5-11-2015 12:00 AM

Quantitative Techniques for Spread Trading in Commodity Markets

Mir Hashem Moosavi Avonleghi
The University of Western Ontario

Supervisor
Dr. Matt Davison
The University of Western Ontario

Graduate Program in Statistics and Actuarial Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

QUANTITATIVE TECHNIQUES FOR SPREAD TRADING IN COMMODITY MARKETS

(Thesis format: Monograph)

by

Mir Hashem Moosavi Avonleghi

Graduate Program in Statistics and Actuarial Science

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

This thesis investigates quantitative techniques for trading strategies on two commodities, the difference of whose prices exhibits a long-term historical relationship known as mean-reversion. A portfolio of two commodity prices with very similar characteristics, the spread may be regarded as a distinct process from the underlying price processes so deserves to be modeled directly. To pave the way for modeling the spread processes, the fundamental concepts, notions, properties of commodity markets such as the forward prices, the futures prices, and convenience yields are described. Some popular commodity pricing models including both one and two factor models are reviewed. A new mean-reverting process to model the commodity spot prices is introduced. Some analytical results for this process are derived and its properties are analyzed. We compare the new one-factor model with a common existing one-factor model by applying these two models to price West Texas Intermediate (WTI) crude oil, and discuss its advantages and disadvantages. We investigate the recent behavioral change in the location spread process between WTI crude oil and Brent oil.

The existing three major approaches to price a spread process namely cointegration, one-factor and two-factor models fail to fully capture these behavioral changes. We, therefore, extend the one-factor and two-factor spread models by including a compound Poisson process where jump sizes follow a double exponential distribution. We generalize the existing one-factor mean-reverting dynamics (Vasicek process) by replacing the constant diffusion term with a nonlinear term to price the spread process. Applying the new process to the empirical location spread between WTI and Brent crude oils dataset, it is shown how the generalized dynamics can rigorously capture the most important characteristics of the spread process namely high volatility, skewness and kurtosis. To consider the recent structural breaks in the location spread between WTI and Brent, we incorporate regime switching dynamics in the generalized model and Vasicek process by including two regimes.

We also introduce a new mean-reverting random walk, derive its continuous time stochastic differential equation and obtain some analytical results about its solution. This new mean-reverting process is compared with the Vasicek process and its advantages discussed. We showed that this new model for spread dynamics is capable of capturing the possible skewness, kurtosis, and heavy tails in the transition density of the price spread process. Since the analytical transition density is unknown for this nonlinear stochastic process, the local linearization method is deployed to estimate the model parameters. We apply this method to empirical data for modeling the spread between WTI crude oil and West Texas Sour (WTS) crude oil.

Finally, we apply the introduced trading strategies to empirical data for the location spread between WTI and Brent crude oils, analyze, and compare the profitability of the strategies. The optimal trading strategies for the spread dynamics in the cointegration approach and the one-factor mean-reverting process are discussed and applied to our considered empirical dataset. We suggest to use the stationary distribution to find opti-

mal thresholds for log-term investment strategies when the spread dynamics is assumed to follow a Vasicek process. To incorporate essential features of a spread process such as skewness and kurtosis into the spread trading strategies, we extend the optimal trading strategies by considering optimal asymmetric thresholds.

Keywords: Pairs trading; Commodity pricing model; Commodity spread process; Convenience Yield; Cointegration; Mean-reversion; Energy markets; Mean-reverting random walk; The Kalman filter; Optimal Trading Strategy; Regime Switching Algorithm; Ornstein-Uhlenbeck process; Crude Oil Futures Prices

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Professor Matt Davison, for his invaluable guidance and continued support, and immense knowledge. This thesis would not have been possible without his substantiable help and his long-lasting patience. I am appreciative of his generosity with his time and for great stimulating discussions.

I wish to extend my warmest thanks for the advice that I received from Dr. Adam Metzler during the time I studied for my Master's degree at Applied Mathematics Department and his leading me to make the right choice. I would like to express grateful thanks to Dr. Mark Reesor for his guidance and leading financial modeling group. I am grateful to all of the people who helped in the various stages of this research. I would also like to express my sincere thanks to the members of my thesis examiners.

I acknowledge the Department of Statistical and Actuarial Sciences and the Faculty of Graduate Studies for their financial support. I would also like to thank to Jennifer Dungavell (Administrative Officer), Jane Bai (Academic Coordinator), and Keri Knox (Professional Education Coordinator) for their assistance.

Special thanks also go to my friends Ali Akbar Mohsenipour, Azaz Bin Sharif, and Mehdi Taheri for their support and encouragement.

Finally, I express my thanks to all my other friends and colleagues who provided great company for all these years I spent at Western. Thank you for making academic life tolerable and for the excellent memories that I will never forget.

Dedicated To:

My Parents

My Sisters

And My Brothers

Thank you for your love, support and encouragement

Contents

List of Figures

List of Tables

Chapter 1

COMMODITY MARKETS

1.1 Introduction

This thesis is about developing stochastic models for commodity prices and for spreads between commodity prices. Based on these models, we build optimized trading strategies on spreads (more detail is given in section 1.6). We begin the thesis with an introduction to commodity markets.

When we review daily financial newspapers, we find that commodity-related news provide many important headlines and columns. For instance, "corn is shining like gold" was one headline in The Globe and Mail on July 04, 2012. The impact of the recent decline in commodity prices are inevitable on national economies. Since countries such as Canada, Russia, and Australia have commodity-based economies, falling commodity prices will cause considerable decline in their gross domestic product (GDP), possibly leading to financial crisis if the decline continues for a long period. For instance, according to *Statistics Canada*, in 2009 approximately 58% of Canada's entire exports were from energy, forestry, mining, and agriculture. Most commodities prices, especially that of crude oil, started to rise between 2002 to mid-2008 and peaked in mid-2008. Although these prices started receding in October 2008 and dramatically dropped when the world faced the financial crisis and the resulting economic recession, their prices remained significantly higher and more volatile compared to 2005 and earlier. Again, the commodity prices have been modestly increasing since mid-2009. These trends are not just of interest to financiers; as a result, food prices extended to an all-time high in February 2011 (United Nations (2011)). This chapter is organized as follows:

First we explain commodities in general. In section 1.3, we describe how commodity markets work. We review the commodity markets properties in section 1.4. Finally, in section 1.5, we compare the commodity markets with equity markets.

1.2 What is a commodity?

In economics, a commodity is a term that refers to any marketable item produced to fulfill wants or needs. Commodities comprise both goods and services, but the term is more particularly used to refer to goods only. The word commodity comes into English from the French word "commodit", which is used to refer to any item that provide some benefits or useful services. The term commodity applies to any good that is interchangeable with another product of the same type without qualitative distinction across a market. In other words, commodities produced by different producers are treated as equivalent apart from their slightly different qualities. For instance, wheat produced by Russia, USA and India is treated as equivalent. Any good which is supplied across the markets without any product differentiation, and for which there exists demand, is a commodity. By this definition, crude oil is a commodity because it does have a single price all around the world on daily basis and the price is determined by supply and demand. Although commodities are uniform, their prices may marginally vary with respect to transportation costs, qualitative differences, currencies exchange rates and delivery places and times. To cover these differences, commodities are usually graded. However, in order to be suitable for trading and deliverable, a commodity must meet minimum acceptance standard which is called "a basic grade", "par grade", or "contract grade" Chatnani (2010). Commodities are usually produced by many different producers in large quantities. Commodity prices fluctuate based on supply and demand. Moreover, a commodity is considered as a consumption asset that its "scarcity" has a significant impact on economic development, international trade, world economic, and political stability. The demand for commodities including energy, grains, and metals as well as the availability of these commodities has significantly increased over the centuries. To sum up the definition of an economic commodity, a commodity is a good that has following properties:

- Produced and sold by different producers
- Between producers of the commodity, its quality is uniform
- Its price is determined based on supply and demand alone

Commodities are categorized as follow:

- **Agricultural commodities:** raw "grains" such as wheat, corn and oats
- **Industrial metals:** raw metals including aluminum, lead, and copper
- **Precious metals:** raw metals including gold, silver, and platinum
- **Meat commodities:** meat commodities are raw products like livestock

- **"Soft" commodities:** including coffee, sugar, and lumber
- **Energy commodities:** including crude oil, gasoline, and heating oil

1.3 Commodity markets:

Commodities spot markets have existed over nearly the entire history of humankind around the world. A commodity market is a physical or virtual marketplace in which buyers and sellers meet in order to trade standardized or graded products. Although "spot" contracts, for immediate delivery of a particular commodity, still exist, most modern trading is done through forward, futures contracts, or derivatives. The contemporary futures markets were established in the Midwestern United States in the 19th century, even though some sort of futures contracts already existed in Europe and Japan centuries ago (Geman (2005)). A forward contract is an over-the-counter (OTC) contract representing an agreement between two parties for the exchange of a commodity at a certain time in the future for a price settled at the time of the agreement. A futures contract is a particular type of forward contract. However, there are some specific differences between futures and forward contracts. First, futures contracts are traded on future exchanges and are standardized contracts; on the other hand, forward contracts are agreements between individual counterparties and are more flexible in their specified terms and conditions. Since forward contracts are private agreements, and usually the exchange of the commodity and its cash value is done at maturity, there is always a possibility that one of the counterparties may default. To resolve this issue, future contracts have clearing houses that guarantee the creditworthiness of transactions, which considerably reduces the probability of default to almost zero.

A clearing house is a financial organization on a financial exchange that is maintained by banks and coordinates the delivery, confirmation and settlement of securities transactions for a commission. In other words, a clearing house settles trading accounts, clears trades, collects and maintains margin money. Furthermore, settlement of forward contracts occur at the end of the contract, but futures contracts are marked-to-market daily, which means that price changes are settled in counterparties margin accounts day by day until the expiry date of the contract. Since futures contracts are standardized with respect to quality, quantity, delivery terms and expiry dates and can be easily transferred to other parties, their liquidities are considerably higher than forward contracts. Forward contracts, on the other hand, can only be traded on the exchange that created that particular forward contract. Sometimes forward contracts are customized over the counter contracts that are difficult to trade at all.

The popularity of commodity derivatives trading has rapidly grown over the years. In recent years, commodity exchanges have witnessed rapid growth in trading volume, the diversity of contracts, wide range of underlying commodities, and market participants. Commodity exchanges facilitate the access of market participants including speculators and hedgers to commodity futures and other derivatives. The numbers of commodity exchanges have increased in recent decades. Most commodity exchanges are located in the world's leading financial centers including New York, Chicago, and London. The Chicago Board of Trade (CBOT) was established in 1848 to trade agricultural products. CBOT invented futures contracts trading in 1865 and is one of the oldest futures and options trading exchanges. We can also mention some other prominent commodity exchanges across the world namely the Chicago Mercantile Exchange (CME) (On July 12, 2007, the CBOT merged with CME to form the CME Group), the New York Mercantile Exchange (NYMEX), the London Metal Exchange (LME), the world's largest markets for commodities (mainly metals) derivatives, and the Indian Commodity Exchange Limited (ICEX).

1.4 Commodity markets properties:

Commodity markets have some properties that make them different from that of equity markets. These properties even vary from one commodity to another. For instance, seasonality is crucial to be considered when we attempt to build a model for agricultural commodities whereas precious metals prices are not generally seasonal even though they may be cyclical. In this section, we review some of these properties.

- **Demand and Supply:** The commodity spot prices are derived based on demand and supply. When the demand is lower than supply, the commodity prices will drop and vice versa. The intersection of the supply and demand curves will form the spot prices. When the commodity price is low, some producers whose production costs are high will decide to stop producing. Therefore, supply will fall. However, low supply will make price trend upward sooner or later.

- **Physical transactions:** The commodity markets are associated with physical delivery. The producers and consumers must agree on the place and time of the commodity exchange, although in the commodity exchanges, there are two types of futures contracts: physical and financial settlements, contracts for differences (CFD) for which there is no physical delivery and only cash value differences will apply to the counterparties accounts. Huge volumes of trades are done by speculators in these types of contracts.

- **Liquidity:** One of the issues of the commodity markets is illiquidity. Once incidents occur, even in well covered markets, the volume of trades will sharply drop

and the spreads (bid and ask) will widen. As a result, the market will be illiquid.

- **Storage:** Although for some commodities such as electricity, storage is either not practical or economical, storage, holding and protecting commodities for future consumption, plays a crucial role in most commodity markets. At every given time, production plus inventory define supply. Low inventory for a given commodity will drive its spot and futures prices up.

- **Volatility:** Major commodity prices are highly volatile. The commodity prices have sharply peaked in recent years even though in some cases such as natural gas, the prices have since retreated. The magnitude of commodities' price rise had extended five times or more in some cases. The volatile behaviors of commodity prices have various reasons. First, demands for commodities, particularly grains, are relatively inelastic- almost constant; as a result, when the supply fluctuates due to some fundamental price drivers such as weather, the commodity prices display volatile behaviors. In recent decades Asian countries, especially China and India, have experienced rapid industrial development. Therefore, demands for major commodities such as metals, and energy as inputs to the production have considerably added and caused price climbs. The increased number of market participants, particularly speculators, and the variety of derivatives on commodities also have added price volatilities in commodity markets. Commodity markets abruptly respond to events and news, and the frequency of news events are very high. For instance, conflicts in crude oil suppliers can make price jump, or drought in main grains producers countries such as USA, and Russia will cause prices to spike. Figure 1.1 depicts daily front contracts futures prices for Gold, Silver, Wheat, and WTI crude oil from 1990-01-02 to 2013-12-31. The empirical data are obtained from the CME group. These snapshots clearly demonstrate how volatile are the commodity markets. For instance, silver was traded around $4 from per ounce 1990 until mid 2005 however it peaked over $46 per ounce in mid-2011.

- **Supply and demand balancing:** Nowadays, supply and demand can be balanced at both the local and global level for major commodities with a few notable exceptions. Some commodities such as electricity, cannot be balanced globally and so electricity is traded on many local markets.

- **Diversification:** Most investors and financial institutions are increasingly considering commodities as assets and are trading commodities to diversify their portfolios.

- **Regulation and intervention:** In order to stabilize the commodity prices in domestic markets, most countries intervene in the commodity markets. For instance,

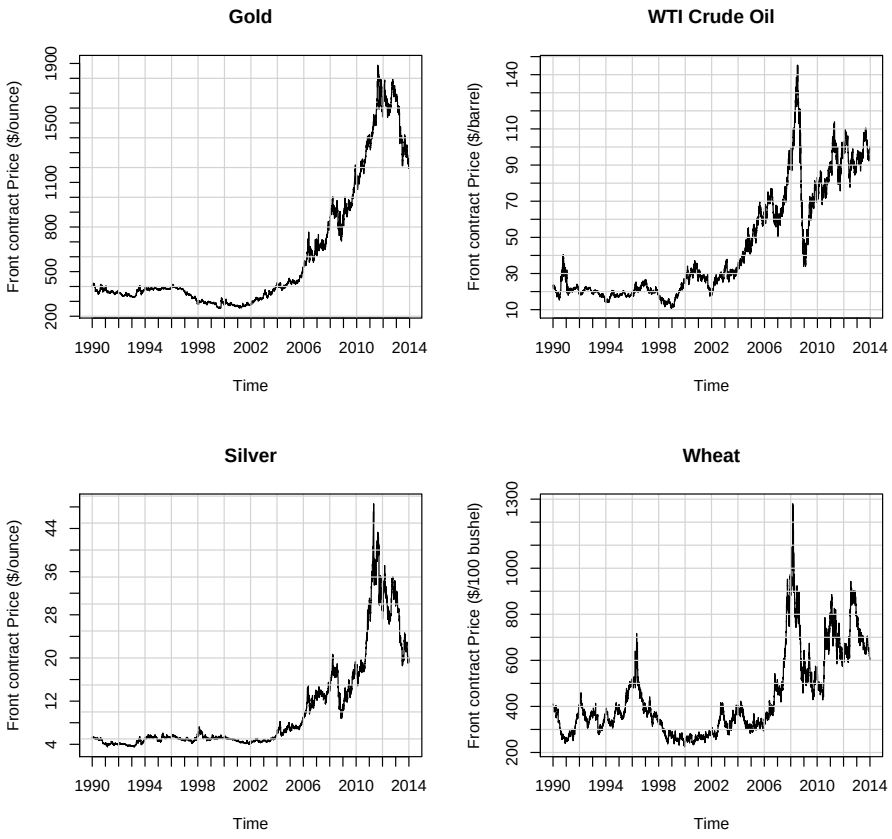

Figure 1.1: The front contracts daily futures prices for Gold, Silver, Wheat, and WTI crude oil from 1990-01-02 to 2013-12-31.

due to adverse weather conditions, in 2012, Russia expected to harvest 10-15 percent less wheat than 2011. Consequently, the Russian agricultural ministry has downgraded its export forecast by 20 percent. The Russian agricultural minister also pointed out that they might use "pinpoint intervention sales to contain domestic prices".

