are implicitly accepting some sort of uncertainty in the model, and the role of the transition error vector (or rather, the covariance matrix $\boldsymbol{Q}_t$) is just to characterize the uncertainty in the model.

The model components $\boldsymbol{Z}_t$, $\boldsymbol{d}_t$, $\boldsymbol{T}_t$, $\boldsymbol{c}_t$, $\boldsymbol{H}_t$ and $\boldsymbol{Q}_t$ are called *system matrices* and are assumed to be deterministic. We do not assume, however, that all the elements in the system matrices are necessarily known. If the system matrices are time-independent, then the model is called *time-invariant*[19]. As we will see in section 4.3.1, the Schwartz two-factor model in the state space form will be time-invariant with system matrices containing unknown model parameters originating from the futures pricing formula.

4.2 Kalman filter

As already pointed out in the introduction of this chapter, the benefit of formulating a state space representation of a model is that a wide array of different optimization routines are applicable for the model. Perhaps the most frequently applied is the Kalman filter. The Kalman filter is a recursive routine for computing the minimum mean square estimator of the state $\boldsymbol{X}_t$ vector at time t, which uses both previous and current information of the state and observations [19]. Different variations of the Kalman filter have been applied over multiple research fields from engineering to mathematical finance[1].

[1] See for example [24, 28, 10, 33, 37].

The Kalman filter is a multi-purpose algorithm in the sense that it not only generates an optimal estimator of the state vector, but also provides means to produce comparative data of the likelihood of the model in question. This enables the implementation of different parameter fitting routines subjected to numerical optimization. This specific usage is in some cases referred to as *the prediction error decomposition* algorithm, and is a central component in our parameter fitting routine.

In this section we will give an explicit form of the recursion equations of the Kalman filter and the prediction decomposition algorithm, and also outline the derivation of both respectively. We will first present a few contributive results and definitions.

Recall the definition of mean square error.

Definition 4.4. *(Mean square error) Let $\boldsymbol{X} \in \mathbb{R}^{n\times 1}$ be a multivariate random variable $\hat{\boldsymbol{X}} \in \mathbb{R}^{n\times 1}$ an estimator of $\boldsymbol{X}$. The mean square error of the estimator is then defined as*

$$\mathrm{MSE}_{\mathbb{P}}\left(\hat{\boldsymbol{X}}\right) = \mathbb{E}_{\mathbb{P}}\left[\left(\hat{\boldsymbol{X}} - \boldsymbol{X}\right)\left(\hat{\boldsymbol{X}} - \boldsymbol{X}\right)^T\right]. \tag{4.5}$$

The *minimum mean square error estimator* (usually referred as the *MMSE estimator*) is then simply defined as

$$\mathrm{MMSE}_{\mathbb{P}}\left(\boldsymbol{X}\right) = \arg\min_{\hat{\boldsymbol{X}}}\left(\mathrm{MSE}_{\mathbb{P}}\left(\hat{\boldsymbol{X}}\right)\right). \tag{4.6}$$

Theorem 4.7. *Assume a setting similar in definition 4.4. Furthermore, assume a set of measurements $\boldsymbol{Y} \in \mathbb{R}^{m\times 1}$. The MMSE estimator, given the observations, is*

$$\mathrm{MMSE}_{\mathbb{P}}\left(\boldsymbol{X}\right) \equiv \mathrm{MMSE}_{\mathbb{P}}\left(\boldsymbol{X};\boldsymbol{Y}\right) = \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{X}|\boldsymbol{Y}\right) \tag{4.8}$$

Proof. See [23]. □

The foundation of the Kalman filter is actually the MMSE estimator; the recursive process determines the MMSE estimator of the state vector given the observations, and updates the estimator accordingly when new observations become available. However, we need to make an additional assumption in addition to the assumption of a model in a state space representation. Namely, the state vector is assumed to be normally distributed, enabling the use of the properties of the normal distribution.

Theorem 4.9. *(The Kalman filter)*

Assume the notation in section 4.1.

Let $\hat{\boldsymbol{X}}_{t|t-1}$ *denote the MMSE estimator of* $\boldsymbol{X}_t$ *of observations* $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{t-1} =: \boldsymbol{Y}_{0:t-1}$*, that is,* $\hat{\boldsymbol{X}}_{t|t-1} = \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{X}_t | \boldsymbol{Y}_{0:t-1}\right)$*. Similarily, denote* $\boldsymbol{P}_{t|t-1} = \mathrm{Cov}_{\mathbb{P}}\left(\boldsymbol{X}_t | \boldsymbol{Y}_{0:t-1}\right)$*, that is, the error covariation matrix.*

Furthermore, assume that $\boldsymbol{X}_t$ *is normally distributed. Then, the following identities hold.*

$$
\begin{aligned}
\hat{\boldsymbol{X}}_{t|t-1} &= \boldsymbol{T}_t \hat{\boldsymbol{X}}_{t-1|t-1} + \boldsymbol{c}_t \\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{T}_t^T + \boldsymbol{Q}_t
\end{aligned} \tag{4.10}
$$

$$
\begin{aligned}
\hat{\boldsymbol{X}}_{t|t} &= \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{P}_{t|t-1} + \boldsymbol{Z}_t^T \boldsymbol{F}_t^{-1} \left(\boldsymbol{Y}_t - \boldsymbol{Z}_t \hat{\boldsymbol{X}}_{t|t-1} - \boldsymbol{d}_t\right) \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1}\left(\boldsymbol{I} - \boldsymbol{Z}_t^T \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t \boldsymbol{P}_{t|t-1}\right)
\end{aligned} \tag{4.11}
$$

where $\boldsymbol{F}_t = \boldsymbol{Z}_t \boldsymbol{P}_{t|t-1} \boldsymbol{Z}_t^T + \boldsymbol{H}_t$.

Equations (4.10) are known as the *prediction equations*, and (4.11) as the *updating equations*. We will now present an outline of the proof of the Kalman filter. For a complete proof, the presented framework is easily extendable through induction.

Consider the case of $t = 1$. From the definition of the model, we see that

$$\boldsymbol{X}_1 = \boldsymbol{T}_1 \boldsymbol{X}_0 + \boldsymbol{c}_1 + \boldsymbol{\tau}_1,$$

and therefore

$$
\begin{aligned}
\hat{\boldsymbol{X}}_{1|0} &= \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{X}_1 | \boldsymbol{Y}_0\right) = \boldsymbol{T}_1 \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{X}_0 | \boldsymbol{Y}_0\right) + \boldsymbol{c}_1 + \underbrace{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{\tau}_1 | \boldsymbol{Y}_0\right)}_{\perp\!\!\!\perp\ 0} \\
&= \boldsymbol{T}_1 \hat{\boldsymbol{X}}_0 + \boldsymbol{c}_1,
\end{aligned} \tag{4.12}
$$

and

$$
\begin{aligned}
\boldsymbol{P}_{1|0} &= \mathrm{Cov}_{\mathbb{P}}\left(\boldsymbol{X}_1 | \boldsymbol{Y}_0\right) = \boldsymbol{T}_1 \mathrm{Cov}_{\mathbb{P}}\left(\boldsymbol{X}_0 | \boldsymbol{Y}_0\right) \boldsymbol{T}_1^T + \mathrm{Cov}_{\mathbb{P}}\left(\boldsymbol{\varepsilon}_0 | \boldsymbol{Y}_0\right) \\
&\overset{\perp\!\!\!\perp}{=} \boldsymbol{T}_1 \boldsymbol{P}_0 \boldsymbol{T}_1^T + \boldsymbol{Q}_1
\end{aligned} \tag{4.13}
$$

Recall that the disturbances in no way dependent of the measurements or model values respectively.

Equations (4.12) and (4.13) prove the prediction equations in the case of $t = 1$. For the updating equations we will need the following lemma.

Lemma 4.14. *Assume that random variables X and Y follow a joint normal distribution, that is,*

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_Y & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_X \end{pmatrix} \right]. \tag{4.15}$$

Then

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left(X|Y\right) &= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}\left(Y - \mu_Y\right) \\ \mathrm{Cov}_{\mathbb{P}}\left(X|Y\right) &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} \end{aligned} \tag{4.16}$$

Proof. Begin by defining $Z = X - \Sigma_{XY}\Sigma_Y^{-1}Y$. From the definition of covariance, we see that

$$\begin{aligned} \mathrm{Cov}_{\mathbb{P}}\left(Z, Y\right) &= \mathrm{Cov}_{\mathbb{P}}\left(X, Y\right) - \Sigma_{XY}\Sigma_Y^{-1}\mathrm{Cov}_{\mathbb{P}}\left(Y, Y\right) \\ &= \Sigma_{XY} - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_Y = \Sigma_{XY} - \Sigma_{XY} \\ &= 0, \end{aligned}$$

and by the properties of joint normal distributions, we see that $Z \perp\!\!\!\perp Y$. Hence,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left(X|Y\right) = \mathbb{E}_{\mathbb{P}}\left(Z|Y\right) + \Sigma_{XY}\Sigma_Y^{-1}Y &\overset{\perp\!\!\!\perp}{=} \mu_X - \Sigma_{XY}\Sigma_Y^{-1}\mu_Y + \Sigma_{XY}\Sigma_Y^{-1}Y \\ = \mu_x + \Sigma_{XY}\Sigma_Y^{-1}\left(Y - \mu_Y\right). \end{aligned}$$

The proof is completed by evaluating

$$\begin{aligned} \mathrm{Cov}_{\mathbb{P}}\left(X|Y\right) &= \mathrm{Cov}_{\mathbb{P}}\left(Z + \Sigma_{XY}\Sigma_Y^{-1}Y, Z + \Sigma_{XY}\Sigma_Y^{-1}Y|Y\right) \\ &= \mathrm{Cov}_{\mathbb{P}}\left(Z|Y\right) + \mathrm{Cov}_{\mathbb{P}}\left(\Sigma_{XY}\Sigma_Y^{-1}Y|Y\right) + \mathrm{Cov}_{\mathbb{P}}\left(\Sigma_{XY}\Sigma_Y^{-1}Y, Z|Y\right) \\ &\quad + \mathrm{Cov}_{\mathbb{P}}\left(Z, \Sigma_{XY}\Sigma_Y^{-1}Y|Y\right) \\ &\overset{\perp\!\!\!\perp}{=} \mathrm{Cov}_{\mathbb{P}}\left(Z\right) \\ &= \Sigma_X + \Sigma_{XY}\Sigma_Y^{-1}\Sigma_Y\Sigma_Y^{-1}\Sigma_{YX} - 2\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} \\ &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}. \end{aligned}$$

□

As direct consequence of the normality assumption of $\boldsymbol{X}_0$, and due to the conservation of normality for linear combinations of normally distributed random variables, we simply argue that $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$ follow a joint normal distribution. By evaluating the covariations in (4.17) and (4.18)

$$\operatorname{Cov}_{\mathbb{P}}(\boldsymbol{X}_1, \boldsymbol{Y}_1) = \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{X}_1, \boldsymbol{Z}_1\boldsymbol{X}_1) = \boldsymbol{P}_{1|0}\boldsymbol{Z}_1^T, \tag{4.17}$$

$$\begin{aligned}\operatorname{Cov}_{\mathbb{P}}(\boldsymbol{Y}_1, \boldsymbol{Y}_1) &= \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{Z}_1\boldsymbol{X}_1 + \boldsymbol{\varepsilon}_1, \boldsymbol{Z}_1\boldsymbol{X}_1 + \boldsymbol{\varepsilon}_1) \\ &= \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{Z}_1\boldsymbol{X}_1, \boldsymbol{Z}_1\boldsymbol{X}_1) + \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{Z}_1\boldsymbol{X}_1, \boldsymbol{\varepsilon}_1) \\ &\quad + \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{\varepsilon}_1, \boldsymbol{Z}_1\boldsymbol{X}_1) + \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_1) \\ &\overset{\perp\!\!\!\perp}{=} \boldsymbol{Z}_1\boldsymbol{P}_{1|0}\boldsymbol{Z}_1^T + \boldsymbol{H}_1 =: \boldsymbol{F}_1,\end{aligned} \tag{4.18}$$

we assert that the explicit form of the joint distribution of $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$ is

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{Y}_1 \end{pmatrix} \sim N\left[\begin{pmatrix} \hat{\boldsymbol{X}}_{1|0} \\ \boldsymbol{Z}_1\hat{\boldsymbol{X}}_{1|0} + \boldsymbol{d}_1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{P}_{1|0} & \boldsymbol{P}_{1|0}\boldsymbol{Z}_1^T \\ \boldsymbol{Z}_1\boldsymbol{P}_{1|0} & \boldsymbol{F}_1 \end{pmatrix}\right]. \tag{4.19}$$

Now, applying lemma 4.14 yields

$$\begin{aligned}\hat{\boldsymbol{X}}_{1|1} &= \mathbb{E}_{\mathbb{P}}(\boldsymbol{X}_1|\boldsymbol{Y}_1) = \hat{\boldsymbol{X}}_{1|0} + \boldsymbol{P}_{1|0}\boldsymbol{Z}_1^T\boldsymbol{F}_1^{-1}\left(\boldsymbol{Y}_1 - \boldsymbol{Z}_1\hat{\boldsymbol{X}}_{1|0} - \boldsymbol{d}_1\right) \\ \boldsymbol{P}_{1|1} &= \operatorname{Cov}_{\mathbb{P}}(\boldsymbol{X}_1|\boldsymbol{Y}_1) = \boldsymbol{P}_{1|0}\left(\boldsymbol{I} - \boldsymbol{Z}_1^T\boldsymbol{F}_1^{-1}\boldsymbol{Z}_1\boldsymbol{P}_{1|0}\right),\end{aligned}$$

which concludes the proof in the case of $t = 1$.

The following step, which we will be omitting, is to prove the complete statement by induction. As this is almost solely a technical feat, we are omitting the next step. The technical details can be found for instance in[33].