1.5 Commodity Markets versus Equity Markets:

Commodity markets are fast growing markets along several aspects including volume of trades, number of participants, and importance for modern economies; however, in comparison to equity markets, commodity markets are still relatively undeveloped. The volume of papers, books, and research devoted to equity markets is very high compared

to that devoted to commodity markets. The following reasons make commodity markets so different from equity markets:

- **Fundamental price drivers:** In commodity markets, fundamental price drivers are many, and sophisticated. These drivers are relatively complex to incorporate into quantitative pricing models whereas in equity markets, these drivers are comparatively few and can be easily included into pricing models. When we attempt to price financial derivatives for which the underlying is a commodity, we have to deal with various issues such as weather, storage, transportation, and technological progress. Commodities are continuously produced and consumed and stored at a storage cost, which means there is a cost to physically holding the commodity. Even this storage cost varies based on the commodity type. Consequently, the forward price models for commodities will be different from the one in equity markets and must be modified to be applied in commodity derivatives. Therefore, these issues make commodity derivatives substantial different from pure financial investment assets. In commodity markets, end users actually consume the commodity. Residential users, for instance, constantly consume energy for cooling in summer and heating in winter. Other industries need energy or other commodities to keep their production line running. Each commodity market participant, suppliers and consumers, alike have their own set of drivers that should translate into pricing models. When an event such as hurricane hits one part of the world, it will abruptly impact commodity markets. Almost none of these kinds of drivers are issues in equity markets. It is not easy to capture these drivers into quantitative models.

- **Mean reversion and economic cycle:** The economic cycle refers to the phenomenon by which an economy fluctuates between states of growth (health) to contraction (recession). The current state of the economic cycle can be inferred from various factors including gross domestic product (GDP), unemployment rates, interest rates, inflation rates, and consumer spending. The economic cycles experience four major phases: expansion, prosperity, contraction and recession. Developed countries may intervene to smooth the up peaks and down peaks. Traditional financial markets are strongly responsive to economic cycles; however, the impact of economic cycles is usually low for many commodities (Pilipovic (2007)). "Mean reversion" is a theory proposing that most of economic markets fluctuates around or move back towards a "mean" or "average" (equilibrium level). The equilibrium level can be the historical average commodity price, return for a stock or interest rates. In pure financial markets such as interest rates, mean reversion occurs due to economic cycle. However, there are totally different reasons for mean reversion in commodity markets. Commodities prices are obtained based on supply and demand. For example, when the supply is excessively higher than the demand for a particular commodity, the suppliers with high cost of production tend to lower their

production; consequently, the consumption will wipe out the excess supply with a return to the more normal price. When the supply is low, the reverse circumstances will happen. The mean reversion can also occur when events such as weather or political situations in supplying countries hit in commodity markets, we will witness a spike (upward or downward) in commodity prices in short period of time. However, the market will eventually return to "normal levels" over a longer period of time. For instance, by applying point estimations, Bessembinder *et al.* (1995) show that 44 percent of a usual front contract crude oil price spike is expected to revert to equilibrium levels over the following eight months.

- **Convenience yield:** As far back as the 1930s and 1940s, some well-known economists had noticed the benefit of the physical ownership of storable commodity in the theory of storage (Geman (2005)). Having the commodity on hand (inventories) enable agents to respond to the unpredictable, and abrupt demand and prevent manufacturing interruption. Like the stock market dividends that are only paid to the owner of stock rather than owner of derivatives such as options, this benefit is paid merely to the holder of the commodity rather than to the holder of futures or forward contracts and is known as *the convenience yield*. The convenience yield is defined as the overall benefits that holder of commodity receive minus the costs especially the cost of storage with exception of the cost of financing. When the spot commodity price unexpectedly peaked, the inventories' owners will tend to short the physical inventories and perhaps replaced by the futures or forward contracts. Conversely, when the commodity spot price dramatically drop, they will most likely decide to increase their inventories. In general, the convenience yield is measured by the existing demand and the accessible supply's balance. In chapter 3, section 3.3, we will explain these ideas in more details.

- **Seasonality:** Many commodities show undoubtable seasonality in prices. Although equity markets exhibit weak seasonality due to investment flows, during different periods of time in year, the supply and demand have dramatic changes in particular commodities in which drive seasonality in commodity markets. For instance heating oil, or agricultural commodities exhibit significant seasonality. Demand for heating oil changes due to weather patterns (demand dramatically rises in winters; conversely, demand drops in summers). seasonality in agricultural commodities is mostly driven by harvest cycles. Seasonality is implemented in models as a deterministic trend function $f(t)$ (usually as a sinusoidal function with semiannual periodicity). In chapter 4, we implement seasonality in the commodity spread model.

The differences between equity and commodity markets are not only limited to the above. Whereas equity markets, in commodity markets, frequency of events are high, liquidity

is low and mostly is decentralized markets.

1.6 Thesis Road Map

This thesis focuses on commodity pricing models, commodity spread pricing models, and optimal trading strategies for the spread processes. In commodity spread trading, we construct a portfolio of two commodities with very similar characteristics sharing common stochastic trends. The two commodities may even have identical usages (for instance, in the location spread between Brent crude oil and West Texas Intermediate (WTI) crude oil) or sometimes, the spread portfolio consists of two derivatives on the same commodity (for instance, the spread between two WTI futures prices with different delivery dates). We will discuss the spread dynamics and optimized strategies on spreads in detail later in this thesis. This thesis is divided into four main parts.

The first part, Chapter 2, describes the concepts, implications and embedded conditions of two estimation algorithms namely Kalman filter and the new local linearization, introduced by Shoji and Ozaki (1998), in detail. These approaches will be deployed to estimate parameters of the stochastic models throughout this study as relevant and are not major contributions of this research, so it can be skipped by readers already familiar with these tools.

Chapter 3 first provides an overview of forward and futures contracts, their differences, and the convenience yield, one of the important state variables in commodity markets. Second, we analyze a simple model to price a commodity, the one-factor mean-reverting commodity pricing model introduced by Schwartz (1997), and the very popular two-factor model proposed by Gibson and Schwartz (1990) to price commodities contingent claims particulary for crude oil. Finally, we introduce the new generalized one-factor commodity pricing model. We investigate the properties of this new dynamical model and show that this one-factor process is able to capture the key characteristics of the dynamics of commodity spot-prices including mean-reversion, heavy tails, skewness and kurtosis. The new stochastic process is a nonlinear process with a unique but unknown transition distribution; therefore, Shoji and Ozaki (1998)'s local linearization method will be deployed to estimate the model's parameters. The new generalized one-factor model is compared with Schwartz's one-factor dynamics by fitting these models to the WTI crude oil's front futures contracts. It will be argued which process is capable to explain the reality of the commodity spot-price process more accurately using both observed and estimated results.

Chapters 4 and 5 discusses the concepts, implications and stochastic models of

spread trading in commodity markets. First, we describe cointegration and its application in pairs trading, a one factor mean-reverting Vasicek process to model the spread process, presented by Elliott *et al.* (2005), and the two factor model proposed by Dempster *et al.* (2008) to model the spot spread process. Second, we apply these three models to our empirical sample data and compare the results. Later, in Chapter 4, we analyze the recent behavioral change in the location spread between WTI crude oil and Brent oil. Since important news can generate a shock in a spread process, we propose to implement a jump, which is compound Poisson process, in the one-factor and two-factor spread models. In these models, jump sizes follow the double exponential distribution introduced by Kou (2002). A new one-factor mean-reverting process is introduced to explain not only the mean-reverting property of the spread process, but also the skewness and the kurtosis characteristics of a spread process. The transition density of this nonlinear mean-reverting stochastic process is unknown so the new local linearization method is deployed to estimate the model's parameters. Since the spread between WTI crude oil and Brent oil has recently experienced a structural break for fundamental reasons, we deploy *Regime-Switching Models* (RSM) in this generalized one-factor mean-reverting dynamics to capture this phenomena in the spread process in Chapter 4. In Chapter 5, we will develop a novel mean-reverting random walk, obtain its continuous time dynamics, and use it to model the spread dynamics. The new mean-reverting process will be compared to Elliott *et al.* (2005)'s one factor model and its advantages and disadvantages are investigated. We will deploy both models: the new one factor mean-reverting model and the Vasicek process, to price the spread between WTI and WTS crude oils. Using both observed and estimated results, we will discuss which process can better describe the reality of the spread process.

Finally, Chapter 6 provides a discussion of specifying some optimized trading strategies for different classes of the spread dynamics. First, we empirically illustrate an existing trading strategies for the cointegration approach and Elliott *et al.* (2005)'s one-factor model. Second, we explain how two different trading strategies can be implemented in these two approaches. We review the optimal trading strategies for the cointegration approach. We also investigate the optimal trading strategies proposed by Bertram (2010) in which he exploits *first passage time results* for the Ornstein-Uhlenbeck (OU) process to find optimal upper and lower boundaries. We propose to deploy the stationary distribution to obtain optimal barriers in the Vasicek dynamics which is appropriate for the long-term investment strategy. Finally, we introduce two new trading strategies for the spread process of WTI crude oil and Brent oil by considering the empirical facts and behavioral process change.

1.7 Conclusion

In this Chapter, we review some of essential notions, concepts, and difficulties that we confront when we study the commodity markets and derivatives on the commodity such as forward or futures pricing. In Chapter 3, we will frequently use these introduced notions and consider these concepts in the commodity pricing models. But first in Chapter 2, we will develop estimation tools including the *Kalman filter algorithm* as well as a new local linearization approach to estimate parameters for SDEs.

Chapter 2

Estimating Time series using the Kalman Filter and the Local Linearization

2.1 Introduction

The Kalman filter (KF), a practical and powerful algorithm for estimating parameters for time series, was introduced by R.E. Kalman (1960) and Kalman and Bucy (1961). This model is designed to estimate parameters from noisy data on a linear system with errors modelled using Gaussian white noise. The filter recursively updates the state variables once new data becomes available. In order to apply the Kalman filter for prediction and smoothing, a dynamical system should be presented in a specific *state space* form. The state space representation is a statistical framework for unobservable variables. Due to the simplicity of applying maximum-likelihood to estimate parameters in the model, the flexibility of algorithm and the ability to manage missing values, this approach is widely applied in time series analysis and econometrics. For instance, one of the implications of commodity pricing models is that crucial state variables such as spot price and convenience yield are usually unobservable. The Kalman filter algorithm is mainly applied to calibrate unobserved state variables. The KF has been explained in many books. Tsay (2010) focus on financial time series applications, Durbin and Koopman (2001) update the approach, and Kim and Nelson (1999) applied the method in economic applications as well to regime switching models. It might be assumed that once a dynamical system is represented in state space form, the Kalman filter algorithm can easily be used to solve the resulting estimation problems. However, two difficulties can arise. First, particularly in some non-linear models, to come up with right state space forms is a challenging task. Second, un-observability of state variables is a common issue for most dynamical sys-

tems. We, therefore, confront various difficulties to apply the Kalman filter algorithm. Despite these hardships, the Kalman filter algorithm is widely used to estimate many sophisticated models. The chapter is organized as follows.

In Section 2.2 we first explain how a dynamical system can be depicted in the state space model and then we introduce the required assumptions for the KF algorithm. In Section 2.3, we derive the Kalman filter algorithm and explain classical inference problems namely filtering, smoothing and forecasting and derive their relevant formulas. We derive the log-likelihood function to estimate parameters using the Kalman filter algorithm in Section 2.4. Finally, in Section 2.4.1, we review the new local linearization method presented by Shoji and Ozaki (1998) that can be applied to estimate parameters of a large class of one-dimensional and nonlinear stochastic differential equation with unique but unknown transition densities.

2.2 Linear State space Models

Many financial time series can be formulated in the state space form including *Autoregressive Integrated Moving Average* ($ARIMA$) models, commodity pricing models, and stochastic volatility models. Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ represent a $k \times 1$ observation vector at time t. This observed vector, $\boldsymbol{y}_t$, can be stated in terms of another, possibly unobserved, $m \times 1$ vector, $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{mt})'$ known as state vector. The state vector $\boldsymbol{\alpha}_t$ usually follows a stochastic process. Also, let $\mathfrak{F}_t$ denote all information at time t. For simplicity, we consider the following convention:

$$\mathfrak{F}_0 = \emptyset, \mathfrak{F}_t = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_t\} = \{\mathfrak{F}_{t-1}, \boldsymbol{y}_t\},$$

The general Gaussian linear state-space model is denoted by the following equations system:

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t, \tag{2.1}$$

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \tag{2.2}$$

where $\boldsymbol{d}_t$ $(m \times 1)$ and $\boldsymbol{c}_t$ $(k \times 1)$ are deterministic vectors,
$\boldsymbol{T}_t$ $(m \times m)$ and $\boldsymbol{Z}_t$ $(k \times m)$ are coefficient matrices,
$\boldsymbol{R}_t$ is a $m \times n$ matrix and
The observation error vector, $\boldsymbol{\varepsilon}_t$ $(n \times 1)$ and process error vector, $\boldsymbol{\eta}_t$ $(k \times 1)$ are Gaussian white noises. More specifically:

$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}_t)$, $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_t)$,

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_\tau^{'}) = \begin{cases} \boldsymbol{H}_t & \text{for } t = \tau \\ \mathbf{0} & \text{elsewhere} \end{cases},$$

$$\text{and, } E(\boldsymbol{\eta}_t \boldsymbol{\eta}_\tau^{'}) = \begin{cases} \boldsymbol{Q}_t & \text{for } t = \tau \\ \mathbf{0} & \text{elsewhere} \end{cases},$$

We assume that $\boldsymbol{Q}_t\,(n \times n)$ and $\boldsymbol{H}_t\,(k \times k)$ are independent positive-definite matrices; however,the independence condition can be relaxed at the cost of some additional conditions as described by Durbin and Koopman (2001).
Equation 2.2 describe the relation at time t between the observation variables $\boldsymbol{y}_t$ and the state variables $\boldsymbol{\alpha}_t$. This equation is called the observation or measurement equation with measurement disturbance $\boldsymbol{\varepsilon}_t$. Equation 2.1 generates the transition of the state variable $\boldsymbol{\alpha}_t$ from period t to period $t+1$ with innovation $\boldsymbol{\eta}_t$, and it is known as the state or transition equation. It is also a first-order Markov chain, given the above assumptions.
The matrices $\boldsymbol{T}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Z}_t$, $\boldsymbol{Q}_t$, and $\boldsymbol{H}_t$ can be functions of some parameters $\boldsymbol{\theta}$ as well as time t. One can estimate the parameters in the matrices using the maximum-likelihood approach as described by Tsay (2010).