The filtering scheme is of course valuable to us; it enables us to perform qualitative measurements of our estimations. As explained in [citation needed] in the case of testing our estimation scheme with synthetic data, the relative error measurements of the filtered state vector versus the synthetic truth gives us an additional mean to verify the quality of our estimation scheme. However, our primary interest in the Kalman filter lies in what might be considered as a by-product of the actual filtering routine. Namely, the Gaussian property of the state vector enables us to compute the likelihood of the model. This is an extremely important feature to us, as this gives us a numeric measure of the quality of our model, which is also comparable to the likelihood of other models with different parameter sets from the parameter space. When this is subjected to a numerical optimization scheme, we are able to obtain an optimal parameter set in the sense of the likelihood function. This is, of course, wishful thinking to some extent; for example, the convergence of the optimization procedure is in no way guaranteed, but still provides a feasible starting point.

Theorem 4.20. *(The likelihood obtained from the Kalman filter)*
The log-likelihood of the model is

$$-\frac{1}{2}\left[nT\log(2\pi)+\sum_{k=0}^{T}\det\left(\boldsymbol{F}_k\right)+\sum_{k=0}^{T}\boldsymbol{V}_k^T\boldsymbol{F}_k^{-1}\boldsymbol{V}_k\right] \tag{4.21}$$

where

$$\boldsymbol{V}_t=\boldsymbol{Y}_t-\boldsymbol{Z}_t\hat{\boldsymbol{X}}_{t-1|t-1}+\boldsymbol{d}_t.$$

Proof. See [19]. □

In our implementation the likelihood will, in fact, be computed iteratively. More on the actual implementation and the use of the calculated likelihood follows in section 4.3.

4.3 The parameter estimation scheme applied to the commodity spot price model

In this section we will present the actual implementation of our parameter estimation scheme. We will begin by deriving the state space form of our analytic commodity spot price model, which enables us to utilize the prediction error decomposition algorithm. Next, we will present some customized procedures, or tweaks, if you may for the prediction decomposition algorithm in order to improve the convergence of the process. Finally we will present the full implementation with each step of the procedure, including e.g. the chosen parameter sets and the choice of the optimization routine.

4.3.1 The state space form of the commodity spot price model

A prerequisite of applying the Kalman filter is to show the joint Gaussian property of the state entities, that is, the spot price and the convenience yield. As it turns out, in order to achieve the Gaussian property we will be modelling the *log*-spot price of the commodity. This will not pose any significant problems, as the transformation is achieved from a simple application of Itô's lemmma. The methodology overall will be quite similar to the one used to derive the futures pricing formula in section 3.1.1.

Recall the commodity spot price model defined in chapter 3.

$$\begin{aligned}
dS_t &= (\mu - \delta_t)\, S_t dt + \sigma_S S_t dW_t^S \\
d\delta_t &= \kappa\,(\alpha - \delta_t)\, dt + \sigma_\delta dW_t^\delta \\
S_0 &\equiv S \\
\delta_0 &\equiv \delta
\end{aligned}$$

We will begin by transforming the stochastic differential equation of the spot price to the corresponding one of the log-spot price. For this we will use Itô's lemma on the function $f = \log$. Simple calculus implies that

$$\begin{aligned}
f_{S_t}(S_t) &= \frac{1}{S_t} \\
f_{S_t S_t}(S_t) &= -\frac{1}{S_t^2}.
\end{aligned}$$

By inserting the above to the formula in Itô's lemma, we see that

$$\begin{aligned}
d\log S_t &= \left[(\mu - \delta_t)\, S_t \frac{1}{S_t} - \frac{1}{2}\sigma_S^2 S_t^2 \frac{1}{S_t^2}\right] dt + \sigma_S S_t \frac{1}{S_t} dW_t^S \\
&= \left(\mu - \delta_t - \frac{1}{2}\sigma_S^2\right) dt + \sigma_S dW_t^S
\end{aligned} \tag{4.22}$$

Theorem 4.23. *(The joint distribution of the log-spot price and the convenience yield)*

$$\begin{pmatrix} \log S_t \\ \delta_t \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_S(t) \\ \mu_\delta(t) \end{pmatrix}, \begin{pmatrix} \Sigma_S(t) & \Sigma_{S\delta}(t) \\ \Sigma_{\delta S}(t) & \Sigma_\delta(t) \end{pmatrix}\right], \tag{4.24}$$

where

$$\begin{aligned}
\mu_S(t) &= \log S + \left(\mu - \frac{1}{2}\sigma_S^2 - \alpha\right) t + \frac{\delta - \alpha}{\kappa}\left(1 - e^{-\kappa t}\right) \\
\mu_\delta(t) &= \alpha + e^{-\kappa t}\,(\delta_t - \alpha) \\
\Sigma_S(t) &= \frac{\sigma_\delta^2}{\kappa}\left[\frac{1}{2\kappa}\left(1 - e^{-2\kappa t}\right) - \frac{2}{\kappa}\left(1 - e^{-\kappa t}\right) + t\right] \\
&\quad + 2\frac{\sigma_S \sigma_\delta \rho}{\kappa}\left(\frac{1 - e^{-\kappa t}}{\kappa} - t\right) + \sigma_S^2 t \\
\Sigma_\delta(t) &= \frac{\sigma_\delta^2}{2\kappa}\left(1 - e^{-2\kappa t}\right) \\
\Sigma_{\delta S}(t) = \Sigma_{S\delta}(t) &= \frac{\sigma_\delta}{\kappa}\left[\left(\sigma_S \rho - \frac{\sigma_\delta}{\kappa}\right)\left(1 - e^{-\kappa t}\right) + \sigma_\delta \frac{1 - e^{-2\kappa t}}{2\kappa}\right].
\end{aligned} \tag{4.25}$$

Proof. We begin again by transforming the model to a system with uncorrelated Brownian motions (see for example section 3.2.2).

$$
\begin{aligned}
d\log S_t &= \left(\mu - \delta_t - \frac{1}{2}\sigma_S^2\right) dt + \sigma_S \left(\sqrt{1-\rho^2} dZ_t^1 + \rho dZ_t^2\right) \\
d\delta_t &= \kappa \left(\alpha - \delta_t\right) dt + \sigma_\delta dZ_t^2
\end{aligned}
$$

As a direct consequence of lemma 3.6 we acquire the solution of the convenience yield.

$$
\delta_t = \alpha + e^{-\kappa t}\left(\delta - \alpha\right) + \sigma_\delta \int_0^t e^{\kappa(s-t)} dZ_s^2 \tag{4.26}
$$

The expectation of the Itô integral and the Itô isometry implies that

$$
\mu_\delta(t) = \mathbb{E}_{\mathbb{P}}\left(\delta_t\right) = \alpha + e^{-\kappa t}\left(\delta - \alpha\right), \tag{4.27}
$$

and that

$$
\begin{aligned}
\Sigma_\delta(t) &= \mathbb{E}_{\mathbb{P}}\left(\delta_t^2\right) - \mu_\delta(t)^2 \\
&= \mu_\delta(t)^2 + \mu_\delta(t)\sigma_\delta \underbrace{\mathbb{E}_{\mathbb{P}}\left(\int_0^t e^{k(s-t)} dZ_s^2\right)}_{=0} + \sigma_\delta \mathbb{E}_{\mathbb{P}}\left[\left(\sigma_\delta \int_0^t e^{k(s-t)} dZ_s^2\right)^2\right] \\
&\quad - \mu_\delta(t)^2 \\
&= \mathbb{E}_{\mathbb{P}}\left[\left(\sigma_\delta \int_0^t e^{k(s-t)} dZ_s^2\right)^2\right] = \sigma_\delta^2 \int_0^t e^{2k(s-t)} ds \\
&= \frac{\sigma_\delta^2}{2\kappa}\left(1 - e^{-2\kappa t}\right)
\end{aligned} \tag{4.28}
$$

By expanding the differential equation of the log-spot price to the explicit form, and by inserting the previously evaluated (see equation (3.18)) value of $\int_0^t \delta_s ds$ into the equation, we assert that

$$
\begin{aligned}
\log S_t = \log S &+ \left(\mu - \alpha - \frac{1}{2}\sigma_S^2\right) t + \frac{\delta - \alpha}{\kappa}\left(1 - e^{-\kappa t}\right) \\
&+ \sigma_S \sqrt{1-\rho^2} Z_t^1 + \int_0^t \sigma_s \rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) dZ_s^2.
\end{aligned} \tag{4.29}
$$

Evaluating the rest of the parameters of the joint distribution is now a straightforward task. By recalling the properties of the expectation of the Itô integral, we simply argue that

$$\mu_S(t) = \mathbb{E}_{\mathbb{P}}\left(\log S_t\right) = \log S + \left(\mu - \alpha - \frac{1}{2}\sigma_S^2\right) t + \frac{\delta - \alpha}{\kappa}\left(1 - e^{-\kappa(t-s)}\right). \tag{4.30}$$

By yet again recalling the Itô isometry, we also assert that

$$\begin{aligned}
\Sigma_S(t) &= \mathbb{E}_{\mathbb{P}}\left(\log^2 S_t\right) - \mu_S(t)^2 \\
&= \mu_S(t)^2 + \mu_S(t)\mathbb{E}_{\mathbb{P}}\left(\sigma_S\sqrt{1-\rho^2}Z_t^1 + \int_0^t \sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) dZ_s^2\right) \\
&\quad + \mathbb{E}_{\mathbb{P}}\left[\left(\sigma_S\sqrt{1-\rho^2}Z_t^1 + \int_0^t \sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) dZ_s^2\right)^2\right] - \mu_S(t)^2 \\
&= \mathbb{E}_{\mathbb{P}}\left[\left(\sigma_S^2\sqrt{1-\rho^2}Z_t^1\right)^2\right] + \mathbb{E}_{\mathbb{P}}\left[\left(\int_0^t \sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) dZ_s^2\right)^2\right] \\
&\quad + \underbrace{\mathbb{E}_{\mathbb{P}}\left[\left(\int_0^t \sigma_S^2\sqrt{1-\rho^2}Z_s^1\right)\left(\int_0^t \sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) dZ_s^2\right)\right]}_{=0 \text{ (recall that the Brownian motions are uncorrelated)}} \\
&= \sigma_S^2\left(1-\rho^2\right)t + \int_0^t \left[\sigma_S\rho - \frac{\sigma_\delta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right)\right]^2 ds \\
&= \sigma_S^2\left(1-\rho^2\right)t + \sigma_S^2\rho^2 t - 2\frac{\sigma_S\sigma_\delta\rho}{\kappa}t + \frac{\sigma_S\sigma_\delta\rho}{\kappa}\int_0^t e^{-\kappa(t-s)}ds \\
+ &\qquad \frac{\sigma_\delta^2}{\kappa^2}\left(t - 2\int_0^t e^{-\kappa(t-s)}ds + \int_0^t e^{-2\kappa(t-s)}ds\right) \\
&= \frac{\sigma_\delta^2}{\kappa}\left[\frac{1}{2\kappa}\left(1 - e^{-2\kappa t}\right) - \frac{2}{\kappa}\left(1 - e^{-\kappa t}\right) + t\right] + 2\frac{\sigma_S\sigma_\delta\rho}{\kappa}\left(\frac{1 - e^{-\kappa t}}{\kappa} - t\right) + \sigma_S^2 t.
\end{aligned} \tag{4.31}$$

A property that remains unproven is the expression of the covariance of the log-spot price and the convenience yield. From the definition of the covariance we get

$$\begin{aligned}
\Sigma_{S\delta}(t) &= \Sigma_{\delta S}(t) = \mathbb{E}_{\mathbb{P}}\left[(\log S_t - \mu_S(t))\,(\delta_t - \mu_\delta(t))\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[\left(\sigma_S\sqrt{1-\rho^2}Z_t^1 + \int_0^t \sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1-e^{-\kappa(t-s)}\right)dZ_s^2\right)\left(\sigma_\delta\int_0^t e^{k(s-t)}dZ_s^2\right)\right] \\
&= \sigma_\delta\mathbb{E}_{\mathbb{P}}\left[\int_0^t \left(e^{-k(t-s)}\right)\left(\sigma_s\rho - \frac{\sigma_\delta}{\kappa}\left(1-e^{-\kappa(t-s)}\right)\right)\left(dZ_t^2\right)^2\right] \\
&= \sigma_\delta\left(\sigma_S\rho\int_0^t e^{-\kappa(t-s)}ds - \frac{\sigma_\delta}{\kappa}\int_0^t e^{-\kappa(t-s)}ds + \frac{\sigma_\delta}{\kappa}\int_0^t e^{-2\kappa(t-s)}ds\right) \\
&= \frac{\sigma_\delta}{\kappa}\left[\left(\sigma_S\rho - \frac{\sigma_\delta}{\kappa}\right)\left(1-e^{-\kappa t}\right) + \sigma_\delta\frac{1-e^{-2\kappa t}}{2\kappa}\right].
\end{aligned} \tag{4.32}$$

Combining equations (4.27), (4.28), (4.30), (4.31) and (4.32) prove the parameters of the joint distribution. By concluding that, as a linear combination of normally distributed random variables is normally distributed, and that the Itô integrals are normally distributed, the joint distribution is Gaussian completes the proof. □

All prerequisites to forming the state space representation of our commodity spot price model and applying the Kalman filter on the transformed model are now complete, and we will now construct the state space form of our model. As we have already worked on the stochastic model in this chapter, we will begin with the transition equation. Note that throughout the rest of this thesis we will be working with evenly distributed time partitions. More precisely, if we are working on a time domain $[0, T]$ with time-steps $\{t_k \in [0, T] \mid k \in \{0, \dots, n\}\}$, the size of the time-steps will be equal, that is,

$$\Delta t_k \equiv \Delta t$$

for all $k \in \{0, \dots, n\}$.

Technically, this is not a complete resemblance of the reality, as we have incompleteness in the data due to holidays and weekends etc. These gaps are, however, negligible, as we pre-process the data and either remove these gaps or fill them with constant values from the previous time-step.

The transition equation

As pointed out on multiple previous occasions, we assume that the increments log-spot price $\log S_t$ is directly dependant of both the convenience and the previous state of the log-spot price (see for example equation (4.29)). Increments of the convenience yield, on the other hand, depend only on the previous value of the convenience yield. Hence, these properties need to be extracted from the system in order to form the transition equation. In order to complete the definition of the system, we see that the transition equation, excluding the impact of the Brownian motion increments for now, needs to be of the form

$$
\begin{aligned}
\boldsymbol{X}_{t_k} &= \begin{pmatrix} \log S_{t_k} \\ \delta_{t_k} \end{pmatrix} \\
\boldsymbol{d}_{t_k} &= \begin{pmatrix} \left(\mu - \alpha - \frac{1}{2}\sigma_S^2\right)\Delta t - \frac{\alpha}{\kappa}\left(1 - e^{-k\Delta t}\right) \\ \alpha\left(1 - e^{-\kappa\Delta t}\right) \end{pmatrix} \\
\boldsymbol{T}_{t_k} &= \begin{pmatrix} 1 & \frac{1}{\kappa}\left(1 - e^{\kappa\Delta t}\right) \\ 0 & e^{-\kappa\Delta t} \end{pmatrix},
\end{aligned}
\tag{4.33}
$$

which can be verified by expanding the expression of the transition equation (with the above values) and comparing the results against equations (4.29) and (4.26). Note that we are following the naming convention established in section 4.1.

The expressions in (4.33) account for the deterministic part of the model. The probabilistic properties, that stem from the Brownian motion increments, is yet to be incorporated into the state space version of our model. The representation framework does not allow an explicit inclusion of these properties. However, as discussed in section 4.1, we can still implicitly incorporate the effect of the Brownian motions by applying the properties of the joint distribution to the error covariance matrix $\boldsymbol{Q}_t$. Recall that the 'errors' in the transition equation are not to be interpreted as something akin to measurement errors, but rather as a descriptive factor of the probabilistic properties in the model. Hence it is natural to define the covariance matrices of the error components and the joint distribution as equal, namely

$$
\boldsymbol{Q}_t = \begin{pmatrix} \Sigma_S(t) & \Sigma_{S\delta}(t) \\ \Sigma_{\delta S}(t) & \Sigma_\delta(t) \end{pmatrix}. \tag{4.34}
$$

Equations (4.33) and (4.34) completely defines the transition equation of our commodity spot price model.

The measurement equation

Recall the pricing formula commodity futures defined in (3.27), namely

$$F(S, \delta, T) = S \exp\left[-\frac{\delta\left(1 - e^{-\kappa T}\right)}{\kappa} + A(T)\right].$$

The log-price of the futures is then

$$\log F(S, \delta, T) = \log S - \frac{\delta\left(1 - e^{-\kappa T}\right)}{\kappa} + A(T). \tag{4.35}$$

For futures on the commodity with maturities $\{T_k\}_{k\in\{1,\dots,n\}}$, the above equation, in matrix form, translates to

$$\begin{pmatrix} \log F(S, \delta, T_1) \\ \vdots \\ \log F(S, \delta, T_n) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1-e^{-\kappa T_1}}{\kappa} \\ \vdots & \vdots \\ 1 & -\frac{1-e^{-\kappa T_n}}{\kappa} \end{pmatrix} \begin{pmatrix} \log S \\ \delta \end{pmatrix} + \begin{pmatrix} A(T_1) \\ \vdots \\ A(T_n) \end{pmatrix}.$$

Let $\hat{F}_t(T_k)$ denote an observed futures price at time t with T_k as the length to maturity. By applying the same logic as in the matrix transformation for the futures price formula for multiple contracts, we see that

$$\begin{aligned} \boldsymbol{Y}_t &= \begin{pmatrix} \log \hat{F}_t(T_1) \\ \vdots \\ \log \hat{F}_t(T_n) \end{pmatrix} \\ \boldsymbol{Z}_t &= \begin{pmatrix} 1 & -\frac{1-e^{-\kappa T_1}}{\kappa} \\ \vdots & \vdots \\ 1 & -\frac{1-e^{-\kappa T_n}}{\kappa} \end{pmatrix} \\ \boldsymbol{c}_t &= \begin{pmatrix} A(T_1) \\ \vdots \\ A(T_n) \end{pmatrix}. \end{aligned} \tag{4.36}$$

While the above derivation of the system matrices is, given the pricing formula, quite effortless, determining the properties of the observation errors is far from trivial. We do not have any measuring equipment per se which could be tested for e.g. typical error ranges, and therefore evaluating the total noise as an accumulation of for example market uncertainty or noise traders[8] is very difficult and way beyond the scope of this thesis. We will therefore make a simple assumption of observation errors, namely that

$$\boldsymbol{H}_t = \begin{pmatrix} h_1 & 0 & 0 & \cdots & 0 \\ 0 & h_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & h_n \end{pmatrix}. \tag{4.37}$$

In other words, we are assuming that the measurement noise of different contracts are independent of each other.

In practice we have actually found that, given the lack of proper tools to measure market noise, it is actually better to use an even simpler form of the error covariations, namely

$$\boldsymbol{H}_t = h_t I_n \tag{4.38}$$

where $h_t \in R_+$ and I_n is the identity matrix of size n. This actually allows us to incorporate the error calibration in the optimization procedure. This mimics the trial and error approach of just testing different combinations of error variances based on prior assumptions.

The equations (4.36), (4.37) and (4.38) completely define the measurement equation, which concludes our definition of the commodity spot price model in the state space form.

4.3.2 On the estimation of distinct parameter subsets

As mentioned in the introduction to this chapter we will be ultimately using a numerical optimization routine to maximize the log-likelihood obtained from the prediction error decomposition algorithm. The optimization routines require, in addition to the objective function, also a starting point for the algorithm (an initial guess, if you may). This starting point plays an essential part in cases where the mapping of the objective function contains multiple local minima and maxima, as it is more likely for the optimization routine to find local extreme values closer to the initial guess opposed to other, perhaps global, extreme values further away from the initial guess. This is especially important where the parameter space is of higher dimension, which is the case in our study. We will therefore pre-process the data in order to find suitable starting points for the optimization algorithm. The pre-processing routines are presented in this section.

We will begin with the convenience yield. As shown in section 4.3.1, Ornstein-Uhlenbeck processes follow a normal distribution. This opens up opportunities for likelihood optimization of the Ornstein-Uhlenbeck processes, equivalent to the methodology we are using to estimate the parameters for our commodity spot price model. The likelihood maximization has previously been subject to research, for example in [14] the max-likelihood is implicitly derived as an optimization of a one-dimensional objective function. However, as we are already forced to utilize a great amount of computing power, this does not seem as an effective solution. Therefore we will be using an explicit solution to the max-likelihood of the discretized Ornstein-Uhlenbeck process derived in [38].

Theorem 4.39. *Let X_t be an Ornstein-Uhlenbeck process of the form*

$$dX_t = \kappa\left(\alpha - X_t\right)dt + \sigma dW_t.$$

Applying the gaussian property of X_t yields a discretized Ornstein-Uhlenbeck of the form

$$X_{t_{k+1}} = \alpha + e^{-\kappa\Delta t}\left(\alpha - X_{t_k}\right) + \sigma\sqrt{\frac{1-e^{-2\kappa\Delta t}}{2\kappa}}Z,$$

where $Z \sim N(0,1)$ and the time partition is equivalent with the one in the state space commodity spot price model. The max-likelihood parameters for the discretized model are then

$$\begin{aligned}
\alpha &= \frac{S_y S_{xx} - S_x S_{xy}}{n\left(S_{xx} - S_{xy}\right) - S_x^2\left(S_x S_y\right)} \\
\kappa &= -\frac{1}{\Delta t}\log\left[\frac{S_{xy} + \alpha\left(n\alpha - S_x - S_y\right)}{S_{xx} - \alpha\left(n\alpha - 2S_x\right)}\right] \\
\sigma &= \sqrt{\frac{2\kappa\left[S_{yy} - 2\phi S_{xy} + \phi^2 S_{xx} - 2\alpha\left(1-\phi\right)\left(S_y - \phi S_x\right) + n\alpha^2\left(1-\phi\right)^2\right]}{n\left(1-\phi^2\right)}}
\end{aligned} \tag{4.40}$$

where

$$\begin{aligned}
\phi &= e^{-\kappa \Delta t} \\
S_x &= \sum_{k=0}^{n-1} X_{t_k} \\
S_{xx} &= \sum_{k=0}^{n-1} X_{t_k}^2 \\
S_y &= \sum_{k=0}^{n-1} X_{t_{k+1}} \\
S_{yy} &= \sum_{k=0}^{n-1} X_{t_{k+1}}^2 \\
S_{xy} &= \sum_{k=0}^{n-1} X_{t_{k+1}} X_{t_k}.
\end{aligned} \tag{4.41}$$

Even though the convenience yield in our model can be considered more like an extension of the usual Ornstein-Uhlenbeck, as the increments in the Brownian motion are partially driven by the commodity spot price process through the correlation, we observed that the results from the above theorem produce valid baseline estimates for the parameters in the convenience yield process.

As for the parameters in the spot price process, we will be using maximum likelihood estimates of geometric Brownian motion. This approach is perhaps even more controversial, as not only is the increments of the Brownian motion partially driven by the increments of the convenience yield due to the correlation, but the drift of the process is also corrected with the convenience yield. However, similar to the estimates on the convenience yield parameters, this is not a major concern as we only need some guidelines for the optimization routine. We will be using the maximum likelihood estimates derived in [4].

Theorem 4.42. *Let X_t be a geometric Brownian motion of the form*

$$dX_t = \mu X_t dt + \sigma X_t dW_t.$$

The maximum likelihood estimators of the parameters of X_t are then

$$\begin{aligned}\hat{\sigma} &= \frac{std(L)}{\sqrt{\Delta t}} \\ \hat{\mu} &= \frac{\mathbb{E}\,(L)}{\Delta t} + \frac{1}{2}\hat{\sigma}^2\end{aligned} \tag{4.43}$$

where

$$L = \left(\log F_{t_{k+1}} - \log F_{t_k}\right)_{k\in\{1,\dots,n\}} \in R^{1\times n-1}$$

is a vector of the log-returns of the observed commodity futures prices and $\mathbb{E}(\cdot)$ *denotes the sample mean.*

4.3.3 A naïve dimension reduction

As previously mentioned, a multidimensional optimization procedure requires a great computational effort in order to achieve acceptable convergence conditions. In the previous section we presented counteractive measures for the convergence issue by calculating optimal starting points for the optimization routine. Beyond these measures, we have implemented something that could be called naïve dimension reduction; we will divide the estimation into estimating different parameter subsets and then combine the results. A possible parameter set structure is presented in the listing (4.1) below.

Listing 4.1: A set structure with two parameter sets

```
{
        "parameter-sets": {
                "set-1": [
                        {μ, σS, ρ, λ},
                        {κ, α, σδ}
                ],
                "set-2": [
                        {μ, σS, κ, α, σδ, ρ, λ}
                ]
        }
}
```

The first parameter is separated into two subsets, where the parameters of the convenience yield process are isolated from the rest of the sets. In the optimization procedure, we would initially keep the convenience yield parameters constant (on par with the initial guess) and only after the first optimization run is complete would the convenience yield parameters be subject to the optimization routine.

The second parameter set corresponds to the usual setting, where parameters are optimized simultaneously.

4.3.4 The complete estimation scheme

Finally, we will summarize the contents in this chapter and present our complete estimation scheme through pseudo-code. The code might be somewhat MATLAB influential, as all analysis conducted in this thesis is based on MATLAB implementations, but should be readable without any prior knowledge of MATLAB syntax.

The full scheme is presented in algorithm 1, and the the prediction error decomposition routine is presented in algorithm 2.

Algorithm 1 Preseting the commodity spot price model parameter estimation scheme. F - futures contract data, CY - convenience yield calculated for futures.

procedure ESTIMATE_PARAMS(F, CY)
    spot_appr ← GET_CONTRACTS_WITH_SHORTEST_TTM(F)
    $\{\mu, \sigma_S, \kappa, \alpha, \sigma_\delta, \rho, \lambda\}$ ← CREATE_INIT_GUESS(*spot_appr*, CY)
    param_sets ← CREATE_PARAM_SETS()
    optim_candidates ← $\{\}$
    for each *set* in *param_sets* **do**
        current_params ← $\{\mu, \sigma_S, \kappa, \alpha, \sigma_\delta, \rho, \lambda\}$
        for each *subset* in *set* **do**
            current_params ← PED_ESTIMATE(*current_params*, *subset*, F, CY)
        end for
        optim_candidates.APPEND(*current_params*)
    end for
    return GET_OPTIMAL_PARAM_SET(*optim_candidates*, *spot_appr*, CY)
end procedure

The algorithm above returns the optimal parameter set from the different candidates that were estimated. The optimality of a parameter set is determined by two conditions

I) The relative error between the filtered state variables and the futures price and convenience yield of the futures with the shortest time to maturity.

- Note that (in the absence of arbitrage opportunities) the futures price converges towards the commodity spot price as the time to maturity tends to zero. This property is therefore useful as a qualitative measure of our estimation.

II) The difference in the mean standard deviation of the futures price and convenience yield of the futures with the shortest time to maturity and simulated trajectories.

The optimization of the objective function in algorithm 2 is calculated utilizing the Nelder-Mead algorithm[30], which is implemented in MATLAB as `fminsearch`.

Algorithm 2 The prediction error decomposition routine for parameter estimation.

procedure PED_ESTIMATE($current\_params, subset, F, CY$)
    $constant\_params \leftarrow current\_params$.COMPLEMENT($subset$)
    function OBJECTIVE_FUN($objective\_params$)
        $all\_params \leftarrow objective\_params$.UNION($constant\_params$)
        $\{\boldsymbol{Y}_t, \boldsymbol{d}_t, \boldsymbol{Z}_t, \boldsymbol{H}_t, \boldsymbol{c}_t, \boldsymbol{T}_t, \boldsymbol{Q}_t\} \leftarrow$ STATE_SPACE(F, CY, all_params)
        $initial\_state \leftarrow [\log(F(1,1)), CY(1,1)]$
        $loglike \leftarrow$ KALMAN_FILTER($initial\_state, \boldsymbol{Y}_t, \boldsymbol{d}_t, \boldsymbol{Z}_t, \boldsymbol{H}_t, \boldsymbol{c}_t, \boldsymbol{T}_t, \boldsymbol{Q}_t$)
        return $-loglike$
    end function
    return $optimize$(@OBJECTIVE_FUN, $current\_params$)
end procedure

Chapter 5

Results

In this chapter we will be presenting the main results of this thesis.