It is assumed that the initial state, $\boldsymbol{\alpha}_1$ is normally distributed with known mean vector and covariance matrix $\boldsymbol{\alpha}_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and is independent of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ for $t > 0$. The state space form for a given dynamic process is not usually unique. For some cases, there are many approaches to find the state space form. However, in some particular cases, finding the state space form can be quite challenging. We present one example to show how we can handle this model.

Example 2.2.1. *State space representation for autoregressive moving-average model (ARMA(p, q)) process:* As we mentioned before, state space forms are not unique. In case of $ARMA$, there are many approaches to express in state space form such as Akaike, Harvey, and Aoki' s approaches. In this example, we explain Harvey's approach for zero mean $ARMA(p,q)$ Harvey (1993).
First consider the $ARMA(p,q)$ process expressed by:

$$\boldsymbol{\phi}(\boldsymbol{B})\boldsymbol{y}_t = \boldsymbol{\theta}(\boldsymbol{B})\boldsymbol{a}_t. \quad t = 0, \pm 1, \ldots \tag{2.3}$$

where $\boldsymbol{\phi}(\boldsymbol{B}) = 1 - \sum_{i=1}^{p} \phi_i \boldsymbol{B}^i$ and $\boldsymbol{\theta}(\boldsymbol{B}) = 1 - \sum_{i=1}^{q} \theta_i \boldsymbol{B}^i$ ($\boldsymbol{B}$ is back-shift operator, $\boldsymbol{B}^i\,\boldsymbol{a}_t = \boldsymbol{a}_{t-i}$) $\{a_t\}$ is a Gaussian white noise series ($\boldsymbol{a}_t \sim \mathcal{WN}(\mu_1, \sigma_a^2)$) and p and q are nonnegative integers.
Consider $m = \max(p, q+1)$; as a result, the $ARMA$ can be rewritten as follows:

$$y_t = \sum_{i=1}^{m} \phi_i y_{t-i} + a_t - \sum_{j=1}^{m-1} \theta_j a_{t-j}. \tag{2.4}$$

where $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$; as a result, the model is denoted as a $ARMA(m, m-1)$ for which some of $\phi_i's$, $\theta_j's$ are zero. Harvey (1993) introduce a state space representation as follows:
He defines state vector $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{mt})'$ in which the first element is $\alpha_{1t} = y_t$ and the rest of the elements can be obtained recursively from the $ARMA(m, m-1)$ model as follows:
Step 1:

$$\begin{aligned} y_{t+1} &= \phi_1 y_t + \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j} + a_{t+1} \\ &= \phi_1 \alpha_{1t} + \alpha_{2t} + \eta_t. \end{aligned} \tag{2.5}$$

where $\alpha_{1t} = y_t$, $\alpha_{2t} = \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j}$ and $\eta_t = a_{t+1}$.

Step 2: Now let us consider $\alpha_{2,t+1}$:

$$\begin{aligned} \alpha_{2,t+1} &= \sum_{i=2}^{m} \phi_i y_{t+2-i} - \sum_{j=1}^{m-1} \theta_j a_{t+2-j} \\ &= \phi_2 y_t + \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j} - \theta_1 a_{t+1} \\ &= \phi_2 \alpha_{1t} + \alpha_{3t} + (-\theta_1)\eta_t. \end{aligned} \tag{2.6}$$

where $\alpha_{3t} = \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j}$.

If we follow steps 1 and 2 repeatedly, we can conclude by induction that α_{lt} and $\alpha_{l,t+1}$ for $(l < m)$ will be given as follows:

$$\begin{aligned} \alpha_{lt} &= \sum_{i=l}^{m} \phi_i y_{t+l-1-i} - \sum_{j=l-1}^{m-1} \theta_j a_{t+l-1-j} \\ \alpha_{l,t+1} &= \phi_l y_t + \sum_{i=l+1}^{m} \phi_i y_{t+l-i} - \sum_{j=l}^{m-1} \theta_j a_{t+l-j} + (-\theta_{l-1}) a_{t+1} \\ &= \phi_l \alpha_{1t} + \alpha_{lt} + (-\theta_{l-1})\eta_t, \end{aligned} \tag{2.7}$$

Finally, $\alpha_{m,t+1}$, is:

$$\alpha_{l,t+1} = \phi_m \alpha_{1t} + (-\theta_{m-1})\eta_t. \tag{2.8}$$

and using the above equations 2.7 , 2.8, it is easy to show that the state space form is:

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}\boldsymbol{\eta}_t, \tag{2.9}$$

$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{\alpha}_t \quad (\boldsymbol{\eta}_t \sim \mathcal{N}(0, \sigma_a^2)). \tag{2.10}$$

where $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{mt})'$, $\boldsymbol{T}$, $\boldsymbol{R}$ and $\boldsymbol{Z}$ are time invariant and are:

$\boldsymbol{Z}_t = (1, 0, 0, \ldots, 0)_{(1\times m)}$,

$$T_{(m\times m)} = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \phi_{m-1} & 0 & 0 & \cdots & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$R_{(m\times 1} = \begin{pmatrix} 1 \\ -\theta_1 \\ -\theta_2 \\ \vdots \\ -\theta_{m-1} \end{pmatrix},$$

In this setting, there is no measurement noise and all the system matrices are constructed by using $ARMA(m, m-1)$ coefficients.

The results of Theorem 2.2.1 is extensively used in the derivation of the Kalman filter algorithm.

Theorem 2.2.1. Suppose that $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, are random vectors such that their joint distributions are multivariate normal with means $\boldsymbol{\mu}_p$ and covariance matrices $\boldsymbol{\Sigma}_{pp}$ where $\boldsymbol{\Sigma}_{pp}$ is nonsingular for $p = x, y, z$ and $\boldsymbol{\Sigma}_{yz}$=$\mathbf{0}$. In this case, we have, as described in Durbin and Koopman (2001):

(i) $E[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\ \boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}\ -\ \boldsymbol{\mu}_y)$

(ii) $Var[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\ \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{xy}'$,

(iii) $E[\ \boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{z}\ ] = E[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] + \boldsymbol{\Sigma}_{xz}\ \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{z}\ -\ \boldsymbol{\mu}_z)$,

(iv) $Var[\ \boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{z}\ ] = Var[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] - \boldsymbol{\Sigma}_{xz}\ \boldsymbol{\Sigma}_{zz}^{-1}\ \boldsymbol{\Sigma}_{xz}'$.

(i) *and* (ii) *are standard results. To prove* (iii)*, we apply* (i) *to vector* $(\boldsymbol{y}, \boldsymbol{z})'$ *which is normal with following mean and covariance matrix:*

$$mean = (\boldsymbol{\mu}_y \ \boldsymbol{\mu}_z), \quad and \ covariance \ matrix = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz} \end{bmatrix}.$$

Since $\boldsymbol{y}$ *and* $\boldsymbol{z}$ *are independent multivariate normal, we have:*

$$\begin{aligned} E[\ \boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{z}\ ] &= \boldsymbol{\mu}_x + \begin{bmatrix} \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{yy}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} - \boldsymbol{\mu}_y \\ \boldsymbol{z} - \boldsymbol{\mu}_z \end{bmatrix} \\ &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y) + \boldsymbol{\Sigma}_{xz}\, \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z) \\ &= E[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] + \boldsymbol{\Sigma}_{xz}\, \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z) \end{aligned}$$

And from condition (ii)*:*

$$\begin{aligned} Var(\ \boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{z}\ ) &= \boldsymbol{\Sigma}_{xx} + \begin{bmatrix} \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{yy}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{xy}^{'} \\ \boldsymbol{\Sigma}_{xz}^{'} \end{bmatrix} \\ &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{xy}^{'} - \boldsymbol{\Sigma}_{xz}\, \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{xz}^{'} \\ &= Var[\ \boldsymbol{x} \mid \boldsymbol{y}\ ] - \boldsymbol{\Sigma}_{xz}\, \boldsymbol{\Sigma}_{zz}^{-1}\, \boldsymbol{\Sigma}_{xz}^{'}. \end{aligned}$$

2.3 Classical inference:

The main objective of a state-space model is to study the evolution of the unobserved state vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\cdots$, $\boldsymbol{\alpha}_n$ for which we use the observed time series vectors $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\cdots$, $\boldsymbol{y}_n$. In the Kalman filter algorithm, we discuss three types of important typical inference problem namely filtering, forecasting, and smoothing:

(i) Filtering means to forecast the state variable $\boldsymbol{\alpha}_t$ given all available information at time t ($\mathfrak{F}_t$). This is because $\boldsymbol{\alpha}_t$ is usually unobservable and we estimate $\boldsymbol{\alpha}_t$ by eliminating the measurement noise from the data,

(ii) Smoothing means to estimate the state variable $\boldsymbol{\alpha}_t$ given all available information at time T ($\mathfrak{F}_T$) where $(T > t)$,

(iii) Forecasting means to predict the state variable $\boldsymbol{\alpha}_{t+h}$ or $\boldsymbol{y}_{t+h}$ given all available information at time t ($\mathfrak{F}_t$) where $h > 0$ and t is the forecasting origin.

We define some notation that we deploy to derive the Kalman algorithm as follows:

- $\mathfrak{F}_t = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_t\} = \{\mathfrak{F}_{t-1}, \boldsymbol{y}_t\}$ be all information available at time t,
- $\boldsymbol{\alpha}_{t|j} = E[\ \boldsymbol{\alpha}_t \mid \mathfrak{F}_j]$, $\boldsymbol{y}_{t|j} = E[\ \boldsymbol{y}_t \mid \mathfrak{F}_j]$ be the conditional means of $\boldsymbol{\alpha}_t$ and $\boldsymbol{y}_t$ given $\mathfrak{F}_j$ respectively,
- $\boldsymbol{\Sigma}_{t|j} = Var[\ \boldsymbol{\alpha}_t \mid \mathfrak{F}_j]$ be conditional variance of $\boldsymbol{\alpha}_t$ given $\mathfrak{F}_j$,
- $\boldsymbol{\nu}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1}$ be the 1-step-ahead forecasting error,
- $\boldsymbol{V}_t = Var[\ \boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1}]$ denotes the conditional covariance matrix of the 1-step-ahead forecast error ($\boldsymbol{\nu}_t$) given $\mathfrak{F}_{t-1}$

Later in this chapter, we will derive all the above mentioned inference problems.

2.3.1 Preliminary Derivations:

Now, we attempt to derive all the model involved elements. Based on the model assumptions, the prediction error $\boldsymbol{\nu}_t$ is independent of $\mathfrak{F}_{t-1}$; as a result, $Var[\ \boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1}] = Var[\ \boldsymbol{\nu}_t\ ]$. Since all distributions in the model are assumed normal, all conditional involved distributions are also normal and we need only study the mean and covariance matrix to determine the distribution of $\boldsymbol{\alpha}_t$. From equations 2.1 and 2.2, we have:

$$\begin{aligned} \boldsymbol{y}_{t|t-1} &= E[\ \boldsymbol{y}_t \mid \mathfrak{F}_{t-1}] = E[\ (\boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t) \mid \mathfrak{F}_{t-1}] \\ &= \boldsymbol{c}_t + \boldsymbol{Z}_t E[\ \boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}] + E[\ \boldsymbol{\varepsilon}_t \mid \mathfrak{F}_{t-1}] \\ &= \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_{t|t-1}. \end{aligned} \tag{2.11}$$

And we have:

$$\begin{aligned} \boldsymbol{\alpha}_{t+1|t} &= E[\ \boldsymbol{\alpha}_{t+1} \mid \mathfrak{F}_t] = E[\ (\boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t) \mid \mathfrak{F}_t] \\ &= \boldsymbol{d}_t + \boldsymbol{T}_t E[\ \boldsymbol{\alpha}_t \mid \mathfrak{F}_t] + \boldsymbol{R}_t E[\ \boldsymbol{\eta}_t \mid \mathfrak{F}_t] \\ &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_{t|t}. \end{aligned} \tag{2.12}$$

Also, based on the model assumptions, we have:

$$\begin{aligned} \boldsymbol{\Sigma}_{t+1|t} &= Var[\ \boldsymbol{\alpha}_{t+1} \mid \mathfrak{F}_t] = Var[\ (\boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t) \mid \mathfrak{F}_t] \\ &= \boldsymbol{T}_t Var[\ \boldsymbol{\alpha}_t \mid \mathfrak{F}_t]\boldsymbol{T}_t^{'} + \boldsymbol{R}_t Var[\ \boldsymbol{\eta}_t \mid \mathfrak{F}_t]\boldsymbol{R}_t^{'} \\ &= \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{T}_t^{'} + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t^{'}. \end{aligned} \tag{2.13}$$

By definition and equations 2.2 and 2.11 , we have:

$$\begin{aligned} \boldsymbol{\nu}_t &= \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1} = \boldsymbol{y}_t - (\boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_{t|t-1}) \\ &= \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t - (\boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_{t|t-1}) \\ &= \boldsymbol{Z}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t. \end{aligned} \tag{2.14}$$

We also have:

$$\begin{aligned} E[\,\boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1}] &= E[\,(\,\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1}\,) \mid \mathfrak{F}_{t-1}] \\ &= E[\,\boldsymbol{y}_t \mid \mathfrak{F}_{t-1}] - \boldsymbol{y}_{t|t-1} = \boldsymbol{y}_{t|t-1} - \boldsymbol{y}_{t|t-1} = \mathbf{0}, \end{aligned} \tag{2.15}$$

$$\begin{aligned} Cov(\,\boldsymbol{\nu}_t \;\; \boldsymbol{y}_j\,) &= E[\boldsymbol{\nu}_t\,\boldsymbol{y}_j] = E[\,E(\boldsymbol{\nu}_t\,\boldsymbol{y}_j \mid \mathfrak{F}_{t-1})] \\ &= E[\boldsymbol{y}_j E(\boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1})] = E[\boldsymbol{y}_j \times \mathbf{0}] = \mathbf{0}, \quad for\ 1 \le j < t. \end{aligned} \tag{2.16}$$

Therefore, $\boldsymbol{\nu}_t$ is independent of $\mathfrak{F}_{t-1}$, and from equation 2.14, we have:

$$\begin{aligned} \boldsymbol{V}_t &= Var[\,\boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1}] = Var(\,\boldsymbol{\nu}_t\,) \\ &= Var[\,\boldsymbol{Z}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t\,] \\ &= \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t^{'} + \boldsymbol{H}_t. \end{aligned} \tag{2.17}$$

By using Theorem 2.2.1 and $\mathfrak{F}_t = \{\mathfrak{F}_{t-1}, \boldsymbol{y}_t\} = \{\mathfrak{F}_{t-1}, \boldsymbol{\nu}_t\}$, we can derive:

$$\begin{aligned} \boldsymbol{\alpha}_{t|t} &= E[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_t\,] = E[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,,\,\boldsymbol{\nu}_t\,] \\ &= E[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,] + Cov(\,\boldsymbol{\alpha}_t\;,\;\boldsymbol{\nu}_t\,)Var(\,\boldsymbol{\nu}_t\,)^{-1}(\,\boldsymbol{\nu}_t - \mathbf{0}\,) \\ &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{C}_t,\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t. \end{aligned} \tag{2.18}$$

where $\boldsymbol{C}_t = Cov(\,\boldsymbol{\alpha}_t\;,\;\boldsymbol{\nu}_t\,)$,
As a result, we apply the above equations 2.18 to derive:

$$\begin{aligned} \boldsymbol{C}_t &= Cov(\,\boldsymbol{\alpha}_t\;,\;\boldsymbol{\nu}_t\,) = E[\,\boldsymbol{\alpha}_t\boldsymbol{\nu}_t^{'}] \\ &= E[\,E(\,\boldsymbol{\alpha}_t\boldsymbol{\nu}_t^{'} \mid \mathfrak{F}_{t-1}\,)] \\ &= E[\,E[\,\boldsymbol{\alpha}_t(\boldsymbol{Z}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t)^{'} \mid \mathfrak{F}_{t-1}]\,] \\ &= E[\,E[\,\boldsymbol{\alpha}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1})^{'}\boldsymbol{Z}_t^{'} \mid \mathfrak{F}_{t-1}]\,] \\ &= E[\,E[\,\boldsymbol{\alpha}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1})^{'} \mid \mathfrak{F}_{t-1}]\,]\,\boldsymbol{Z}_t^{'} \\ &= \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'} \end{aligned} \tag{2.19}$$

Since $\boldsymbol{H}_t$ is assumed nonsingular, $\boldsymbol{V}_t$ is nonsingular too. This assumption can sometimes be relaxed; see Durbin and Koopman (2001). Applying equations 2.12 and 2.18, we can obtain:

$$\begin{aligned} \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_{t|t} \\ &= \boldsymbol{d}_t + \boldsymbol{T}_t(\,\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{C}_t\,\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t) \\ &= \boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{T}_t\,\boldsymbol{C}_t\,\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t \\ &= \boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{K}_t\,\boldsymbol{\nu}_t, \quad for\ t = 1, 2, \ldots, n. \end{aligned} \tag{2.20}$$

With

$$\boldsymbol{K}_t \;=\; \boldsymbol{T}_t\,\boldsymbol{C}_t\,\boldsymbol{V}_t^{-1} \;=\; \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}, \tag{2.21}$$

in equation 2.20, we obtain $\boldsymbol{\alpha}_{t+1|t}$ as a linear function of $\boldsymbol{\alpha}_{t|t-1}$ and $\boldsymbol{\nu}_t$. $\boldsymbol{K}_t$ is called the Kalman gain at time t. Using Theorem 2.2.1 , we obtain:

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t} &= Var[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_t] = Var[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,,\,\boldsymbol{\nu}_t\,] \\
&= Var[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}] \;-\; Cov(\,\boldsymbol{\alpha}_t\;,\;\boldsymbol{\nu}_t\,)Var(\,\boldsymbol{\nu}_t\,)^{-1}Cov(\,\boldsymbol{\alpha}_t\;,\;\boldsymbol{\nu}_t\,)^{'} \\
&= \boldsymbol{\Sigma}_{t|t-1} \;-\; \boldsymbol{C}_t\,\boldsymbol{V}_t^{-1}\,\boldsymbol{C}_t^{'} \\
&= \boldsymbol{\Sigma}_{t|t-1} \;-\; \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{Z}_t\,\boldsymbol{\Sigma}_{t|t-1}.
\end{aligned} \tag{2.22}$$

Substituting equation 2.22 into equation 2.13, we have:

$$\begin{aligned}
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t(\,\boldsymbol{\Sigma}_{t|t-1} \;-\; \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{Z}_t\,\boldsymbol{\Sigma}_{t|t-1}\,)\boldsymbol{T}_t^{'} \;+\; \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'} \\
&= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,(\,\boldsymbol{T}_t \;-\; \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\;\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{Z}_t\,)^{'} \;+\; \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'} \\
&= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,(\,\boldsymbol{T}_t \;-\; \boldsymbol{K}_t\,\boldsymbol{Z}_t\,)^{'} \;+\; \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'} \\
&= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'} \;+\; \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'}, \quad for\ t = 1, 2, \ldots, n,
\end{aligned} \tag{2.23}$$

where

$$\boldsymbol{L}_t \;=\; \boldsymbol{T}_t \;-\; \boldsymbol{K}_t\,\boldsymbol{Z}_t.$$

2.3.2 The Kalman Filter Recursion:

Considered together, the previous derived equations give the Kalman filter for the general state-space model defined in equations 2.1 and 2.2. These equations provide us the means to recursively revise our knowledge of the state space system each time new observations are revealed. Collecting the derived equations and assuming that mean vector, $\boldsymbol{\alpha}_{1|0}$ and the variance covariance matrix, $\boldsymbol{\Sigma}_{1|0}$ for the initial state are given, we have the Kalman filter algorithm as follows:

$$\begin{aligned}
\boldsymbol{\nu}_t &= \boldsymbol{y}_t \;-\; \boldsymbol{c}_t \;-\; \boldsymbol{Z}_t\boldsymbol{\alpha}_{t|t-1}, \\
\boldsymbol{V}_t &= \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t^{'} \;+\; \boldsymbol{H}_t, \\
\boldsymbol{K}_t &= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}, \\
\boldsymbol{L}_t &= \boldsymbol{T}_t \;-\; \boldsymbol{K}_t\,\boldsymbol{Z}_t, \\
\boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_{t|t-1} \;+\; \boldsymbol{K}_t\,\boldsymbol{\nu}_t, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'} \;+\; \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'}, \quad for\ t = 1, 2, \ldots, n.
\end{aligned} \tag{2.24}$$

Durbin and Koopman (2001) prove that when the normality assumption of the observations do not hold, but $\boldsymbol{Z}_t$ and $\boldsymbol{T}_t$ are independent of previous $\boldsymbol{y}_t$'s, $\boldsymbol{\alpha}_{t+1|t}$, the estimate of $\boldsymbol{\alpha}_{t+1}$, minimizes the mean square-error under some appropriate assumptions. It is worth emphasizing that if we relax the normality assumption for the involved variables, the resulting estimate is still statistically optimal in the sense of minimum mean-square errors. In the Kalman filter algorithm, sometimes we also attempt to estimate filtered values, $\boldsymbol{\alpha}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}$; therefore, we add "contemporaneous" filtering equations 2.18 and 2.22 to the algorithm. The modified algorithm is:

$$\begin{aligned}
\boldsymbol{\nu}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t\boldsymbol{\alpha}_{t|t-1}, \\
\boldsymbol{C}_t &= \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}, \\
\boldsymbol{V}_t &= \boldsymbol{Z}_t\,\boldsymbol{C}_t + \boldsymbol{H}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{C}_t,\, \boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{Z}_t\,\boldsymbol{\Sigma}_{t|t-1}, \\
\boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_{t|t}, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'} + \boldsymbol{R}_t\,\boldsymbol{Q}_t\,\boldsymbol{R}_t^{'}, \quad for\ t = 1, 2, \ldots, n.
\end{aligned} \tag{2.25}$$

We should emphasize that the only difficult step in the Kalman filter algorithm is to calculate the determinant and inverse of $\boldsymbol{V}_t$. If we fail to compute either determinant or inverse of $\boldsymbol{V}_t$, the algorithm will halt.

2.3.3 State Smoothing:

As we mentioned before smoothing means to estimate $\boldsymbol{\alpha}_t$ given $\mathfrak{F}_T\,(\,\hat{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_{t|T}\,)$ and its associated error variance matrix $\boldsymbol{\Sigma}_{t|T}$ when $t < T$. In other word, smoothing means to estimate the unobserved $\boldsymbol{\alpha}_t$ given the entire observations series $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_T\}$.
Now let $\boldsymbol{x}_t$ be the state forecast error which is:

$$\boldsymbol{x}_t \;=\; \boldsymbol{\alpha}_t \;-\; \boldsymbol{\alpha}_{t|t-1}.$$

Therefore, the variance matrix of $\boldsymbol{x}_t$ is:

$$Var(\,\boldsymbol{x}_t \mid \mathfrak{F}_{t-1}) \;=\; Var[\,\boldsymbol{\alpha}_t \;-\; \boldsymbol{\alpha}_{t|t-1} \mid \mathfrak{F}_{t-1}] \;=\; Var(\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,) \;=\; \boldsymbol{\Sigma}_{t|t-1}$$

Using equations 2.1 , 2.20 and 2.14, we have:

$$\begin{aligned}
\boldsymbol{x}_{t+1} &= \boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_{t+1|t} \\
&= \boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_t + \boldsymbol{R}_t\,\boldsymbol{\eta}_t - (\,\boldsymbol{d}_t + \boldsymbol{T}_t\,\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{K}_t\,\boldsymbol{\nu}_t\,) \\
&= \boldsymbol{T}_t\,(\,\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}\,) + \boldsymbol{R}_t\,\boldsymbol{\eta}_t - \boldsymbol{K}_t\,\boldsymbol{\nu}_t \\
&= \boldsymbol{T}_t\,\boldsymbol{x}_t + \boldsymbol{R}_t\,\boldsymbol{\eta}_t - \boldsymbol{K}_t\,(\,\boldsymbol{Z}_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t) \\
&= \boldsymbol{L}_t\,\boldsymbol{x}_t + \boldsymbol{R}_t\,\boldsymbol{\eta}_t - \boldsymbol{K}_t\,\boldsymbol{\varepsilon}_t.
\end{aligned} \tag{2.26}$$

where

$$\boldsymbol{L}_t \;=\; \boldsymbol{T}_t \;-\; \boldsymbol{K}_t\,\boldsymbol{Z}_t,$$

As a result, the state space form for $\boldsymbol{\nu}_t$ is :

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{L}_t\,\boldsymbol{x}_t \;+\; \boldsymbol{R}_t\,\boldsymbol{\eta}_t \;-\; \boldsymbol{K}_t\,\boldsymbol{\varepsilon}_t, && (2.27)\\ \boldsymbol{\nu}_t &= \boldsymbol{Z}_t\,\boldsymbol{x}_t \;+\; \boldsymbol{\varepsilon}_t. && (2.28) \end{aligned}$$

where

$$\boldsymbol{x}_1 \;=\; \boldsymbol{\alpha}_1 \;-\; \boldsymbol{\alpha}_{1|0},$$

Note that in equations 2.15 and 2.16 , we show that 1-step-ahead forecast errors $\{\boldsymbol{\nu}_t\}$ are independent of each other and the forecast errors, $\boldsymbol{\nu}_t$ is also independent of $\mathfrak{F}_{t-1}$.
Now we focus on state smoothing:
We wish to estimate $\boldsymbol{\alpha}_{t|T}$ where $(T > t)$. By applying Theorem 2.2.1 to the joint distribution of $\boldsymbol{\alpha}_t$ and

$\{\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t+1}, \cdots, \boldsymbol{\nu}_T\}$ given $\mathfrak{F}_{t-1}$, we have:

$$\begin{aligned} \boldsymbol{\alpha}_{t|T} &= E[\,\boldsymbol{\alpha}_t \,|\, \mathfrak{F}_{t-1}\,,\, \boldsymbol{\nu}_t, \cdots, \boldsymbol{\nu}_T\,]\\ &= E[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,] + \sum_{j=1}^{T} Cov(\,\boldsymbol{\alpha}_t\,,\,\boldsymbol{\nu}_j\,)Var(\,\boldsymbol{\nu}_j\,)^{-1}\,\boldsymbol{\nu}_j\\ &= E[\,\boldsymbol{\alpha}_t \mid \mathfrak{F}_{t-1}\,] + \sum_{j=1}^{T} Cov(\,\boldsymbol{\alpha}_t\,,\,\boldsymbol{\nu}_j\,)\,\boldsymbol{V}_j^{-1}\,\boldsymbol{\nu}_j. \qquad (2.29) \end{aligned}$$

To obtain $Cov(\,\boldsymbol{\alpha}_t\,,\,\boldsymbol{\nu}_j\,)$, we use equation 2.28 as follows:

$$\begin{aligned} Cov(\,\boldsymbol{\alpha}_t\,,\,\boldsymbol{\nu}_j\,) &= E[\,\boldsymbol{\alpha}_t\,\boldsymbol{\nu}_j^{'}\,]\\ &= E[\,\boldsymbol{\alpha}_t\,(\,\boldsymbol{Z}_j\,\boldsymbol{x}_j \;+\; \boldsymbol{\varepsilon}_j\,)^{'}\,]\\ &= E[\,\boldsymbol{\alpha}_t\,\boldsymbol{x}_j^{'}\,]\,\boldsymbol{Z}_j^{'}, \quad for\ j \;=\; 1, 2, \ldots, n. \qquad (2.30) \end{aligned}$$

Also, using equation 2.27, we have:

$$\begin{aligned}
E[\,\boldsymbol{\alpha}_t\,\boldsymbol{x}_t^{'}\,] &= E[\,\boldsymbol{\alpha}_t\,(\,\boldsymbol{\alpha}_t\,-\,\boldsymbol{\alpha}_{t|t-1}\,)^{'}\,] \\
&= Var(\,\boldsymbol{\alpha}_t\,)\,=\,\boldsymbol{\Sigma}_{t|t-1}, \\
E[\,\boldsymbol{\alpha}_t\,\boldsymbol{x}_{t+1}^{'}\,] &= E[\,\boldsymbol{\alpha}_t\,(\,\boldsymbol{L}_t\,\boldsymbol{x}_t\,+\,\boldsymbol{R}_t\,\boldsymbol{\eta}_t\,-\,\boldsymbol{K}_t\,\boldsymbol{\varepsilon}_t\,)^{'}\,] \\
&= \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'}, \\
E[\,\boldsymbol{\alpha}_t\,\boldsymbol{x}_{t+2}^{'}\,] &= \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'}\,\boldsymbol{L}_{t+1}^{'}, \\
&\vdots \\
E[\,\boldsymbol{\alpha}_t\,\boldsymbol{x}_T^{'}\,] &= \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'}\,\boldsymbol{L}_{t+1}^{'}\cdots\boldsymbol{L}_{T-1}^{'}.
\end{aligned} \tag{2.31}$$

Note that all the above expressions are conditional on $\mathfrak{F}_{t-1}$.
By substituting the previous equations 2.31 and 2.30 into equation 2.29, we obtain:

$$\begin{aligned}
\boldsymbol{\alpha}_{T|T} &= \boldsymbol{\alpha}_{T|T-1}\,+\,\boldsymbol{\Sigma}_{T|T-1}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T, \\
\boldsymbol{\alpha}_{T-1|T} &= \boldsymbol{\alpha}_{T-1|T-2}\,+\,\boldsymbol{\Sigma}_{T-1|T-2}\,\boldsymbol{Z}_{T-1}^{'}\,\boldsymbol{V}_{T-1}^{-1}\,\boldsymbol{\nu}_{T-1} \\
&\quad +\,\boldsymbol{\Sigma}_{T-1|T-2}\,\boldsymbol{L}_{T-1}^{'}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T, \\
&\vdots \\
\boldsymbol{\alpha}_{t|T} &= \boldsymbol{\alpha}_{t|t-1}\,+\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{\nu}_t \\
&\quad +\,\boldsymbol{\Sigma}_{T-1|T-2}\,\boldsymbol{L}_{T-1}^{'}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T\,+\cdots \\
&\quad +\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{L}_t^{'}\,\boldsymbol{L}_{t+1}^{'}\cdots\boldsymbol{L}_{T-1}^{'}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T, \\
&\quad\quad for\ t\,=\,T-2,T-3,\ldots,1.
\end{aligned} \tag{2.32}$$