We will begin by presenting estimation results acquired from running the estimation routine on synthetic data. Working with synthetic data is important in many aspects, including debugging, model verification and calibration. Calibrating a model with actual data is seldom conducive, as working with actual data is often very challenging even with an optimized procedure. With synthetic data one is able to work in a controlled environment where the number of free variables can be reduced. Also, it is generally effortless to create synthetic (noisy) data that resembles the actual target data.

In section 5.2 we will present estimation results on Brent oil futures contract data. This specific case is used to demonstrate the effectiveness of the procedure on actual data, and is a typical case study of the portfolio sample basket consisting of 26 different commodity futures.

Finally, in section 5.1 we will present results of the distributions of the returns of the simulated futures trajectories.

5.1 Synthetic data - a proof of concept

In this section we will present a typical case study with synthetic data. The data consists of spot price data, convenience yield data and futures contract data of four futures with different time to maturity. The data was generated for a three year period in business days, a total of $3 \cdot 260$ datapoints for each dataset. The spot price and convenience yield data were simulated utilizing the Milstein discretization of our stochastic model of the spot price dynamics, in similar fashion to the simulations presented in figures 3.1 and 3.2. The futures contract prices were generated with the futures pricing formula (see (3.27)) added with Gaussian noise with variance relative to the magnitude of the price data. The relation of the contracts in the terms of the maturity lengths are presented in figure 5.1.

The timeline of the simulated data is somewhat arbitrary, a fixed year for the simulations was decided for only illustrative purposes.

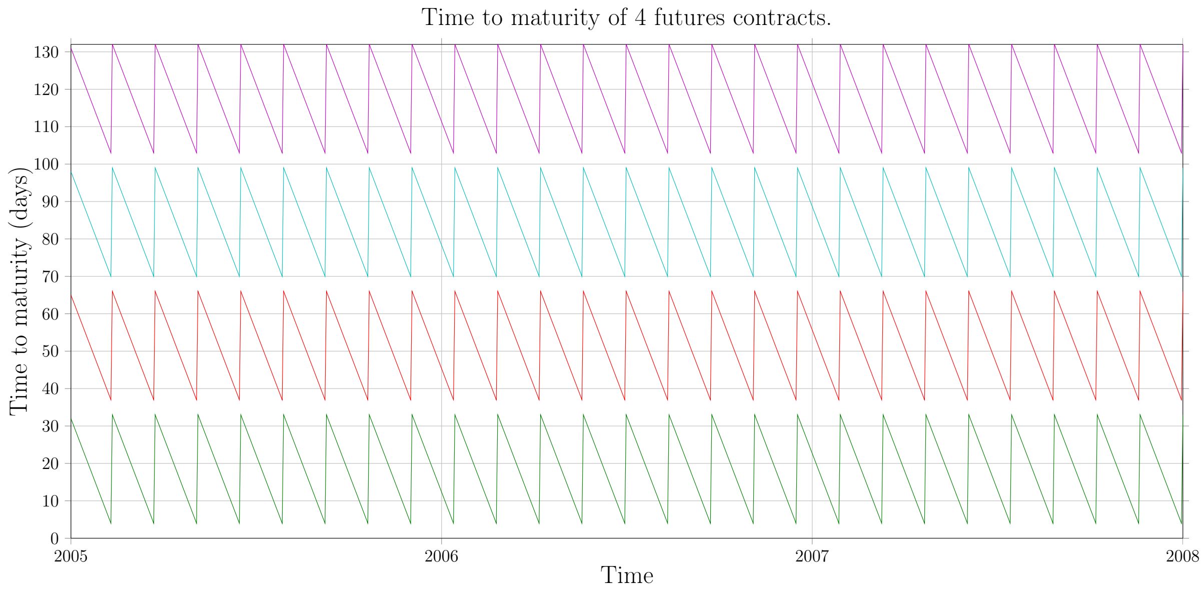


Figure 5.1: The time to maturities over the four different simulated futures contracts. Note that the sawtooth-like shape of the maturities resembles the reality quite well, as there is only a limited amounts of contracts on the market at a time.

Figure 5.2 illustrates the simulated data. Note that the dashed and colored lines present the term structure of the simulated futures. The price data for the contracts is simulated for all timesteps, we have only limited the amount of displayed term structure curves to a weekly basis for illustrative purposes. Note how the noise in the term structure is clearly visible; if no error was applied one would expect to see smooth exponential curves instead of the various irregularities visible in the term structure in the spot price graph.

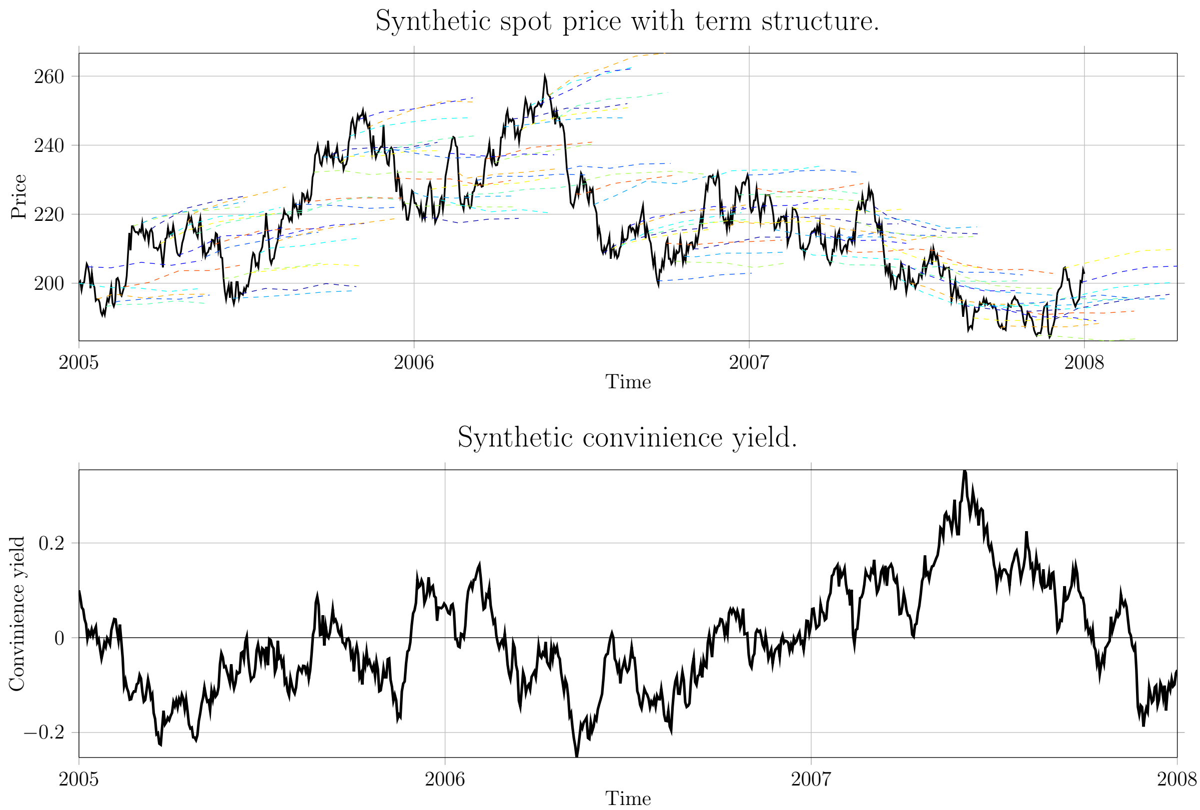


Figure 5.2: The simulated data. The dashed lines in the spot price plot resemble the term structure at the time-step in question (recall the definition of the futures curves given in section 2.4). That is, each dashed is composed of the four contract prices (shifted on the time-axis according to the time to maturity) and a linear interpolation between these data points.

The estimation procedure was run with a relative tolerance threshold at 10^{-7} and with a maximum 5000 iterations for the optimization algorithm. The trace of the optimization algorithm is visualized in figures 5.4 and 5.3. Figure 5.3 shows the evolution of the objective function subject to the optimization routine, while figure 5.4 presents the movement of the optimization routine in the parameter space. The trace shows that even with an effectively small relative tolerance threshold, the convergence is quite towards the threshold limit is quite fast. However, approximately 250 iterations close to the final parameter value indicates that further decreasing the threshold limit is not necessary.

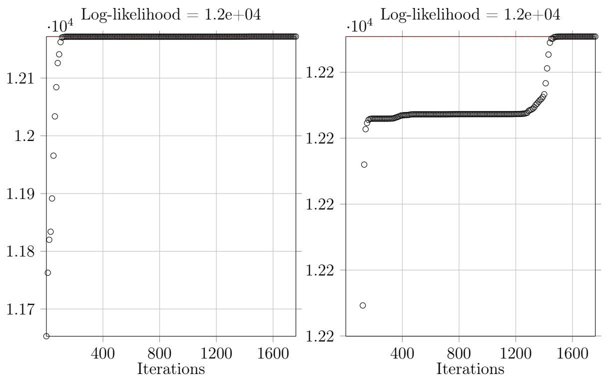


Figure 5.3: Trace of the objective function subject to the optimization routine of the simulated data.

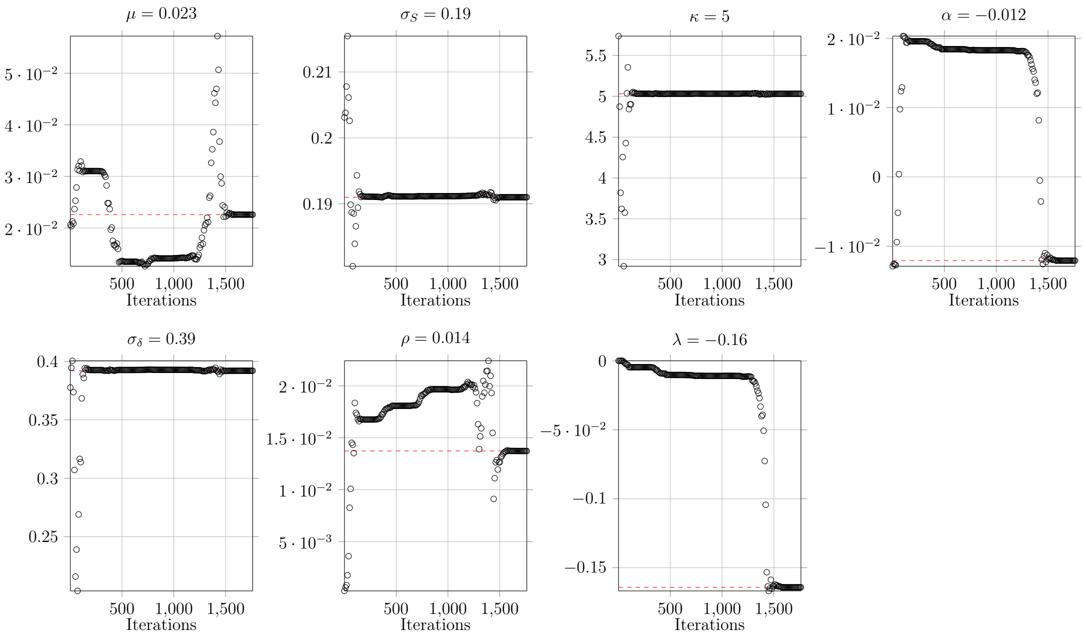


Figure 5.4: Trace of the optimization routine of the simulated data in the parameter space.

It is noteworthy how the initial guess provided to the optimization routine is astonishingly close to the parameter estimation result. This indicates that our actions in section

4.3.2 towards a more stable estimation routine are having a significant impact.

The parameter set acquired from the fitting routine is compared to the actual parameter set in table 5.1. Even though the relative error might seem a bit high, especially for ρ and λ, note that the overall size of the absolute error in the parameters is very small.

	μ	σ_S	κ	α	σ_δ	ρ	λ
parameters	0.03	0.2	5	0	0.35	0.1	-0.1
estim. parameters	0.0226	0.191	5.03	-0.012	0.392	0.014	-0.164
absolute error	0.007	0.009	0.030	0.012	0.042	0.086	0.064
relative error	24.7%	4.5%	0.6%	-	12.0%	86.3%	64.2%

Table 5.1: Comparison of the actual parameters and the results.

The filtered spot price and convenience yield in figure 5.5 fit very accurately with the generated data. Recall that in the case of synthetic data the comparison between the filtered price dynamics and the observed dynamics is especially useful as the observed dynamics is the actual truth, which is not the case with real data.

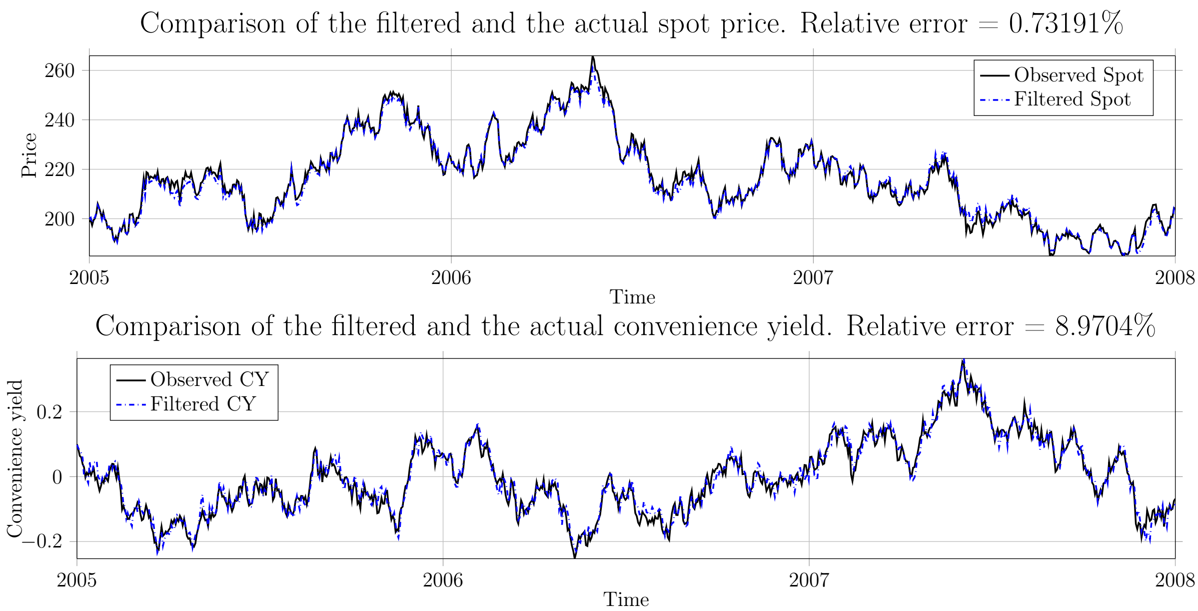


Figure 5.5: A comparison with the filtered state and the actual price dynamics.