Note: When $t\,=\,T$, $\boldsymbol{L}_t^{'}\,\boldsymbol{L}_{t+1}^{'}\cdots\boldsymbol{L}_T^{'}\,=\,\boldsymbol{I}_m$.
The smoothed state vectors, $\boldsymbol{\alpha}_{t|T}$, can be depicted as follows:

$$\boldsymbol{\alpha}_{t|T}\,=\,\boldsymbol{\alpha}_{t|t-1}\,+\,\boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{q}_t. \tag{2.33}$$

where

$$\begin{aligned}
\boldsymbol{q}_{T-1} &= \boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T, \\
\boldsymbol{q}_{T-2} &= \boldsymbol{Z}_{T-1}^{'}\,\boldsymbol{V}_{T-1}^{-1}\,\boldsymbol{\nu}_{T-1}\,+\,\boldsymbol{L}_{T-1}^{'}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T, \\
&\vdots \\
\boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{\nu}_t\,+\,\boldsymbol{L}_t^{'}\,\boldsymbol{Z}_{t+1}^{'}\,\boldsymbol{V}_{t+1}^{-1}\,\boldsymbol{\nu}_{t+1}\,+\cdots \\
&\quad +\,\boldsymbol{L}_t^{'}\,\boldsymbol{L}_{t+1}^{'}\cdots\boldsymbol{L}_{T-1}^{'}\,\boldsymbol{Z}_T^{'}\,\boldsymbol{V}_T^{-1}\,\boldsymbol{\nu}_T,\quad for\ t\,=\,T-2,T-3,\ldots,1.
\end{aligned} \tag{2.34}$$

By applying equations 2.34, we can obtain the smoothed state vectors, $\boldsymbol{\alpha}_{t|T}$, by using backward recursion as follows:

$$\begin{aligned} \boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t^{'}\,\boldsymbol{V}_t^{-1}\,\boldsymbol{\nu}_t + \boldsymbol{L}_t^{'}\,\boldsymbol{q}_t, \\ \boldsymbol{\alpha}_{t|T} &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\,\boldsymbol{q}_t, \quad for\ t = T, T-1, \ldots, 1. \end{aligned} \tag{2.35}$$

where $\boldsymbol{q}_T = 0$.
It is worth mentioning that $\boldsymbol{\alpha}_{t|t-1}$, $\boldsymbol{\Sigma}_{t|t-1}$, $\boldsymbol{L}_t$, and $\boldsymbol{V}_t$ are outputs of the Kalman filter algorithm.

2.3.4 Forecasting:

In forecasting, we would like to predict the state variable $\boldsymbol{y}_{t+i}$ given all available information at time t $(\mathfrak{F}_t)$ for $i = 1, 2, \ldots, h$ where $h > 0$ and t is the time at which the forecast is made. It is easy to show that the i-step-ahead forecast, $\hat{\boldsymbol{y}}_{t+i}$ is the expectation of $\boldsymbol{y}_{t+i}$ given $(\mathfrak{F}_t)$ which is $\hat{\boldsymbol{y}}_{t+i} = E[\,\boldsymbol{y}_{t+i} \mid \mathfrak{F}_t]$. By deploying the Kalman filter in equation 2.24, we derive these predictions, $\hat{\boldsymbol{y}}_{t+i}$, and the covariance matrices of their prediction errors by considering $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_t\}$ as missing values (Tsay (2010)) as follows:
By using the equation 2.11 , the 1-step-ahead prediction is:

$$\begin{aligned} \hat{\boldsymbol{y}}_{t+1} &= E[\,\boldsymbol{y}_{t+1} \mid \mathfrak{F}_t] \\ &= \boldsymbol{c}_{t+1} + \boldsymbol{Z}_{t+1}\,\boldsymbol{\alpha}_{t+1|t}, \end{aligned} \tag{2.36}$$

which $\boldsymbol{\alpha}_{t+1|t}$ is the output of the Kalman filter. Its prediction error is:

$$\begin{aligned} \hat{\boldsymbol{e}}_{t+1} &= \boldsymbol{y}_{t+1} - \hat{\boldsymbol{y}}_{t+1} \\ &= \boldsymbol{Z}_{t+1}\,(\,\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_{t+1|t}) + \boldsymbol{\varepsilon}_{t+1}. \end{aligned} \tag{2.37}$$

Using equation 2.37, the covariance matrix of the 1-step-ahead of prediction errors is:

$$\begin{aligned} Var[\,\hat{\boldsymbol{e}}_{t+1}\,] &= \boldsymbol{V}_{t+1} = Var(\,\boldsymbol{\nu}_{t+1}\,) \\ &= \boldsymbol{Z}_{t+1}\boldsymbol{\Sigma}_{t+1|t}\boldsymbol{Z}_{t+1}^{'} + \boldsymbol{H}_{t+1}. \end{aligned} \tag{2.38}$$

We repeat the derivation in equation 2.36 for the i-step-ahead prediction, so obtaining:

$$\begin{aligned} \hat{\boldsymbol{y}}_{t+i} &= E[\,\boldsymbol{y}_{t+i} \mid \mathfrak{F}_t] \\ &= \boldsymbol{c}_{t+i} + \boldsymbol{Z}_{t+i}\,\boldsymbol{\alpha}_{t+i|t}, \end{aligned} \tag{2.39}$$

The prediction error for this i-step-ahead prediction is:

$$\begin{aligned} \hat{\boldsymbol{e}}_{t+i} &= \boldsymbol{y}_{t+i} - \hat{\boldsymbol{y}}_{t+i} \\ &= \boldsymbol{Z}_{t+i}\,(\,\boldsymbol{\alpha}_{t+i} - \boldsymbol{\alpha}_{t+i|t}) + \boldsymbol{\varepsilon}_{t+i}. \end{aligned} \tag{2.40}$$

Using equation 2.40, the covariance matrix of the i-step-ahead prediction error is :

$$\begin{aligned} Var[\,\hat{\boldsymbol{e}}_{t+i}\,] &= \boldsymbol{V}_{t+i} = Var(\,\boldsymbol{\nu}_{t+i}\,) \\ &= \boldsymbol{Z}_{t+i}\boldsymbol{\Sigma}_{t+i|t}\boldsymbol{Z}_{t+i}^{'} + \boldsymbol{H}_{t+i}. \end{aligned} \tag{2.41}$$

From equation 2.12, we also obtain:

$$\boldsymbol{\alpha}_{t+i+1|t} = \boldsymbol{d}_{t+i} + \boldsymbol{T}_{t+i}\,\boldsymbol{\alpha}_{t+i|t}. \tag{2.42}$$

Using equation 2.42, we obtain:

$$\boldsymbol{\alpha}_{t+i+1} - \boldsymbol{\alpha}_{t+i+1|t} = \boldsymbol{T}_{t+i}\,(\,\boldsymbol{\alpha}_{t+i} - \boldsymbol{\alpha}_{t+i|t}) + \boldsymbol{R}_{t+i}\,\boldsymbol{\eta}_{t+i}. \tag{2.43}$$

Finally, taking the covariance from equation 2.43, given $\mathfrak{F}_t$, we obtain:

$$\boldsymbol{\Sigma}_{t+i+1|t} = \boldsymbol{T}_{t+i}\,\boldsymbol{\Sigma}_{t+i|t}\,\boldsymbol{T}_{t+i}^{'} + \boldsymbol{R}_{t+i}\,\boldsymbol{Q}_{t+i}\boldsymbol{R}_{t+i}^{'}. \tag{2.44}$$

Note that equations 2.39 and 2.44 are the iterations of the Kalman filter in the step, $t+i$, for $i = 1\,,\,2\,,\,\ldots\,,\,h$ with the assumption that $\boldsymbol{\nu}_{t+i} = \boldsymbol{0}$ and $\boldsymbol{K}_{t+i} = \boldsymbol{0}$.

2.4 Parameter Estimation:

We attempt to estimate the parameters of the model in equations 2.1 and 2.2 using the Kalman filter to evaluate the log-likelihood function. The model matrices $\boldsymbol{T}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Z}_t$,$\boldsymbol{Q}_t$, and $\boldsymbol{H}_t$ are the functions of some parameters $\boldsymbol{\theta}$ in which we want to estimate using maximum likelihood method. As we mentioned prior in this chapter all involved distributions are either normal or multivariate normal distributions. In equations 2.15 , 2.16 and 2.17, we show that $\boldsymbol{\nu}_t$ is independent of $\mathfrak{F}_{t-1}$ and $\boldsymbol{\nu}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_t) = (2\pi)^{\frac{-k}{2}} \mid \boldsymbol{V}_t \mid^{\frac{-1}{2}} e^{\boldsymbol{\nu}_t^{'}\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t}$. Therefore, the likelihood function is:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}_1\,,\,\boldsymbol{y}_2\,,\,\cdots\,,\,\boldsymbol{y}_n) &= f(\boldsymbol{y}_1\,,\,\boldsymbol{y}_2\,,\,\cdots\,,\,\boldsymbol{y}_n \mid \boldsymbol{\theta}) \\ &= \prod_{t=1}^{n} f(\boldsymbol{y}_t \mid \mathfrak{F}_{t-1}, \boldsymbol{\theta}) \\ &= \prod_{t=1}^{n} f(\boldsymbol{\nu}_t \mid \mathfrak{F}_{t-1}, \boldsymbol{\theta}) \\ &= \prod_{t=1}^{n} \left((2\pi)^{\frac{-k}{2}} \mid \boldsymbol{V}_t \mid^{\frac{-1}{2}} e^{\boldsymbol{\nu}_t^{'}\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t} \right). \end{aligned} \tag{2.45}$$

Taking the logarithm of the likelihood function in equation 2.45, we derive the log-likelihood function as follows:

$$\ln \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}_1 \, , \, \boldsymbol{y}_2 \, , \, \cdots \, , \, \boldsymbol{y}_n) = -\frac{kn}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \mid \boldsymbol{V}_t \mid -\frac{1}{2}\sum_{t=1}^{n} \boldsymbol{\nu}_t^{'}\boldsymbol{V}_t^{-1}\boldsymbol{\nu}_t. \tag{2.46}$$

As a result, the Kalman filter algorithm can be applied to recursively generate the log-likelihood function. We illustrate the Kalman filter algorithm and estimate the parameters of the simulated $ARMA(2,1)$ given in example 2.4.1.

Example 2.4.1. Consider the $ARMA(2,1)$ time series in equation 2.47. We simulate a sample size of 1000 for the $ARMA(2,1)$ model using 5610 as a seed for the distribution innovation ε_t. We present the $ARMA(2,1)$ model in the state space form applying Harvey's approach as explained in example 2.2.1. Finally, we fit the simulated data to the $ARMA(2,1)$ model using the Kalman filter algorithm (by deploying FKF package in R programming written by Luethi *et al.*) and reported the results as follows:

$$y_t \;\; = \;\; 0.6\, y_{t-1} \; - \; 0.2\, y_{t-2} \; + \; \varepsilon_t \; - \; 0.2\, \varepsilon_t, \quad \textit{where} \;\; \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2 = 0.8), \tag{2.47}$$

Figure 2.1 shows the time plot of simulated data: Plot (a) shows the simulated series and its corresponding filtered series derived by Kalman filter algorithm . Plot (b) checks for linear serial dependence in the residual series through autocorrelation function (ACF). The state space representation is given in equations 2.9 and 2.10 where the $\boldsymbol{T}$, $\boldsymbol{R}$ and $\boldsymbol{Z}$ are time invariant and are:

$$\boldsymbol{Z} \; = \; (\, 1 \, , \, 0 \,) \, , \quad \boldsymbol{T} \; = \; \begin{bmatrix} 0.6 & 0 \\ -0.2 & 0 \end{bmatrix}, \text{ and } \quad \boldsymbol{R} \; = \; \begin{bmatrix} 1 \\ -0.2 \end{bmatrix}. \tag{2.48}$$

The model parameters are estimated and given in the table 2.1.

Table 2.1: Estimated parameters for 1000 sample size data generated based on the $ARMA(2,1)$ model in equation 2.47 using the Kalman filter algorithm.

	ϕ_1	ϕ_2	θ_1	σ_ε
value	0.6763	-0.2548	0.2676	0.8665
Sd.Err	0.161	0.0597	0.1647	0.0194

2.4.1 New Local Linearization Method

Estimating parameters for a SDE from discrete observations is usually a practical, crucial, complicated task, and numerically unstable in many cases. Since a wide range of

practical and flexible SDEs are nonlinear ones for which the transition densities are not known, we are limited to using numerical or approximation methods. In some of the dynamics, studied in this thesis, such as those given by equations 3.62, 4.59 and 5.27, we are dealing with SDEs with nonlinear and unknown transition densities. We are, therefore, limited to numerical or approximation methods. Some of these methods can be categorized by simulations based methods (Gallant and Tauchen (1996)), nonparametric methods (Stanton (1997)), approximating the transition densities (Aït-Sahalia (2002)), and local linearization method (Ozaki (1992)). In this research, we deploy the new local linearization method (NLLM) introduced by Shoji and Ozaki (1998) to estimate parameters of mentioned SDEs earlier in this section. The new local linearization method is a pseudo-maximum likelihood estimation method for a usually nonlinear SDE with discretely observed data. The main idea behind this method is to approximate the original nonlinear SDE by a linear diffusion process. This method is a generalized version of the local linearization method introduced by Ozaki (1992). This method can be applied to a one-dimensional diffusion process of the following form:

$$dY_t = f(Y_t, t, \boldsymbol{\theta})\, dt + g(Y_t, \boldsymbol{\theta})\, dW_t, \tag{2.49}$$

where
$f(Y_t, t, \boldsymbol{\theta})$ is a two times continuously differentiable function with respect to Y_t,
$f(Y_t, t, \boldsymbol{\theta})$ and $g(Y_t, \boldsymbol{\theta})$ are continuously differentiable functions with respect to t and Y_t respectively, where $\boldsymbol{\theta}$ is the vector of the model parameters.
By transforming using the *Lamperti transformation* $\left(x = \int \frac{1}{g(y,\boldsymbol{\theta})} dy\right)$ and applying Itô's lemma, Shoji and Ozaki (1998) show that this process, 2.49 can be transformed into a more tractable diffusion process of the form as follows:

$$dX_t = h(X_t, t, \boldsymbol{\theta})\, dt + \sigma\, dW_t. \tag{2.50}$$

Applying Itô's lemma to the drift function $h(X_t, t, \boldsymbol{\theta})$, we derive:

$$dh(X_t, t, \boldsymbol{\theta}) = \left(\frac{\partial h(X_t, t, \boldsymbol{\theta})}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 h(X_t, t, \boldsymbol{\theta})}{\partial x^2}\right) dt + \frac{\partial h(X_t, t, \boldsymbol{\theta})}{\partial x} dX_t. \tag{2.51}$$

Assuming $\frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial x}$, $\frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial t}$ and $\frac{\partial^2 h(X_s, s, \boldsymbol{\theta})}{\partial x^2}$ are constant, $h(X_t, t, \boldsymbol{\theta})$ can be linearized in terms of X_t and t if $t \in [s, s + \Delta s)$ and Δs to be quite small. Therefore, we have:

$$\begin{aligned} h(X_t, t, \boldsymbol{\theta}) \simeq \quad & h(X_s, s, \boldsymbol{\theta}) + \left(\frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 h(X_s, s, \boldsymbol{\theta})}{\partial x^2}\right)(t - s) \\ & + \frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial x}(X_t - X_s). \end{aligned} \tag{2.52}$$

Therefore, we have:

$$h(X_t, t, \boldsymbol{\theta}) \simeq L_s X_t + M_s t + N_s, \tag{2.53}$$

where L_s, M_s, and N_s are constant and equal:

$$\begin{aligned} L_s &= \frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial x}, \\ M_s &= \frac{\sigma^2}{2} \frac{\partial^2 h(X_s, s, \boldsymbol{\theta})}{\partial x^2} + \frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial t}, \\ N_s &= h(X_s, s, \boldsymbol{\theta}) - X_s \frac{\partial h(X_s, s, \boldsymbol{\theta})}{\partial x} - s\, M_s. \end{aligned}$$

As a result, we can consider the following bilinear SDE instead of the original nonlinear SDE in 2.50 when $t \in [s, s + \Delta s)$:

$$dX_t = (L_s X_t + M_s t + N_s)\, dt + \sigma\, dW_t. \tag{2.54}$$

Let $Z_t = e^{-L_s t} X_t$, and applying Itô' lemma, it is easy to show that this SDE, 2.54 has an explicit solution as follows:

$$Z_t = Z_s + \int_s^t (uM_s + N_s) e^{-L_s u}\, du + \sigma \int_s^t e^{-L_s u} \mathrm{d}W_u. \tag{2.55}$$

Consequently, after discretization and simplification, we obtain:

$$X_t = X_s + \frac{h(X_s, s, \boldsymbol{\theta})(e^{L_s(t-s)} - 1)}{L_s} + \frac{M_s(e^{L_s(t-s)} - 1 - L_s(t-s))}{L_s^2} + \sigma \int_s^t e^{L_s(t-u)} \mathrm{d}W_u. \tag{2.56}$$

Equation 2.56 shows that the conditional distribution of X_t given X_s is normally distributed with mean and variance :

$$\begin{aligned} E_s = E[X_t | X_s] = X_s &+ \frac{h(X_s, s, \boldsymbol{\theta})(e^{L_s(t-s)} - 1)}{L_s} \\ &+ \frac{M_s(e^{L_s(t-s)} - 1 - L_s(t-s))}{L_s^2}, \end{aligned} \tag{2.57}$$

$$V_s = Var[X_t | X_s] = \frac{\sigma^2}{2L_s} (\, e^{2L_s\,(t-s)} - 1\,). \tag{2.58}$$

We note that Shoji and Ozaki (1998) show that since the new local linearization method considers the stochastic behavior of X_t when $h(X_t, t, \boldsymbol{\theta})$ is discretized locally, its accuracy is higher than two other methods namely the local linearization method (Ozaki (1992)) and the Euler method particularly when the time step size, Δs is relatively big. This method is also fairly simple to implement. Shoji and Ozaki (1998) also show that if

time t approaches s, the absolute mean and square values of the one-step forecast errors converge to zero with $O((t-s)^2)$ and $O((t-s)^3)$ respectively.

Since the distribution of $X_t \mid X_s$ is approximately normal, for observations series $\{x_1, x_2, \cdots, x_n\}$, the log-likelihood function can be obtained as follows (See Shoji and Ozaki (1998) for more detail):

$$\begin{aligned} \ln \mathcal{L}(\boldsymbol{\theta} \mid x_1 \ , \ x_2 \ , \ \cdots \ , \ x_n) &= -\frac{1}{2}\sum_{i=1}^{n-1}\left\{\frac{(x_{i+1}-E_i)^2}{V_i} + \ln(2\Pi V_i)\right\} \\ &+ \ \ln(p(x_1)), \end{aligned} \tag{2.59}$$

where $p(x_1)$ is the density function of X_1.

2.5 Conclusion

In time series, the Kalman filter is a well-known and powerful algorithm to estimate parameters by using maximum-likelihood function. In this chapter, we reviewed the concepts, implications, embedded conditions, the state space representations, and the algorithm itself. We argue that the Kalman filter algorithm is mostly deployed in the models that the state variables are unobservable. We also describe the new local linearization method that can be used to calibrate parameters of a nonlinear and one-dimensional SDE when its transition density is unique but unknown. In remaining chapters, whenever we want to estimate a dynamics's parameters, we deploy one of these two outstanding methods based on characteristics of the SDE, as relevant.

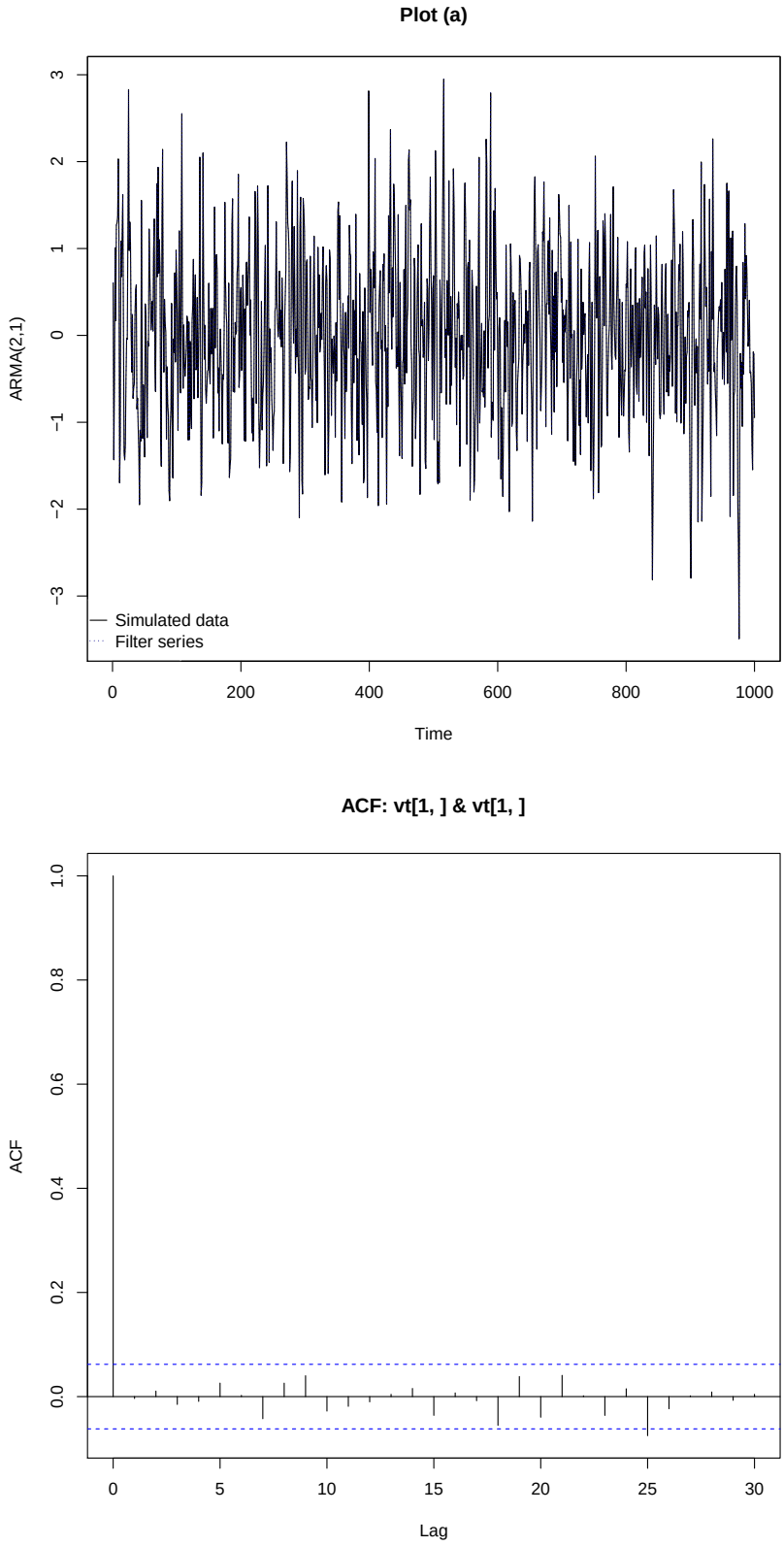

Figure 2.1: Time plot of the simulated data and fitting model to the $ARMA(2,1)$ model in equation 2.47: Plot (a) shows the 1000 simulated series and its corresponding filtered series. Plot (b) checks for linear serial correlation in the residual series through 'acf'. Here, we use 5610 as a seed to generate this series.

Chapter 3

NEW AND EXISTING COMMODITY PRICE MODELS

3.1 Introduction

Commodity markets are increasingly important in modern economies. Nowadays, market participants are widely applying stochastic commodity price models. However, these models have many features which make them difficult to analyze. Several factors make commodity models more complicated than other assets models. As we discussed in Chapter 1, commodity markets require us to deal with issues including many influential fundamental prices drivers, seasonality, frequent price jumps, decentralization, low liquidity, and stochastic convenience yields and interest rates. Moreover, for commodity based assets, the main state variables such as spot price and convenience yield are unobservable (as only forward/futures are traded by speculators and price discovery happens in these markets), escalating the complexity level. This is why we need the Kalman filter algorithm. Commodities can be categorized into various classes namely agricultural, energy, precious metals, and base metals. Although each commodity has its own particular properties, and we should consider these characteristics in our modeling, we attempt to establish a general structure which can be applicable in the modeling of derivatives with any kind of underlying commodities. In this Chapter we will review several popular models. We also propose a new one-factor mean-reverting process to model a commodity price.

We review forward and futures contracts and their differences in Section 3.2. In Section 3.3, convenience yield, one of the important state variables in commodity markets, will be introduced. In Section 3.4, we explain a simple model to price a commodity in which the commodity spot price is considered to follow geometric Brownian motion. We review the one-factor mean-reverting commodity pricing model proposed by Schwartz

(1997) in Section 3.5. In Section 3.6, we analyze the most popular two-factor model introduced by Gibson and Schwartz (1990) to price commodity contingent claims particulary for crude oil. Finally, in Section 3.7, a new generalized one-factor commodity pricing model is introduced. Properties of this new dynamical model will be investigated and it will be shown that this one-factor process is able to capture the key characteristics of the dynamics of commodity spot-prices including mean-reversion, heavy tails, skewness and kurtosis. The new stochastic process is a nonlinear process with a unique but unknown transition distribution; therefore, the new local linearization method, introduced by Shoji and Ozaki (1998), will be deployed to estimate the model's parameters. The new generalized one-factor model is compared with Schwartz's one-factor dynamics by fitting these models to the WTI crude oil's front futures contracts. It will be argued which process is capable to explain the reality of the commodity spot-price process more accurately using both observed and estimated results.

3.2 Futures Contracts versus Forward contracts

As we discussed in Section 1.3, a forward contract is a non-standardized contract between two parties to exchange an asset or a commodity at specified price at a fixed date in the future, known as *the maturity or delivery date*. A futures contract is a special type of forward contract with some crucial features. In futures contracts, the underlying commodity or asset's quantity, quality and delivery location are standardized which are not generally the case for forward contracts. Unlike forward contracts, futures contracts are traded in centralized exchanges. Futures contracts are *marked-to-market* contracts and their prices are updated periodically (usually daily), which means that day by day, price changes are applied to counterparties' accounts through *clearing houses* of exchanges until maturity. At maturity, the future price is simply its spot price, whereas in forward contracts, the agreement only settles once at maturity. Therefore, futures contracts' counterparties are not exposed to *credit risk*. Although both contracts have a zero initial values, the futures and forward contracts' cash flows are different and in order to maintain margin, a futures contract makes parties to deposit interim payments in their margin accounts during its life. Two important reasons why the futures and forward prices with identical maturity and underlying asset are not the same: include stochastic interest rates and credit risk.

3.2.1 Forward Prices

For simplicity, consider $f(t,T)$ to be the forward price at time t on non-dividend stock maturing at date T. Define $B(t,T)$ to be the price of a risk-free zero-coupon bond at time t with maturity T and face value \$1. We also assume that there is no credit risk in the forward market and here, we ignore transaction costs and taxes. To find the forward

price, we deploy a no-arbitrage argument, which is based on *the law of one price* (if two investment strategies have identical payoffs, their current value must be the same as well). We construct a strategy at time t as follows:

- Short one unit of stock and receive $S(t)$,
- Invest dollar amount of $S(t)$ in zero-coupon bond $B(t,T)$,
- Long a forward contract $f(t,T)$ with $S(t)$ as underlying.

As we can see, the amount of investment in this strategy is zero at time t. Since there is no interim payment, the payoff of the strategy at maturity T will be:

- Use stock from long forward to close short,
- Your investment will grow to $\frac{S(t)}{B(t,T)}$,
- Pay the amount of $f(t,T)$ in exchange for receiving one unit of stock.

Based on a no-arbitrage argument, the cost of the strategy must be zero at time T as well; therefore, we have:

$$f(t,T) = \frac{S(t)}{B(t,T)}. \tag{3.1}$$

As we can see if the interest rate is the constant r and same for all maturities, the forward price will be $f(t,T) = S(t)\,e^{r(T-t)}$. We should note that the forward price $f(t,T)$ at time t is the value of a contract that pays $\frac{S(T)}{B(t,T)}$ at maturity T (Cox *et al.* (1981)) .

3.2.2 Futures Prices

Consider the futures price $F(t,T)$ at time t maturing at date T on a non-dividend paying stock $S(t)$. We consider that futures contracts are rolled-over daily, which means profit or loss is credited or charged to the counterparties' margin accounts daily. We also consider that time interval $[t,T]$ is portioned daily,which means $t_1 = t$, $t_2 = t+1$, $\ldots$, $t_n = T$. Let $B(t_i, t_{i+1})$ be the one-period, $[t_i, t_{i+1}]$, zero-coupon bond price. We calculate how much a futures contract $F(t,T)$ at maturity T will pay using the following strategy:

- Invest $F(t,T)$ amount at time t repeatedly in one-period zero-coupon bonds until maturity time T,
- At each time t_i $i = 1, 2, \ldots, n-1$, buy (long) $\frac{1}{\prod_{j=1}^{i} B(t_j, t_{j+1})}$ number of the futures contract $F(t_i, T)$ and liquidate (close) the position at time t_{i+1},

- At time t_{i+1}, $i = 1, 2, \ldots, n-1$, we will have $\frac{F(t_{i+1},T)-F(t_i,T)}{\prod_{j=1}^{i} B(t_j,t_{j+1})}$ profit or loss from closing long position in the future contract that we bought prior day. Each time, we invest this payoff in one-period zero-coupon bond continually until maturity T. Final value of this investment will be:

$$\frac{F(t_{i+1},T) - F(t_i,T)}{\prod_{j=1}^{i} B(t_j,t_{j+1})} \frac{1}{\prod_{j=i+1}^{n-1} B(t_j,t_{j+1})} = \frac{F(t_{i+1},T) - F(t_i,T)}{\prod_{j=1}^{n-1} B(t_j,t_{j+1})}.$$

At maturity T, the strategy will have the following payoff:

$$\frac{F(t,T)}{\prod_{j=1}^{n-1} B(t_j,t_{j+1})} + \sum_{i=1}^{n-1} \frac{F(t_{i+1},T) - F(t_i,T)}{\prod_{j=1}^{n-1} B(t_j,t_{j+1})} = \frac{F(T,T)}{\prod_{j=1}^{n-1} B(t_j,t_{j+1})} = \frac{S(T)}{\prod_{j=1}^{n-1} B(t_j,t_{j+1})}. \tag{3.2}$$

Therefore, a long position in a futures contract $F(t,T)$ at time t, will generate the payoff equal to that given by equation 3.2.
Cox *et al.* (1981) shows that under the risk neutral measure Q, the futures price at time t with maturity T must be equal to $F(t\,,\,T\,,\,S_t) = E_{t,Q}[\,S_T\,]$. We will apply this result to derive futures prices throughout this chapter and the remaining chapters.