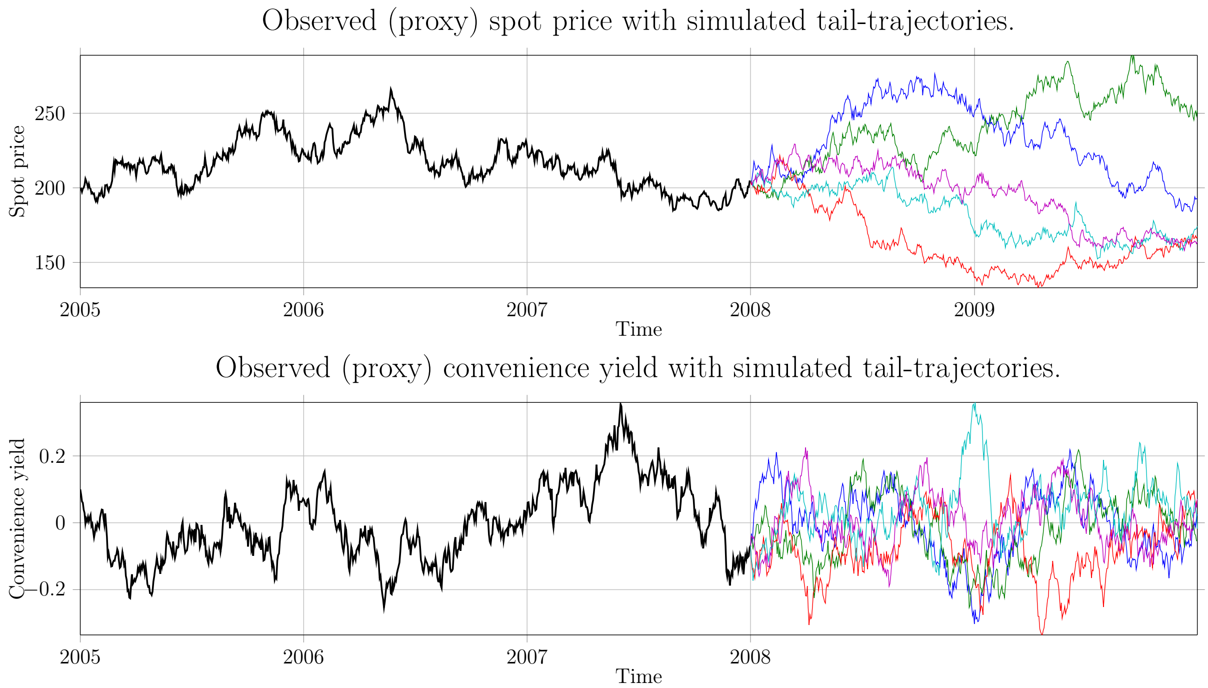


Figure 5.6: Simulated spot price and convenience yield trajectories.

Figure 5.6 illustrates simulated trajectories of both the spot price and the convenience yield after the final observation. The simulations seem to preserve the properties of the observed entities in all the aspects of the model quite well. This observation is supported by the tabulated mean standard deviations in table 5.3.

Δt	spot	simulated spot	CY	simulated CY
5	2.43	2.31	0.0203	0.0221
20	4.39	4.44	0.0409	0.0407
60	7.75	7.57	0.0588	0.0630
130	9.68	11.52	0.0824	0.0822
260	13.83	16.53	0.0898	0.0977

Table 5.2: Mean standard deviations. The mean is calculated as a moving average where the window size is denoted by Δt days.

The mean standard deviation of the simulated entities was taken over 100 sample trajectories that were simulated over a two year period.

The minor divergence of $\Delta t \in \{130, 260\}$ in the table above can be explained by the lack of data; synthetic data was generated for only 3 years, so a moving average with a window size of over 130 days does not provide any particularly useful information. On the other hand, simulations for over a couple of months or so can not even be expected to be very accurate. As pointed out in the model description chapter, one should remember that the assumption of constant system parameters is a simplification, and that it would be more realistic to model the parameters as time dependant.

5.2 Estimation of Brent oil commodity spot price

In this section we will be presenting results on actual data. We will showcase in detail the estimation procedure of Brent crude oil, a typical case study that should give a good insight on the portfolio consisting of 26 different commodities.

The starting point was a raw contract price matrix, from which 4 contracts were extracted with time to maturities illustrated in figure 5.7.

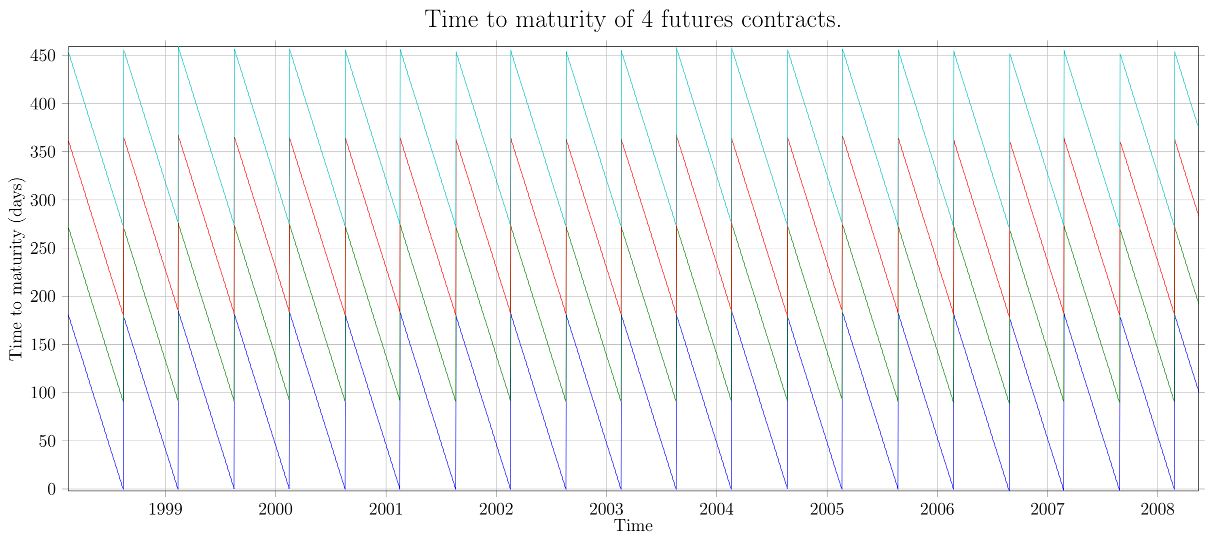


Figure 5.7: The time to maturities of four extracted Brent crude oil contracts.

The term structure is illustrated in figure 5.8.

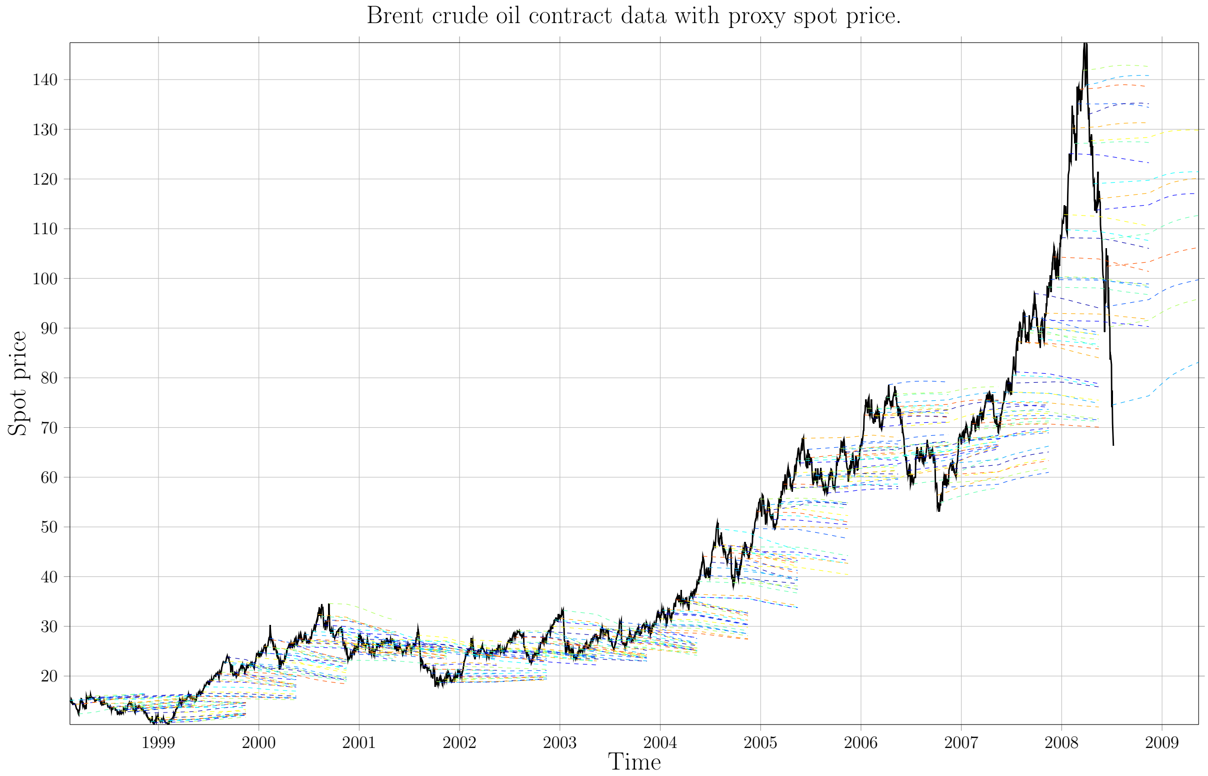


Figure 5.8: The term structure of Brent crude oil.

It is important to note that, in contrast to the case study of synthetic data, the spot price displayed the figure above is only an approximate of the actual spot price. Anything else would be absurd, as the whole problem setting is based on the fact that commodity spot prices are unobservable. As already pointed out in chapter 4, we will be using the futures contracts with the shortest time to maturity as a proxy for spot price. This is a feasible approximation as, in the absence of arbitrage opportunities, it is clear that the futures price should converge to the spot price at maturity.

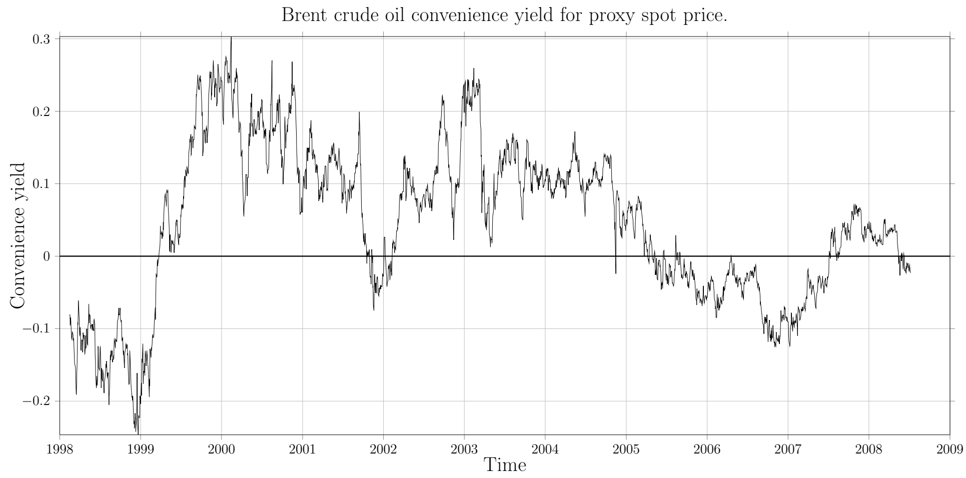


Figure 5.9: The convenience yield calculated for Brent oil futures with shortest time to maturity.

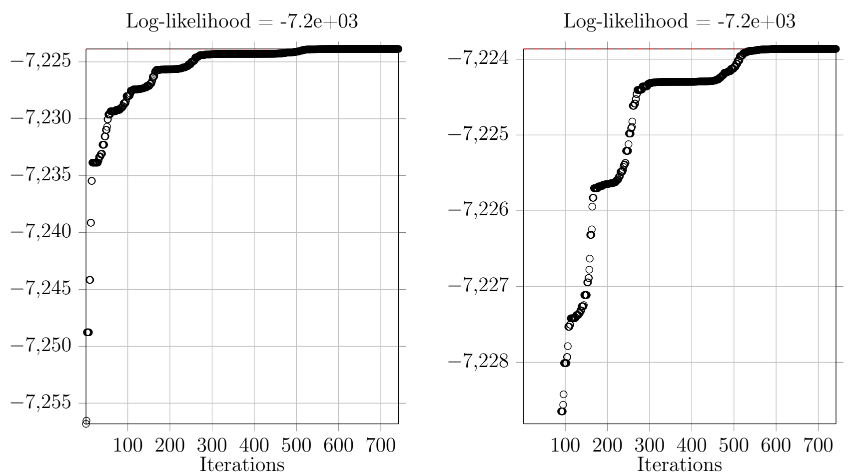


Figure 5.10: Trace of the objective function subject to the optimization routine for Brent crude oil.