3.2.3 Relationship between Futures Prices and Forward Prices:

One of the crucial relationship between futures and forward prices when their underlying and maturity time T are identical is that if risk-free interest rate is constant and equal for all maturities, futures and forward prices are equal. We can easily see this relationship as follows:
When interest rate is constant, all one-period zero-coupon bonds, $B(t_i, t_{i+1})$ for $i = 1, 2, \ldots, n-1$ are equal and we have:

$$B(t,T) = \prod_{i=1}^{n-1} B(t_i,t_{i+1}). \tag{3.3}$$

Results in equations 3.1, 3.2 and 3.3 prove this claim. In general, Cvitanic and Zapatero (2004) show that futures and forward prices for the same maturity and underlying has following relationship:

$$f(t,T,S_t) - F(t,T,S_t) = \frac{\mathrm{Cov}_{t,Q}[S(T), B(0,T)]}{E_{t,Q}[B(0,T)]}. \tag{3.4}$$

Clearly, this result shows that if and only if the interest rate and the underlying asset or commodity are uncorrelated under Q measure, the futures and forward prices are equal. This is obviously the case when the interest rate is constant.

3.2.4 Forward Price Curve

Forward (futures) prices, bringing liquidity and considered as way of *price discovery*, are the most crucial part of any risk management and derivatives pricing. The forward price curves give the market perception of the futures prices of their underlying assets or commodities to the markets' participants. The forward curve for a particular underlying commodity or asset can be defined as a collection of existing forward prices as a function of their maturity time. In other words, a mark-to-market forward curve demonstrates the future prices of multiple contracts with the same underlying asset but different maturities. Construction of forward price curves is a complicated process, because the resulting curves must not only match the existing market prices, but also, when there are limited available prices, the forward curve model should be the best representation of market behavior. Pilipovic (2007) claims that in energy commodity markets, since the correlations of interest rate and energy futures prices are close to zero and the credit risk associated to forward prices can be resolved through collateral such as bonds, the futures price and its corresponding forward price are approximately equal. As a result, market participants may use futures price and forward price notions interchangeably. It is worth mentioning that, in the case of crude oil, if the oil price increases and stays high for a relatively long time, it will impact inflation and the short interest rate will rise as inflation consequence so some type of structured correlation is indicated. Generally, the construction of an appropriate forward curve is a sophisticated task and involves making assumptions, interpreting market data, and comprehension of the market price behavior. We should also associate the impact of important factors such as convenience yield, seasonality, and arrival of highly influential news in the forward curves. Forward curves for nonseasonal commodities such as crude oil usually have different shapes namely *contango, backwardation, or more complex humped shapes*. If a forward curve is a increasing (decreasing) function of expiry dates it is called contango (backwardation) respectively. Figure 3.1 shows four different shapes of forward curves for WTI crude oil in different times. As we can see at time 5/22/2008, the forward curve shows backwardation for the near-portion followed by contango for the longer-term portion. Here, since WTI crude oil is not seasonal, we build its forward curve using the *gam (generalized additive models)* function in R *mgcv* package utilizing the *smoothing B-spline* option in the function. We can also construct forward curves using a specified model for forward prices. For instance, in the rest of this chapter, we explore three different models to price commodities in which we derive their futures prices formulas. After calibrating these models, we can easily build their theoretical forward curves.

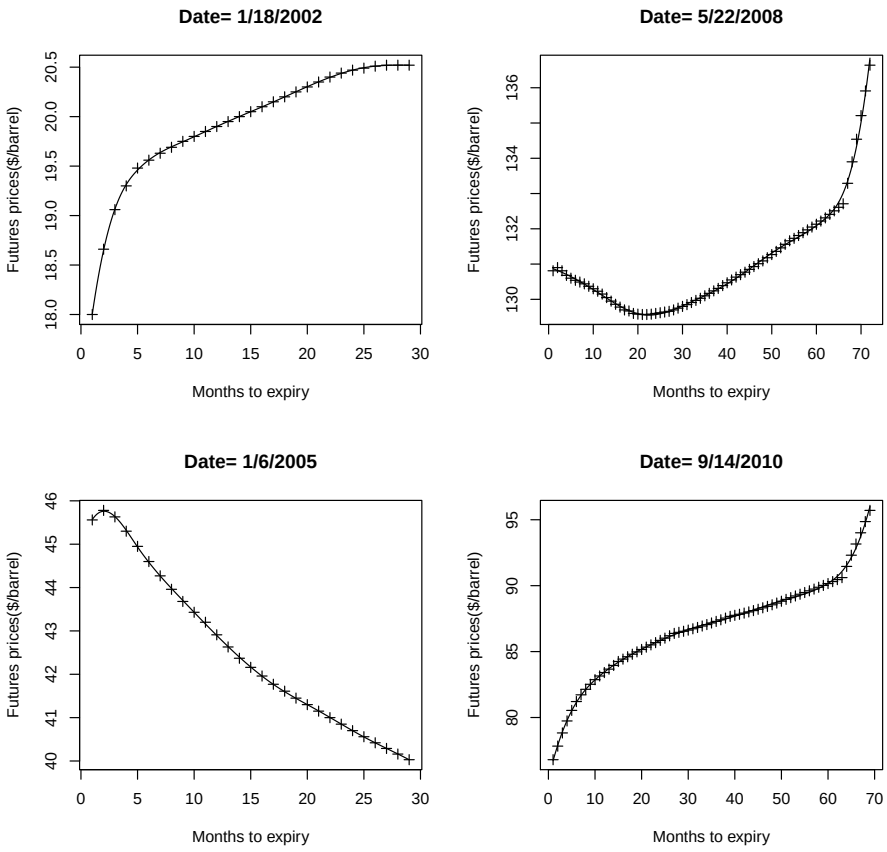

Figure 3.1: The observed futures prices for WTI crude oil in given dates and their corresponding fitted forward curves using smoothing B-spline.

3.3 Convenience Yield

Convenience yield is a fundamental state variable for storable commodities such as crude oil and its derivatives. The convenience yield is the net benefit associated with physically ownership of a commodity, rather than its futures contract, less storage costs but excluding financing costs and it can be positive or negative. In other words, when the commodity (energy) is scarce and demand is high, industrial consumers are willing to pay a premium (convenience yield) to get the minimum required commodity (energy) to run their plants delivered today to avoid losses caused by stopping their production. Since there is a strong negative correlation between convenience yield and the level of inventories, the correlation between spot price and convenience yield is positive. In most commodity pricing models, the convenience yield is considered as formally equivalent to dividends in models of dividends-paying stocks. However, the convenience yield is

not directly observable and, must be estimated from market data based on the specified forward (futures) pricing models. In the rest of this chapter, we estimate convenience yield using the specified forward price model. Here, we estimate convenience yield using empirical observed futures prices $F(t,T_i),\ i=1,2,\ldots,n$, at time t. For a storable commodity, the forward price $F(t,s)$ on S_t $(t<s)$ can be defined as follows:

$$F(t,s) \;=\; S_t\, e^{\int_t^s (\, r_v - \delta_v \,)\, dv}. \tag{3.5}$$

we solve equation 3.5 to derive $\delta_{t,s}$ as follows:

$$\delta_{t,s} \;= r_{t,s} - \frac{\partial \ln(F(t,s))}{\partial s}. \tag{3.6}$$

where $r_{t,s}$ is the continuously compounded short interest rate applied to the interval of $[t,s]$. To calculate implied convenience yield using equation 3.6, we need the forward curve. We fitted the forward curve with smoothing B-splines using the observed futures prices, $F(t,T_i),\ i=1,2,\ldots,n$, in section 3.2.4. Figure 3.2 depicts the implied convenience yields derived using the futures contracts of 2 and 3 months to maturities for the daily futures prices of WTI crude oil from January 2, 2001 to September 6, 2011. We deploy the constructed forward curve (from section 3.2.4), the estimated yield curve, and equation 3.6 to derive the implied convenience yield curve. Figure 3.3 demonstrates the implied convenience yields using the derivatives of fitted forward curves which are plotted in figure 3.1 for WTI crude oil in given dates.

It is worth mentioning that if we assume convenience yield constant in the commodity contingent claims, the forward curves will be monotonic either contango or backwardation and the model will fail to fully explain the reality of the forward curves which includes more complex humped shapes. As we can see here, to estimate the implied convenience yield, we need at least two futures prices (ideally two consecutive prices). As we can notice from figures 3.2, unfortunately, the implied convenience yield derived by using 2 and 3 months to maturities futures prices does not agree with the one derived by using 11 and 12 months maturities futures prices. One solution for this issue is introduced by Schwartz (1997). He estimates convenience yield using five different futures prices for shorter maturity contracts and longer maturity contracts separately.

3.4 Preliminary model:

In this simplest class of models, we have only one factor (state variable), commodity price, S_t which is assumed to follow a stochastic process. We consider that interest rate, r and convenience yield, δ are constant. The spot price assumed to follow geometric Brownian motion as follows:

$$dS_t = (\,\mu - \delta\,)S_t dt + \sigma S_t dW_t, \tag{3.7}$$

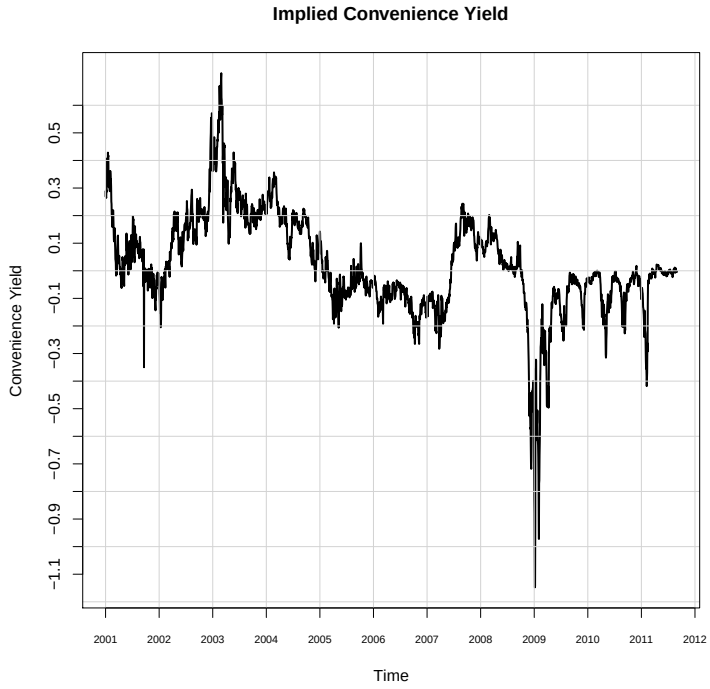

Figure 3.2: Implied convenience yields derived using the futures contracts of 2 and 3 months to maturities for the daily futures prices of WTI crude oil staring from January 2, 2001 to September 6, 2011. For simplicity, we considered the short interest rates r_t constant and equal to 0.05.

where

S_t is the spot price at time t,
μ is the drift of the process and constant,
σ is the volatility of the process and constant,
δ is the convenience yield,
dW_t is the increment of a standard Brownian motion.

Suppose that F_t is the value of forward contract (futures contract) on S_t at time t with expiry time T. By using Ito's lemma, we have:

$$dF_t \;=\; \left[\frac{\partial F_t}{\partial S_t} (\,\mu - \delta\,)\, S_t \;+\; \frac{\partial F_t}{\partial t} \;+ \frac{\partial^2 F_t}{2\partial {S_t}^2}\; \sigma^2\, S_t^2 \right] dt \;+\; \frac{\partial F_t}{\partial S_t}\, \sigma\, S_t\, dW_t. \qquad (3.8)$$

We construct a portfolio to eliminate the risk contribution of the Brownian motion as follows:

- sell (short) one F_t contract ($-F_t$)

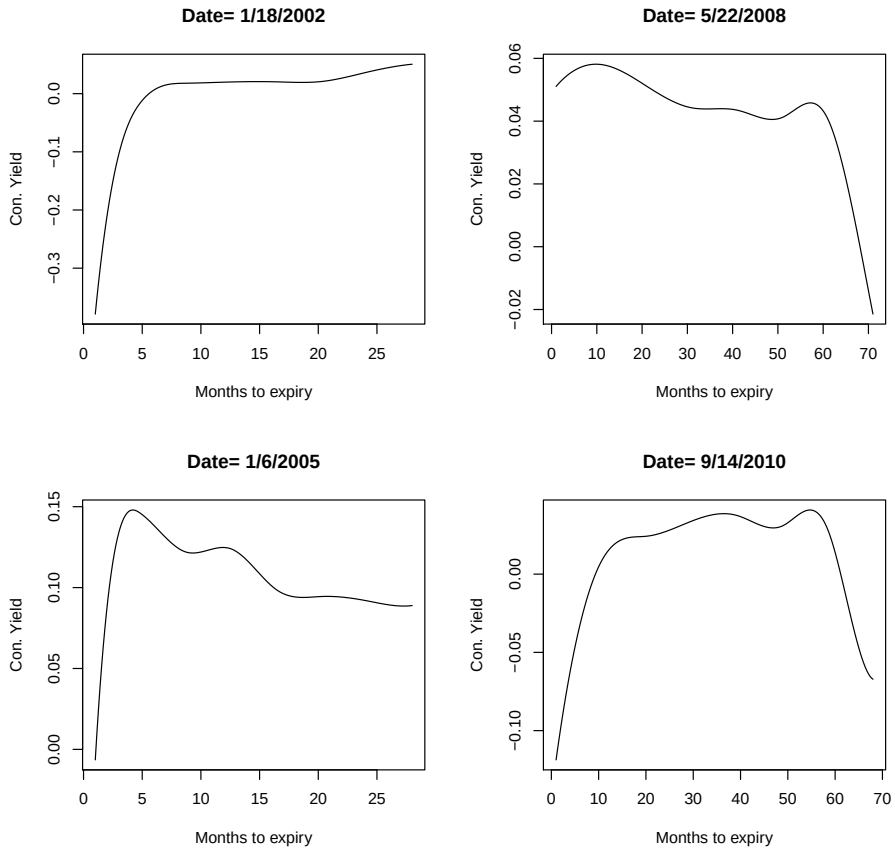

Figure 3.3: The interpolated convenience yields using smoothing B-spline for WTI crude oil in given dates, which their corresponding fitted forward curves are plotted in figure 3.1. For simplicity, in these graphs, we considered the short interest rates $r_{t,s}$ to be constant and equal to 5% for all dates and maturities.