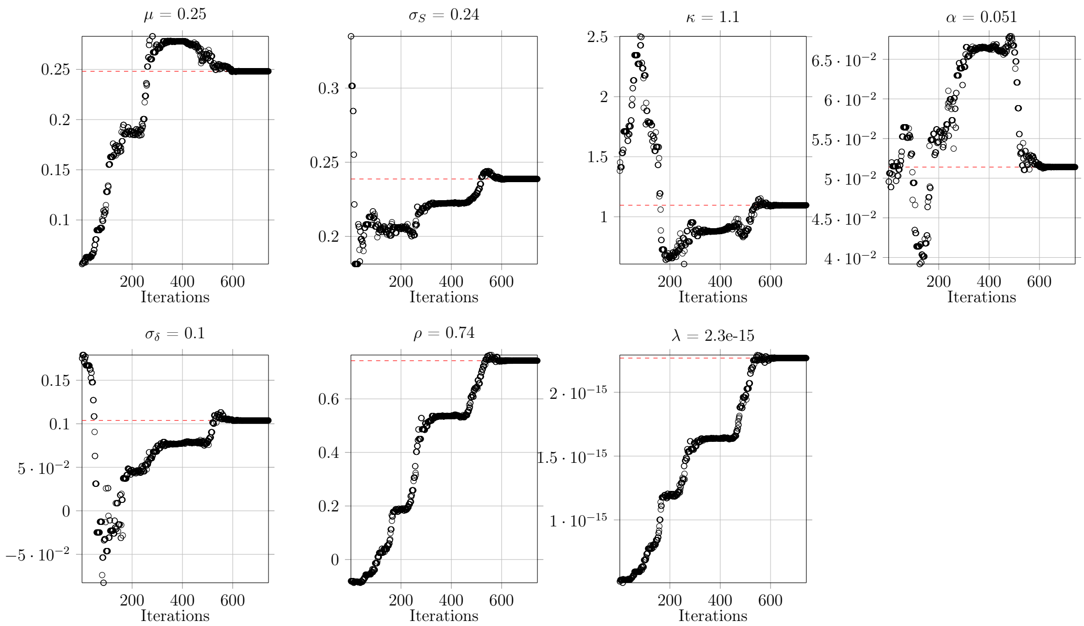


Figure 5.11: The trace of the optimization routine for estimating the parameters for Brent crude oil.

The optimization procedure was run with a similar setup as in the case of synthetic data. Perhaps a bit surprisingly, the procedure convergence significantly faster than in the synthetic study. No absolute reasons for this behaviour could be pinpointed.

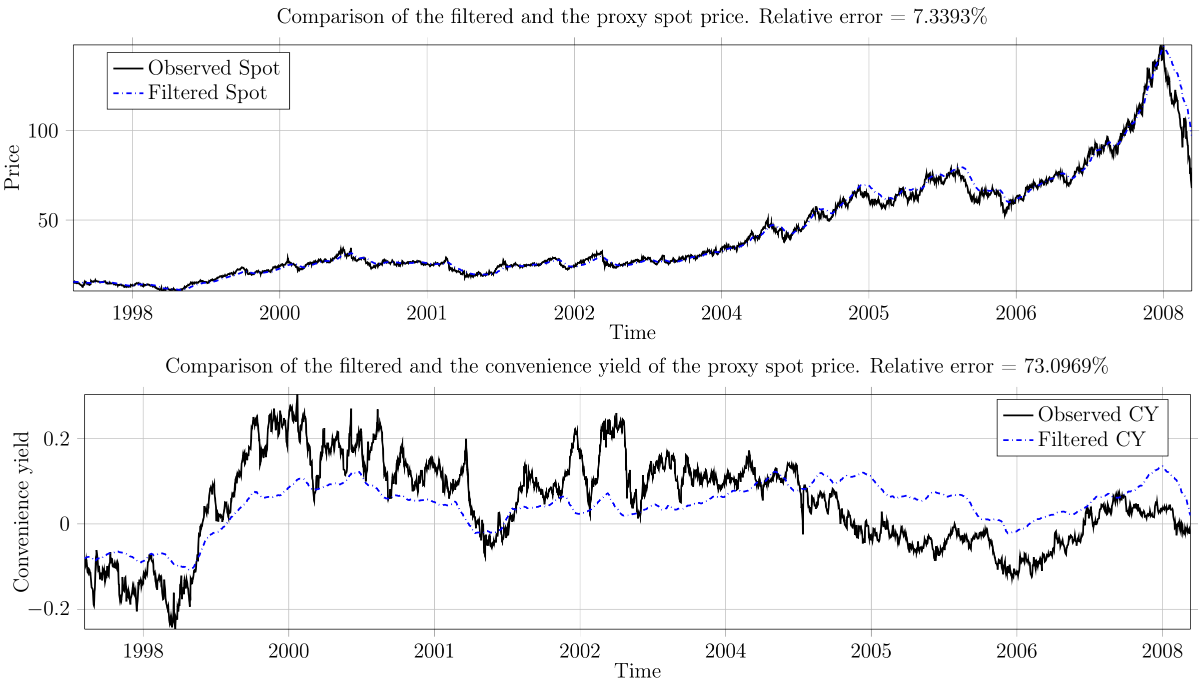


Figure 5.12: A comparison with the filtered state and Brent oil futures with shortest time to maturity.

Figure 5.12 illustrates almost a one to one relation of the proxy spot price and the filtered spot state. On the other hand the filtered convenience yield state seems to diverge quite much from the convenience yield of the proxy spot price. This is in fact quite typical for the optimization routine; the process tends to overfit the spot price at the cost of the convenience yield process. However, the convenience yield process poses similar properties as the proxy convenience yield, the relative error is within the tolerable error range, and it is important to remember comparing the filtered state to the proxy spot dynamics does not convey the absolute truth.

This point can portrayed by comparing the estimated term structure with the observed one.

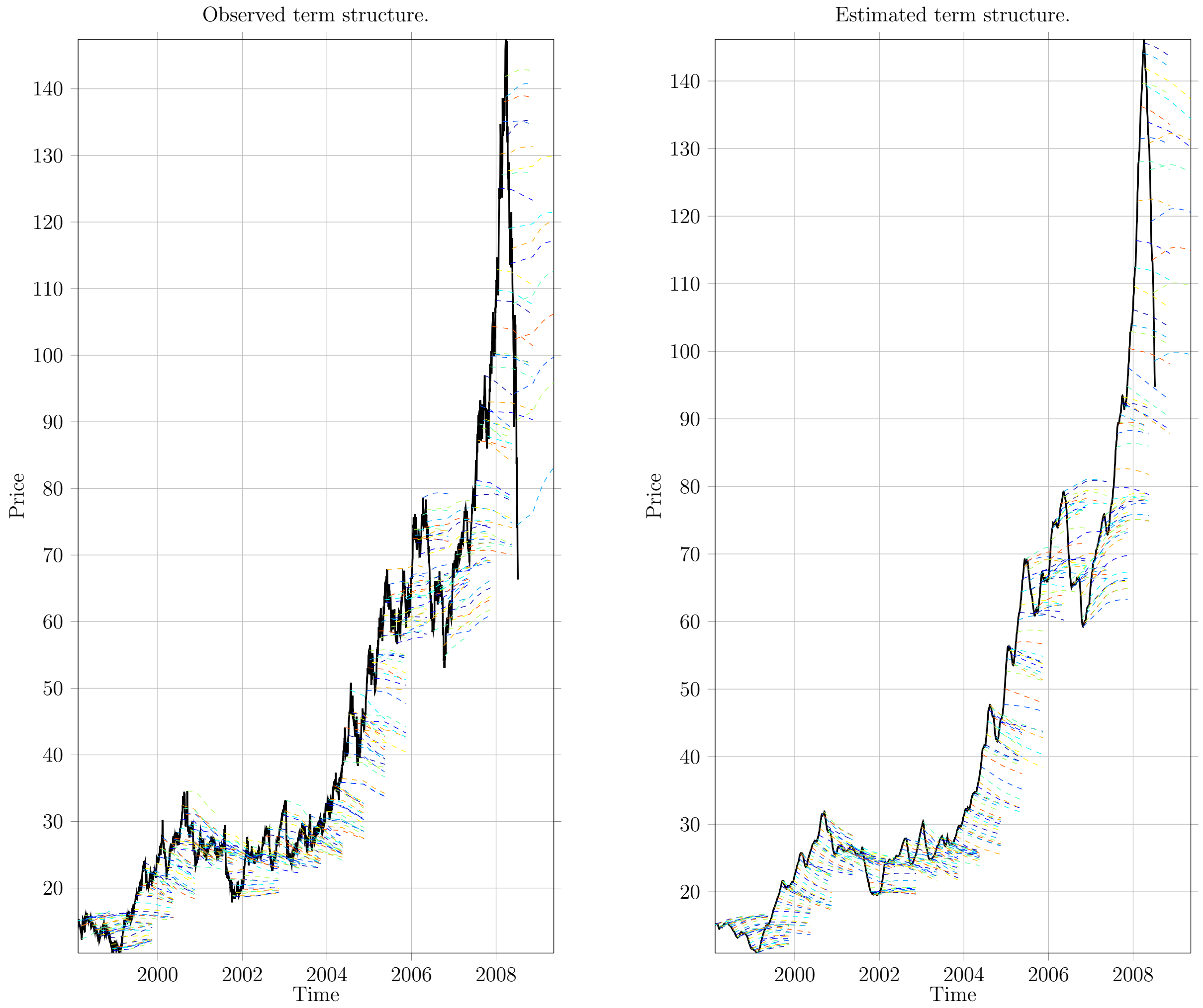


Figure 5.13: A comparison with the estimated and observed term structure.

The estimated term structure is modelled using the filtered state vectors to construct the spot price and convenience yield. The parameters obtained from the optimization procedure are then utilized in the futures pricing formula to construct the term structure. The above graph shows that the estimated term structure possesses very similar properties to the observed term structure with proxy spot price and convenience yield.

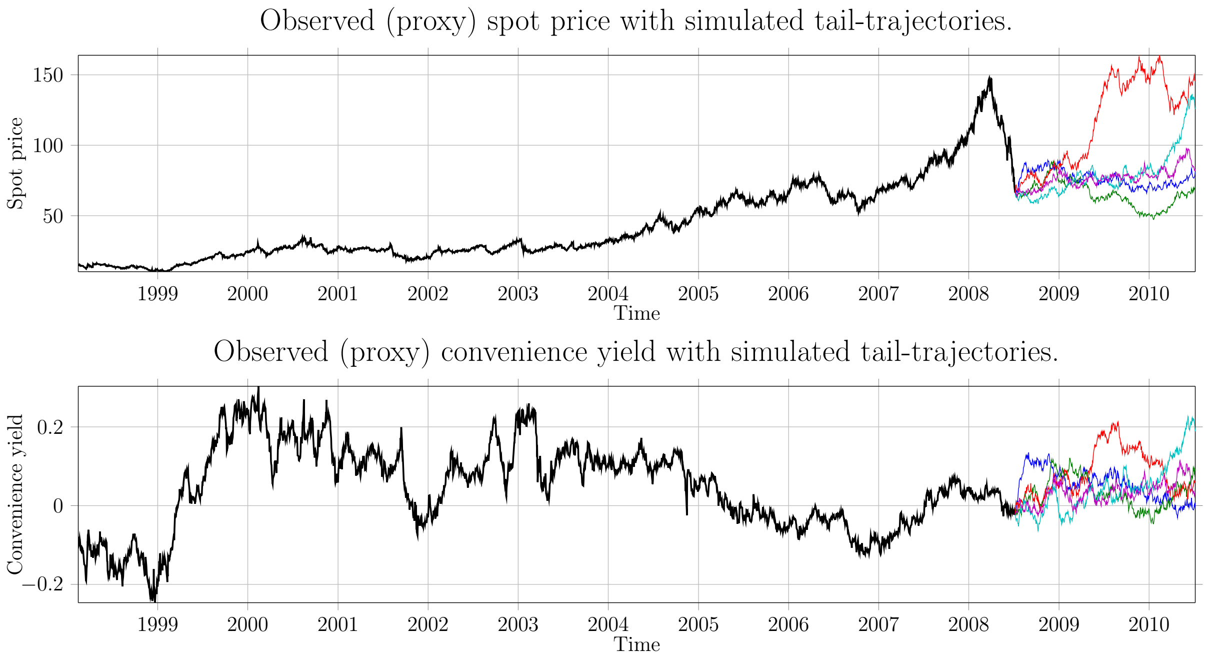


Figure 5.14: Simulated Brent oil spot price and convenience yield trajectories.

Finally, we conducted comparative analysis on the deviations of the simulated trajectories and the historic data. The simulated spot price tail trajectories seem to have higher deviations compared to the historic data. The high peak at the end of the historic data is the culprit in this case, as it impacts strongly on estimation procedure, and especially on the spot volatility.

Δt	spot	simulated spot	CY	simulated CY
5	0.74	1.27	0.0092	0.0058
20	1.40	2.43	0.0175	0.0109
60	2.80	4.12	0.0279	0.0176
130	4.14	6.26	0.0372	0.0254
260	6.08	9.56	0.0498	0.0330

Table 5.3: Mean standard deviations. The mean is calculated as a moving average where the window size is denoted by Δt days.

5.3 Distribution results of the returns on the carry trading strategy

The main goal in this thesis is to quantify basis risk by simulating the spot price dynamics for the commodities in a trading portfolio with a fixed commodity basket. In other words, we will be generating tail trajectories for the different commodities and then apply the trading strategy to the generated spot price and convenience yield trajectories. This process is repeated for distinct dates within a predefined time domain. The parameter estimation procedure will also be repeated for each time-step, which will mimic the time-dependent nature of the parameters of the stochastic model. This procedure is summarized by algorithm 3.

Algorithm 3 The procedure conducted in the distributional analysis. This process generates the return data, which needs to be post-processed in order to acquire significant data from the ditributions.

procedure GENERATE_DISTS($timesteps, F, CY$)
    $dists \leftarrow \{\}$
    for t in $timesteps$ **do**
        $\{\mu, \sigma_S, \kappa, \alpha, \sigma_\delta, \rho, \lambda\} \leftarrow$ ESTIMATE_PARAMS(F, CY)
        $trajs \leftarrow$ GENERATE_TRAJS($\mu, \sigma_S, \kappa, \alpha, \sigma_\delta, \rho, \lambda, F(t), CY(t)$)
        for i in NUMEL($trajs$) **do**
            $dists(t, i) \leftarrow$ CARRY_RETURNS($trajs(i)$)
        end for
    end for
    return $dists$
end procedure

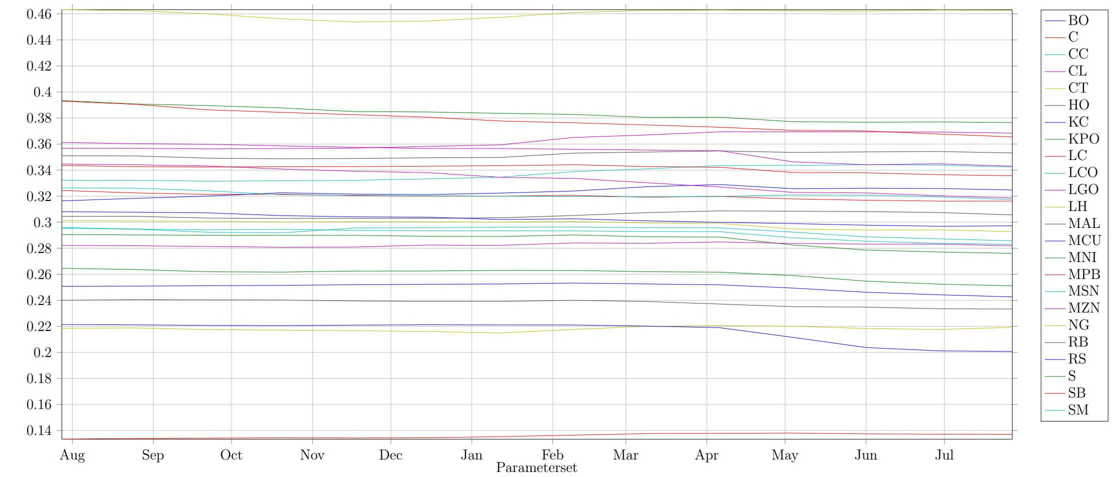
The commodity basket consists of the commodities listed in table 5.4.

Abbreviation	Real name
KPO	Palm oil
CT	Cotton
C	Corn
W	Wheat
KC	Coffee
BO	Soybean oil
LC	Live cattle
CC	Cocoa
HO	Heating oil
NG	Natural gas
RB	Unleaded gasoline
RS	Canola
SM	Soybean meal
LCO	Brent oil
CL	Light crude oil
S	Soybeans
MCU	Copper
SB	Sugar
MAL	Aluminium
MZN	Zinc
MSN	Tin
MPB	Lead
MNI	Nickel
LGO	Crude oil
LH	Lean hogs

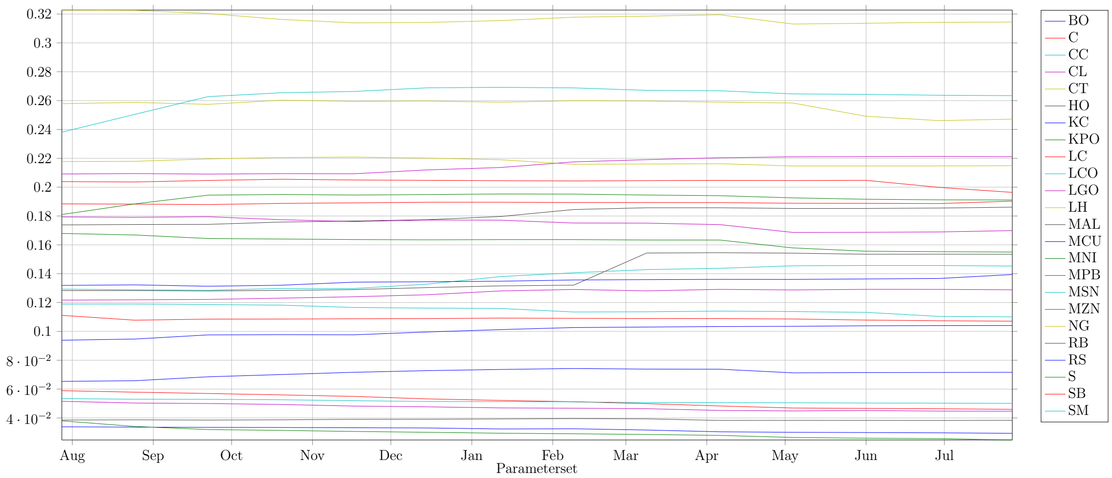
Table 5.4: Listing commodity names with abbreviations.

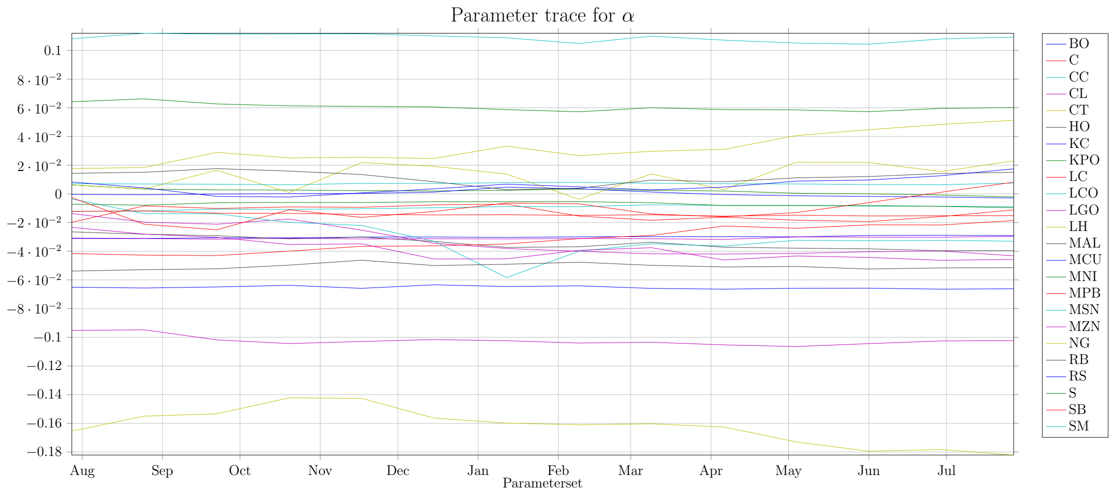
The parameter estimations were run from 2014-7-28 to 2015-7-28 with $\Delta t = 10$. The figures below illustrate the parameter value evolution within the time domain.

Parameter trace for σ_S

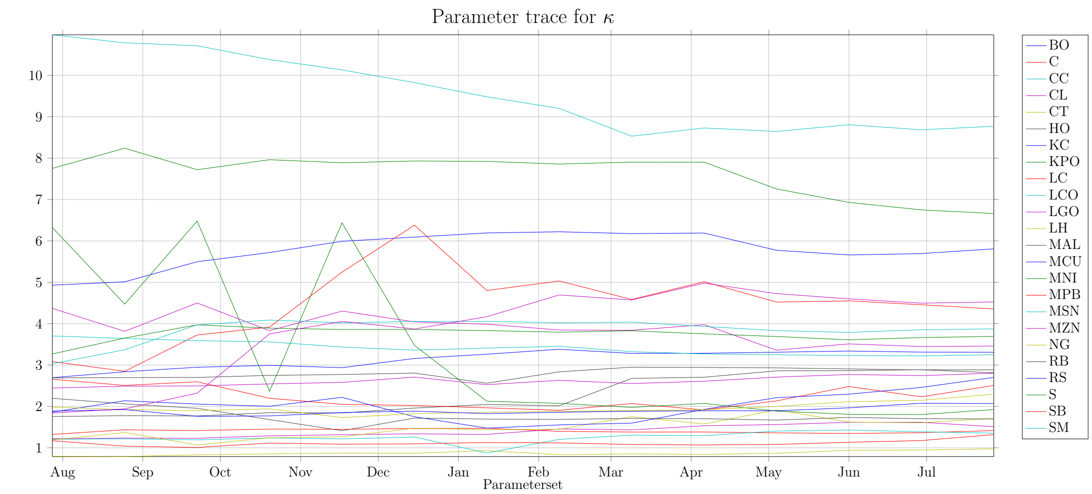


Parameter trace for σ_δ

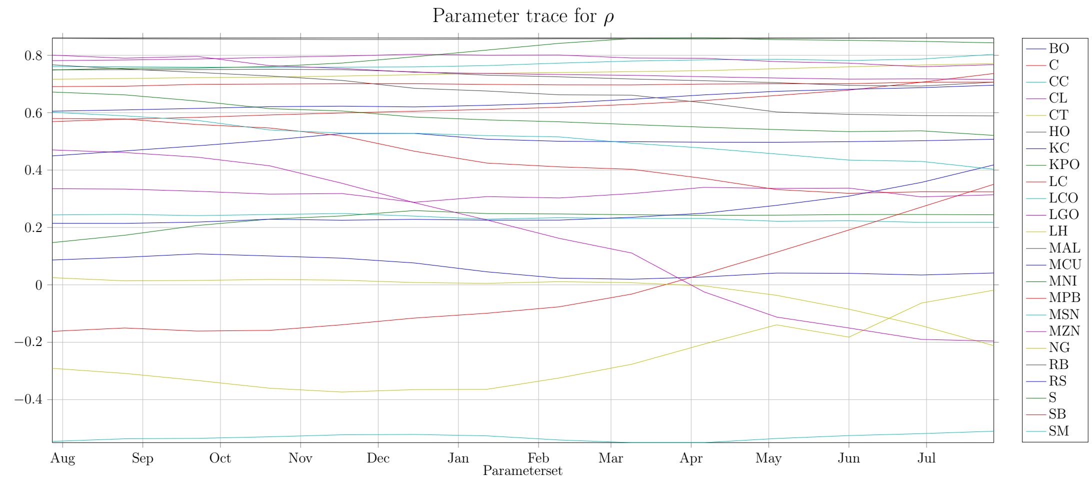




Parameter trace for α
0.1
$8 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
0
$-2 \cdot 10^{-2}$
$-4 \cdot 10^{-2}$
$-6 \cdot 10^{-2}$
$-8 \cdot 10^{-2}$
−0.1
−0.12
−0.14
−0.16
−0.18
Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul
Parameterset
BO C CC CL CT HO KC KPO LC LCO LGO LH MAL MCU MNI MPB MSN MZN NG RB RS S SB SM



Parameter trace for κ
10
9
8
7
6
5
4
3
2
1
Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul
Parameterset
BO C CC CL CT HO KC KPO LC LCO LGO LH MAL MCU MNI MPB MSN MZN NG RB RS S SB SM



Parameter trace for ρ
0.8
0.6
0.4
0.2
0
−0.2
−0.4
Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul
Parameterset
BO C CC CL CT HO KC KPO LC LCO LGO LH MAL MCU MNI MPB MSN MZN NG RB RS S SB SM

With the acquired parameter sets, we ran the process described in algorithm 3 on a daily basis for each commodity. The simulated trajectories were all generated for a two month period ($\Delta t = 20$). Ten thousand trajectories were generated for each timestep. These results were then aggregated, according to the portfolio allocations for each commodity, to the portfolio level in order to be able to conduct distributional analysis on a portfolio level.

The figure 5.15 illustrates the (excess) kurtosis and skewness of the returns distributions on the portfolio level. An observed outlier is also marked in the figure.

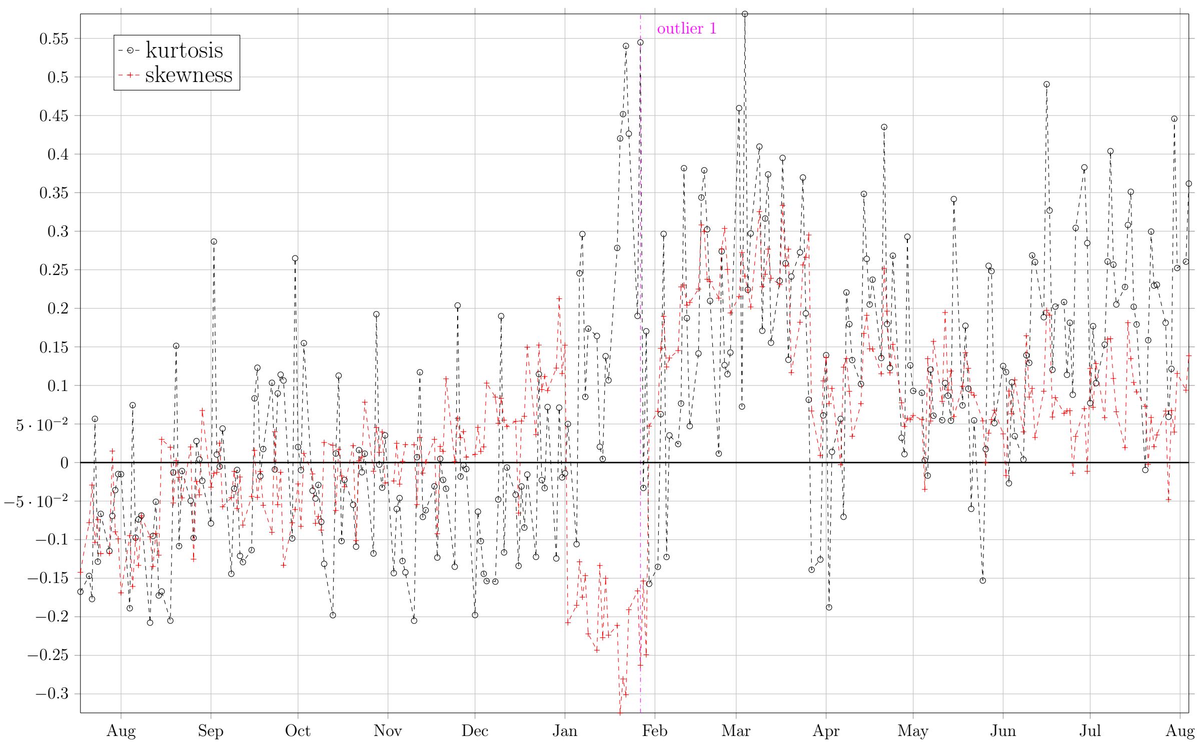


Figure 5.15: The kurtosis and skewness of the returns distribution on the portfolio level.

The statistics for the returns distributions are quite noisy; a moving monthly average of the distributions smoothens the noise to an acceptable level.

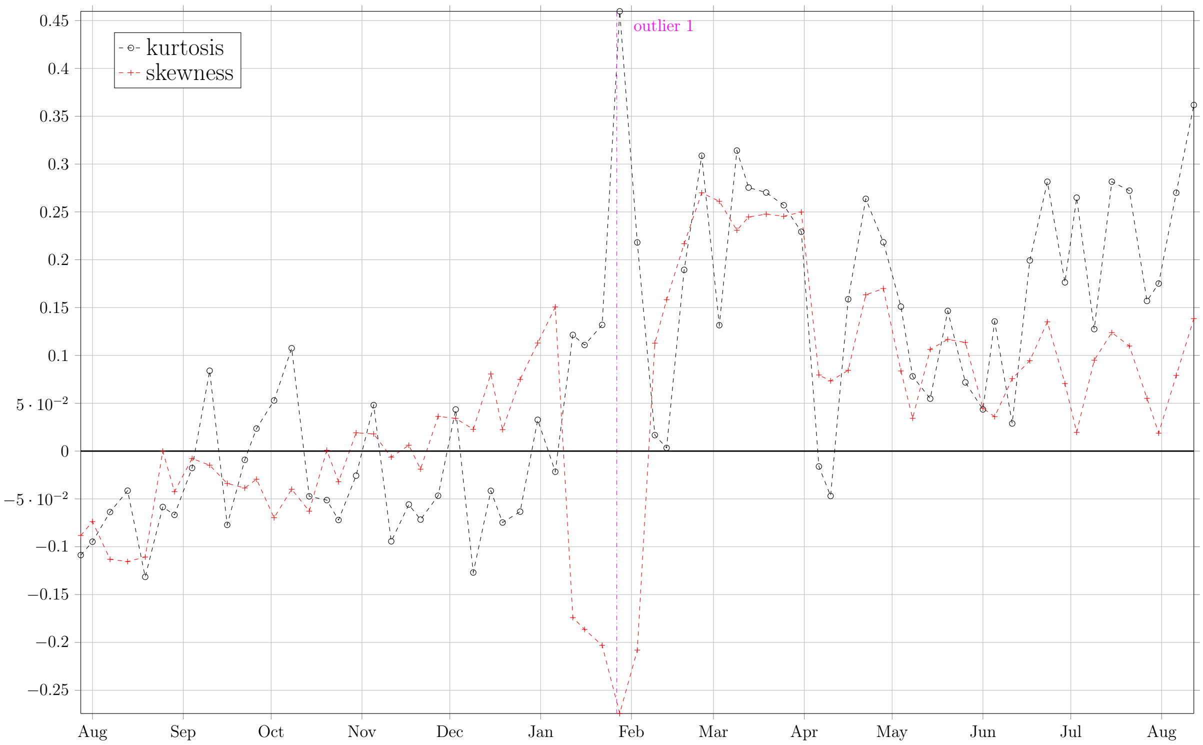


Figure 5.16: A moving average of the kurtosis and skewness of the returns distribution on the portfolio level. The length of the time window is one month.

Chapter 6

Discussion

In this chapter we will contemplate on the validity, optimality and, in general, the pros and cons of our approach, and asses the overall quality of the results presented in the previous chapter. We will also attempt to make suggestions for future work by referencing to recent development within the field which could possibly be incorporated to our procedure of quantifying basis risk.

The fundamental basis of the whole distributional analysis procedure is built on the parameter estimation scheme; without being able to accurately estimate the system parameters of the stochastic model of the commodity spot price dynamics, any attempts to simulate the term structure would be futile. Therefore we feel that the parameter estimation scheme deserves a section of it's own, and the parameter estimation scheme will be analysed in section 6.1.

The analysis of the distributional analysis procedure and of the results acquired in the previous chapter is conducted in section 6.2.

6.1 Parameter estimation

The results of applying the parameter estimation scheme on generated synthetic data speaks for itself. The well behaved pattern of the optimization trace combined with a good accuracy in the estimation results indicate that the estimation procedure might actually be quite robust considering the relatively high parameter set dimension. These results are of course, as pointed out in the previous chapter, only valid as a proof of concept; we should be expecting an even higher level of irregularities in the term structures of actual futures data, which can potentially lead to serious convergence issues in the estimation routine. The process of running an estimation procedure on synthetic data generated by the same model used in the estimation procedure is sometimes, and especially within the field of inverse problems referred to as *inverse crime*[29]. This is considered a kind of a false positive; the quality of the results is positively influenced by the underlying model used in the data generation phase.

The results in the case of Brent crude oil however further promote the robustness of the estimation scheme. The procedure converges to a reasonable dataset that performs well against the historic data. Especially the simulated tail trajectories express similar properties to the historic data both in the intuitive visual sense and numerically by comparing the moving average of the standard deviations of the trajectories. The characteristics of the estimation scheme are actually clearly visible in this case. The aggressive peak of the futures price at the end of the historic dataset has a significant impact in the procedure, as it tends to favor more recent characteristics of the data over the ones present earlier in the dataset. This explains the slight divergence of the average standard deviations; the peak in the futures price positively influences the volatility.

Even though we are very satisfied with the robustness and accuracy of the results of the parameter estimation scheme in the case of Brent crude oil, especially as it is a typical representative of the commodities in the commodity basket of the investment portfolio, we do recognize some shortcomings in the procedure. We noticed that the historic dataset needs to of a remarkable size. In the case of Brent crude oil, we have roughly a decades worth of historic data, which is not always the case. We observed that a dataset with less than a couple years of historic data (with approximately four different contracts) the estimation procedure would show monumental difficulties in converging to a reasonable parameter set. Thus we encountered issues when attempting to perform more thorough distributional analysis as we lacked the sufficient datasets.

We also observed that cases where the evolution of the convenience yield of the commodity was seasonal (typical for e.g. crops) were cumbersome. The problem is that the model can not recognize seasonality, but rather interprets this periodic behaviour as an increase in the volatility of the convenience yield, resulting in obscure results in the estimation procedure. Therefore pre-processing measures in the case of seasonal data are imperative. This is an known issue of the model, and has influenced the development of alternative models that take the seasonality into account, for example [34], that uses deterministic trigonometric functions to describe the seasonality. The trade-off is an increase in the dimension of the parameter space, which is not a trivial issue, as the problems described in the paragraph above indicate.

Also, our choice of the optimization algorithm (Nelder-Mead) is a simplex-based optimization routine. An alternative for future work would be to implement the parameter estimation using some other, perhaps more sophisticated optimization algorithm, for example [22], which does not rely on gradient computations. The optimization is implemented in MATLAB as `direct` [13].

Even though we recognize that the estimation scheme has it's flaws, we feel that it is nevertheless the most valuable asset in this thesis; especially the robustness of the procedure makes the current form of the estimation procedure a compelling choice for the baseline for further development.

6.2 Distribution analysis

The results from the distribution analysis are somewhat noisy, but still signal a positive response; the skewness and the (excess) kurtosis of the return distribution, around the time of the observed outlier in the actual trading situation, indicate that the return distribution would possess properties of a left fat tailed distribution in this time window. The significance of this phenomenon is underlined by the fact that these properties are very distinct in this particular timeframe. The skewness, in particular, exhibits a relatively strong negative trend that is exclusive to this time domain. In an actual trading situation, this information could function as a trigger for proactive measures against the more probable losses in the trading strategy.

We do, however, realize that these results are inconclusive and that the process itself is, in it's current form, not easily adapted as a tool for real life continuous analysis. The above results should be reproduced on a larger scale with more extensive data in order to verify the significance of the critical statistics of the distribution. As described in the previous section, due to convergence issues with limited data, this is not a possibility within the scope of this thesis and is left for future work. It could be useful to also extract additional statistics from the distributions, such as standard deviations, and analyse the significance of these as a measure of risk.

In the production of the distributions, 10000 to 15000 trajectories were generated for each commodity. The analysis procedure will face performance issues when this path generating process is repeated for each timestep. The large number of trajectories might be considered an overkill, but in order to get consistent results, we observed that a significant number of trajectories is required. Large inconsistencies were present for distributions generated from less than 1000 trajectories. Hence, determining the sufficient amount of generated trajectories would be a worthwhile feat left for future work.

References

[1] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[2] P. Bjerksund and Institutt for foretaksøkonomi. *Contingent Claims Evaluation when the Convenience Yield is Stochastic: Analytical Results*. Institutt for foretaksøkonomi, Institute of Finance and Management Science, 1991.

[3] M.J. Brennan, N.I. Crew, and John E. Anderson Graduate School of Management at UCLA. *Hedging Long Maturity Commodity Commitments with Short-dated Contracts*. ECGI working paper series in finance. John E. Anderson Graduate School of Management at UCLA, 1995.

[4] Damiano Brigo, Antonio Dalessandro, Matthias Neugebauer, and Fares Triki. A stochastic processes toolkit for risk management. *Available at SSRN 1109160*, 2007.

[5] Robert Brown. Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine, or Annals of Chemistry, Mathematics, Astronomy, Natural History and General Science*, 4(21):161–173, 1828.

[6] Rohan Christie-David and Mukesh Chaudhry. Coskewness and cokurtosis in futures markets. *Journal of empirical finance*, 8(1):55–81, 2001.

[7] R.A. Dana, A. Kennedy, and M. Jeanblanc. *Financial Markets in Continuous Time*. Springer Finance. Springer Berlin Heidelberg, 2007.

[8] J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Noise trader risk in financial markets. *Journal of political Economy*, pages 703–738, 1990.

[9] M.A.H. Dempster and S.R. Pliska. *Mathematics of Derivative Securities*. Isaac Newton Institute for Mathematical Sciences Cambridge: Publications of the Newton Institute. Cambridge University Press, 1997.

[10] Rached Dhaouadi, Ned Mohan, and Lars Norum. Design and implementation of an extended kalman filter for the state estimation of a permanent magnet synchronous motor. *Power Electronics, IEEE Transactions on*, 6(3):491–497, 1991.

[11] Francis X Diebold. *Elements of forecasting*. Citeseer, 1998.

[12] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4, 1905.

[13] Daniel E Finkel. Direct optimization algorithm user guide. *Center for Research in Scientific Computation, North Carolina State University*, 2, 2003.

[14] José Carlos Garcıa Franco. Maximum likelihood estimation of mean reverting processes. *Real Options Practice*, 2003.

[15] Ana-Maria Fuertes, Joëlle Miffre, and Georgios Rallis. Tactical allocation in commodity futures markets: Combining momentum and term structure signals. *Journal of Banking & Finance*, 34(10):2530–2548, 2010.

[16] R Gibson and E Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3):959–976, 1990.

[17] Mike Giles. Improved multilevel monte carlo convergence using the milstein scheme. *Monte Carlo and Quasi-Monte Carlo Methods*, pages 343–358, 2006.

[18] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2003.

[19] A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1990.

[20] J. Hull. *Options, Futures and Other Derivatives*. Options, Futures and Other Derivatives. Pearson/Prentice Hall, 2009.

[21] Huu Tue Huynh, Van Son Lai, and Issouf Soumare. *Stochastic Simulation and Applications in Finance with MATLAB Programs*. Wiley Publishing, 2009.

[22] Donald R Jones, Cary D Perttunen, and Bruce E Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.

[23] Steven M Kay. Fundamentals of statistical signal processing, volume i: estimation theory. 1993.

[24] Young-Real Kim, Seung-Ki Sul, and Min-Ho Park. Speed sensorless vector control of induction motor using extended kalman filter. *Industry Applications, IEEE Transactions on*, 30(5):1225–1233, 1994.

[25] F.C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, 2012.

[26] Annika Lang, Pao-Liu Chow, and Jürgen Potthoff. Almost sure convergence of a semidiscrete milstein scheme for spdes of zakai type. *Stochastics: An International Journal of Probability and Stochastics Processes*, 82(3):315–326, 2010.

[27] Robert Liptser and Albert N Shiryaev. *Statistics of Random Processes: I. General Theory*, volume 5. Springer Science & Business Media, 2013.

[28] Larry Matthies, Takeo Kanade, and Richard Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238, 1989.

[29] Jennifer L Mueller and Samuli Siltanen. *Linear and nonlinear inverse problems with practical applications*, volume 10. Siam, 2012.

[30] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[31] Philip E Protter. *Stochastic integration and differential equations*, volume 21. Springer, 2013.

[32] Diana R. Ribeiro and Stewart D. Hodges. A two-factor model for commodity prices and futures valuation. *SSRN eLibrary*, 2004.

[33] Maria Isabel Ribeiro. Kalman and extended kalman filters: Concept, derivation and properties. *Institute for Systems and Robotics*, 43, 2004.

[34] M Richter and C Sorensen. Stochastic volatility and seasonality in commodity futures and options: The case of soybeans (working paper). *Frederiksberg, Denmark: Copenhagen Business School*, 2002.

[35] Fabrice Douglas Rouah. The feynman-kac theorem.

[36] Eduardo S Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–73, 1997.

[37] Carsten Sørensen. Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393–426, 2002.

[38] Thijs van den Berg. Calibrating the ornstein–uhlenbeck (vasicek) model. *Web page http://www. sitmo. com/article/calibrating-the-ornstein-uhlenbeck-model/. View date October 24th*, 2011.

[39] Ubbo F Wiersema. *Brownian motion calculus*. John Wiley & Sons, 2008.

[40] D. Williams. *Probability with Martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.

[41] Andreas Winkelbauer. Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*, 2012.