- buy (long) $\frac{\partial F_t}{\partial S_t}$ amount of the underlying commodity, $S_t \left(+ \frac{\partial F_t}{\partial S_t} S_t \right)$

Define Π_t to be the value of the portfolio; as a result, we have

$$\Pi_t = -F_t + \frac{\partial F_t}{\partial S_t} S_t, \tag{3.9}$$

The change in the value of the portfolio, $d\Pi_t$ in the time interval dt is given by:

$$d\Pi_t = -dF_t + \frac{\partial F_t}{\partial S_t} dS_t, \tag{3.10}$$

We substitute equations 3.7 and 3.8 into 3.10 to obtain:

$$\begin{aligned} d\Pi_t &= -\left[\frac{\partial F_t}{\partial S_t}(\mu-\delta)S_t + \frac{\partial F_t}{\partial t} + \frac{\partial^2 F_t}{2\partial {S_t}^2}\sigma^2 S_t^2\right]dt - \frac{\partial F_t}{\partial S_t}\sigma S_t dW_t \\ &\quad + \frac{\partial F_t}{\partial S_t}[(\mu-\delta)S_t dt + \sigma S_t dW_t] \\ &= \left[-\frac{\partial F_t}{\partial t} - \frac{\partial^2 F_t}{2\partial {S_t}^2}\sigma^2 S_t^2\right]dt. \end{aligned} \tag{3.11}$$

Since equation 3.11 does not depend on dW_t, the portfolio must be riskless during dt. Because part of the return comes from holding $\frac{\partial F_t}{\partial S_t}$ amount of the commodity, S_t and this ownership of the commodity earns at the rate of the convenience yield which is $\delta \frac{\partial F_t}{\partial S_t} S_t\, dt$, we have:

$$d\Pi_t = r\Pi\, dt_t - \delta \frac{\partial F_t}{\partial S_t} S_t\, dt, \tag{3.12}$$

Notice that although convenience yield is not cash but rather insurance like. Substituting equations 3.9 and 3.11 into 3.12, we derive:

$$\left[\frac{\partial F_t}{\partial t} + \frac{\partial^2 F_t}{2\partial {S_t}^2}\sigma^2 S_t^2\right]dt = r\left[F_t - \frac{\partial F_t}{\partial S_t} S_t\right]dt + \delta \frac{\partial F_t}{\partial S_t} S_t\, dt, \tag{3.13}$$

As a result, the partial differential equation satisfied by the forward contract value is:

$$\frac{\partial F_t}{\partial t} + (r-\delta) S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial^2 F_t}{2\partial {S_t}^2}\sigma^2 S_t^2 - r F_t = 0. \tag{3.14}$$

With terminal condition $F_T(T,T) = S_T$.
From equation 3.14, we can see that S_t under a risk neutral measure Q has the following dynamics:

$$dS_t = (r-\delta)S_t dt + \sigma S_t\, dW_t^Q, \tag{3.15}$$

where dW_t^Q is the increment of a standard Brownian motion under the risk neutral measure Q. Let $X_t = \ln S_t$. By applying Ito's lemma, we have:

$$\begin{aligned} dX_t &= (r - \delta - \frac{\sigma^2}{2})\, dt + \sigma\, dW_t^Q, \\ &\text{or} \\ X_T &= X_t + (r - \delta - \frac{\sigma^2}{2})(T-t) + \sigma W_{T-t}^Q. \end{aligned} \tag{3.16}$$

At time t, X_t is observed. Since X_t is an affine function of the normal variable W^Q_{T-t} and this is independent of $\mathfrak{F}_t$, the conditional distribution of X_T given $\mathfrak{F}_t$ (all information available at time t) under the risk-neutral measure Q is normal with following mean and variance:

$$\begin{aligned} E_Q[\,X_T\,|\,\mathfrak{F}_t\,] &= X_t + (\,r - \delta - \frac{\sigma^2}{2}\,)\,(T-t\,), \\ Var_Q[\,X_T\,|\,\mathfrak{F}_t\,] &= \sigma^2\,(T-t). \end{aligned} \tag{3.17}$$

Since $X_t = \ln S_t$, the conditional distribution of S_T given $\mathfrak{F}_t$ under the risk-neutral measure Q is log-normal with the same parameters. So the futures price (the forward price) equals to:

$$\begin{aligned} F_t &= E_Q[\,S_T\,|\,\mathfrak{F}_t\,] = e^{\{\,E_Q[\,X_T\,|\,\mathfrak{F}_t\,] + \frac{1}{2}\,Var_Q[\,X_T\,|\,\mathfrak{F}_t\,]\,\}}, \\ &= e^{\{\,X_t + (\,r-\delta-\frac{\sigma^2}{2}\,)\,(T-t) + \frac{1}{2}\,\sigma^2\,(T-t)\,\}}, \\ &= e^{\,X_t + (\,r-\delta\,)\,(T-t)}, \\ &= S_t\,e^{(\,r-\delta\,)\,(T-t)}. \end{aligned} \tag{3.18}$$

We can also derive the equation 3.18 using equation 3.15. Equation 3.15 shows that S_t is log-normally distributed and using the moment generating function for normal distribution, we obtain:

$$\begin{aligned} S_T &= S_t\,e^{\{\,(\,r-\delta-\frac{\sigma^2}{2}\,)\,(T-t) + \sigma\,W^Q_{T-t}\,\}}, \\ F_t &= E_Q[\,S_T\,|\,\mathfrak{F}_t\,] = S_t\,e^{(\,r-\delta\,)\,(T-t)}. \end{aligned}$$

3.5 One-factor model for the commodity spot price:

Schwartz (1997) introduced a mean-reverting one-factor stochastic process for a commodity spot price. In this model, he attempts to consider geometric Brownian motion for a commodity spot price, S_t by implementing mean-reversion which modified to the long-term value μ of the drift. He assumed that commodity spot price follows the dynamics:

$$dS_t = \kappa(\,\mu - \ln S_t\,)S_t dt + \sigma S_t dW_t, \tag{3.19}$$

where

S_t is the spot price at time t,
μ is the long-run mean,
σ is the volatility of the process,
κ is the rate of mean reversion,

dW_t is the increment of a standard Brownian motion.

In this model the only state variable is the spot price S_t, which is the key in future prices. Let X_t denote the logarithm of the spot price S_t at time $t\,(X_t = \ln(S_t))$ then we apply Ito's lemma on 3.19 to show that the log spot price follows an Ornstein-Uhlenbeck process: $\frac{\partial X_t}{\partial S_t} = \frac{1}{S_t}$, $\frac{\partial^2 X_t}{\partial {S_t}^2} = \frac{-1}{S_t^2}$

$$\begin{aligned} dX_t &= (\kappa(\mu - X_t)\frac{S_t}{S_t} - \frac{1}{2S_t^2}\sigma^2 S_t^2)\,dt + \sigma\frac{S_t}{S_t}\,dW_t, \\ &= (\kappa(\mu - X_t) - \frac{\sigma^2}{2})dt + \sigma\,dW_t, \\ &= \kappa(\mu - \frac{\sigma^2}{2\kappa} - X_t)dt + \sigma\,dW_t, \\ \text{Let: } \alpha &= \mu - \frac{\sigma^2}{2\kappa}, \quad \text{As a result, we have :} \\ dX_t &= \kappa(\alpha - X_t)dt + \sigma\,dW_t. \end{aligned} \tag{3.20}$$

where

$\alpha = \mu - \frac{\sigma^2}{2\kappa}$ is the long-run mean of log-price X_t.

In this model, interest rate r is considered constant and convenience yield δ is not included in the model. Neither of these assumptions are reasonable, particularly that of the convenience yield. Under a risk-neutral measure Q, the Ornstein-Uhlenbeck process 3.20, can be written as:

$$dX_t = \kappa(\alpha^* - X_t)\,dt + \sigma\,dW_t^*, \tag{3.21}$$

where

$\alpha^* = \alpha - \lambda$,

λ denotes the market price of risk (assumed constant),

dW_t^* denotes the increment of a standard Brownian motion under risk-neutral measure Q.

Now, let $Z_t = e^{\kappa t}X_t$ and apply Ito's lemma to the equation 3.21 to obtain:

$$dZ_t = \kappa\alpha^* e^{\kappa t}\,dt + \sigma e^{\kappa t}\,dW_t^*, \tag{3.22}$$

We integrate the prior equation 3.22 from t to T where $T > t$ to obtain:

$$Z_T = Z_t + \alpha^*(e^{\kappa T} - e^{\kappa t}) + \sigma\int_t^T e^{\kappa u}\mathrm{d}W_u^*. \tag{3.23}$$

Since $X_T = e^{-\kappa T} Z_T$, we replace in the prior equation 3.23 to obtain:

$$X_T = e^{-\kappa (T-t)} X_t + \alpha^* (1 - e^{-\kappa (T-t)}) + \sigma e^{-\kappa T} \int_t^T e^{\kappa u} \mathrm{d}W_u^*. \tag{3.24}$$

At time t , X_t is observed and $X_T|X_t$ is normally distributed; therefore, using the prior equation 3.24, the conditional distribution of X_T given $\mathfrak{F}_t$ under the risk-neutral measure Q is normal with the mean and variance as follows:

$$\begin{aligned} E_Q[X_T \mid \mathfrak{F}_t] &= e^{-\kappa (T-t)} X_t + \alpha^* (1 - e^{-\kappa (T-t)}), \\ Var_Q[X_T \mid \mathfrak{F}_t] &= \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa (T-t)}). \end{aligned} \tag{3.25}$$

Since $X_t = \ln (S_t)$, the conditional distribution of S_T given $\mathfrak{F}_t$ under the risk-neutral measure Q is log-normal with the same parameters. Therefore, the futures price (forward price) is:

$$\begin{aligned} F (t , T) &= E_Q[S_T \mid \mathfrak{F}_t] = e^{\{ E_Q[X_T \mid \mathfrak{F}_t] + \frac{1}{2} Var_Q[X_T \mid \mathfrak{F}_t] \}} \\ &= e^{\{ e^{-\kappa (T-t)} \ln S_t + \alpha^* (1 - e^{-\kappa (T-t)}) + \frac{\sigma^2}{4\kappa} (1 - e^{-2\kappa (T-t)}) \}}. \end{aligned} \tag{3.26}$$

Taking logarithms in equation 3.26, we obtain:

$$\ln F(t , T) = e^{-\kappa (T-t)} \ln S_t + \alpha^* (1 - e^{-\kappa (T-t)}) + \frac{\sigma^2}{4\kappa} (1 - e^{-2\kappa (T-t)}). \tag{3.27}$$

Applying Ito's lemma to equation 3.27, in the risk neutral measure Q confirms that $d \ln F(t , T , X_t)$ where $X_t = \ln S_t$ satisfies:

$$\begin{aligned} d \ln F(t , T , X_t) &= \frac{\partial \ln F}{\partial t} dt + \frac{\partial \ln F}{\partial X} dX_t + \frac{1}{2} \frac{\partial^2 \ln F}{\partial X^2} dX_t^2 \\ &= -\frac{\sigma^2}{2} e^{-2\kappa (T-t)} dt + \sigma e^{-\kappa (T-t)} \mathrm{d}W_t^*. \end{aligned} \tag{3.28}$$

The result in equation 3.28 shows that the variance of the log futures returns in this model is:

$$\sigma_{lnF}^2 = \sigma^2 e^{-2\kappa (T-t)}. \tag{3.29}$$

Equation 3.29 shows that as the maturity approaches to infinity, the volatility of the log returns of future price converges to zero.

3.5.1 The model drawbacks

One of the main drawback of this model is that the term structure of future prices is not flexible enough to capture all possible shapes in real world. To show this we take the limit of $F(t, T)$ when T approaches infinity ($T \to +\infty$) , we obtain

$$F(t, +\infty) = \exp(\alpha^* + \frac{\sigma^2}{4\kappa}). \tag{3.30}$$

As we can see in the equation 3.30, $F(t, +\infty)$ is independent of the spot price S_t. As a result, when the spot price is greater than this infinite maturity future price this model will be in backwardation and when the spot price is less than $F(t, +\infty)$, the model will be in contango. As a result, the term structure of future price in not flexible enough to match the reality of the commodities' future markets. Another issue in this model is that uncertainty on the commodities's derivatives can not be summarized in one factor: (spot price) and we should consider other drivers such as convenience yield and risk-free interest rate in the modeling (Pilipovic (2007)).

3.5.2 Parameter Estimation:

One of the implications in commodities pricing is that the state variables (in this model, the spot price) are not straightly observable. We can generally consider the future front contract as an approximation to the spot price in which is extensively traded in certain exchanges. The Kalman filter algorithm is generally deployed to estimate unobserved state variables. As a result, we construct the state space form (the transition and measurement equations) for this model as follows:

3.5.2.1 Transition Equation:

We obtain the transition equation by exact discretization of equation 3.24 for this model in the following form:

$$\boldsymbol{X}_{t+1} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{X}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, \quad t = 1, \ldots, NT, \tag{3.31}$$

where

$$\begin{aligned} \boldsymbol{d}_t &= \alpha(1 - e^{-\kappa \Delta t}), \quad \boldsymbol{T}_t = e^{-\kappa \Delta t}, \\ \boldsymbol{R}_t &= \sigma \sqrt{\frac{1 - e^{-2\kappa \Delta t}}{2\kappa}}, \end{aligned} \tag{3.32}$$

Here $\boldsymbol{\eta}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\eta}_t) = 0$ and $Var(\boldsymbol{\eta}_t) = 1$.

3.5.2.2 Measurement Equation:

The future prices in commodity market can be noisy. Consequently, the measurement equation can be obtained using equation 3.27 plus Gaussian white noise as follows:

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, NT, \tag{3.33}$$

where

$$\begin{aligned}
\boldsymbol{y}_t \quad & \text{is an } N \times 1 \text{ vector of observations, and } \boldsymbol{c}_t \,\&\, \boldsymbol{Z}_t \text{ are } N \times 1 \text{ vectors,} \\
\boldsymbol{y}_t &= \left[\ln F(t\,,\, T_i) \right], \quad i = 1, \ldots, N, \\
\boldsymbol{c}_t &= \left[\alpha^* \left(1 - e^{-\kappa (T_i - t)} \right) + \frac{\sigma^2}{4\kappa} \left(1 - e^{-2\kappa (T_i - t)} \right) \right], \quad i = 1, \ldots, N, \\
\boldsymbol{Z}_t &= \left[e^{-\kappa (T_i - t)} \right], \quad i = 1, \ldots, N,
\end{aligned} \tag{3.34}$$

Here $\boldsymbol{\varepsilon}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\varepsilon}_t) = 0$ and $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{H}$. We will use these equations in 3.5.3 below when we estimate parameters.

3.5.3 Empirical Results

One of the most important energy assets for the world's economy is crude oil. Recently, there has been wide interest in trading of financial derivatives such as futures and options on futures with oil as their underlying asset. We consider the daily observations of WTI crude oil futures prices from January 2, 2001 to September 6, 2011. The dataset we used is from the Chicago Mercantile Exchange (CME). We use this daily empirical data to estimate the model's parameters. We used futures contracts of 1,2,3,4 and 5 month to maturity. Table 3.1 summarizes the calibration results for one-factor commodity pricing model. Figure 3.4, plot (a) depicts the comparison between the filtered spot price and the front contract futures prices realizations and plot (b) demonstrates the prediction errors corresponding to the first and the second futures contracts. Here, in order to decrease the number of parameters to estimate, we assume the variance of the futures contracts' disturbances (measurement errors) with all maturities to be identical which means $\boldsymbol{H} = \zeta^2 \boldsymbol{I}_N$. Notice that when we estimate parameters in the one and two factor commodity pricing models, we may sometimes confront multiple local maxima resulting in inaccurate and nonsensical estimation. Therefore, we need to check parameters estimation results using different initial values and holding some parameters constant until we reach a confidence level (based on personal experience and heuristical optimization) that the estimation results are both logically and statistically reasonable. Notice that the estimation standard error of λ in table 3.1 is relatively high. This occurs because

of the term $\alpha^* = \alpha - \lambda = \mu - \frac{\sigma^2}{2\kappa} - \lambda$ in the measurement equation 3.33 generating alternative optimal solutions for the MLE function of the model.

Table 3.1: Estimated parameters for one-factor model using the futures contracts of 1,2,3,4 and 5 months to maturities (the daily prices of WTI crude oil stating from January 2, 2001 to September 6, 2011).

	μ	σ	κ	λ	ζ
value	4.543	0.3018	0.28	0.4876	0.0341
Sd.Err	0.3309	0.0064	0.00	0.3311	0.0022

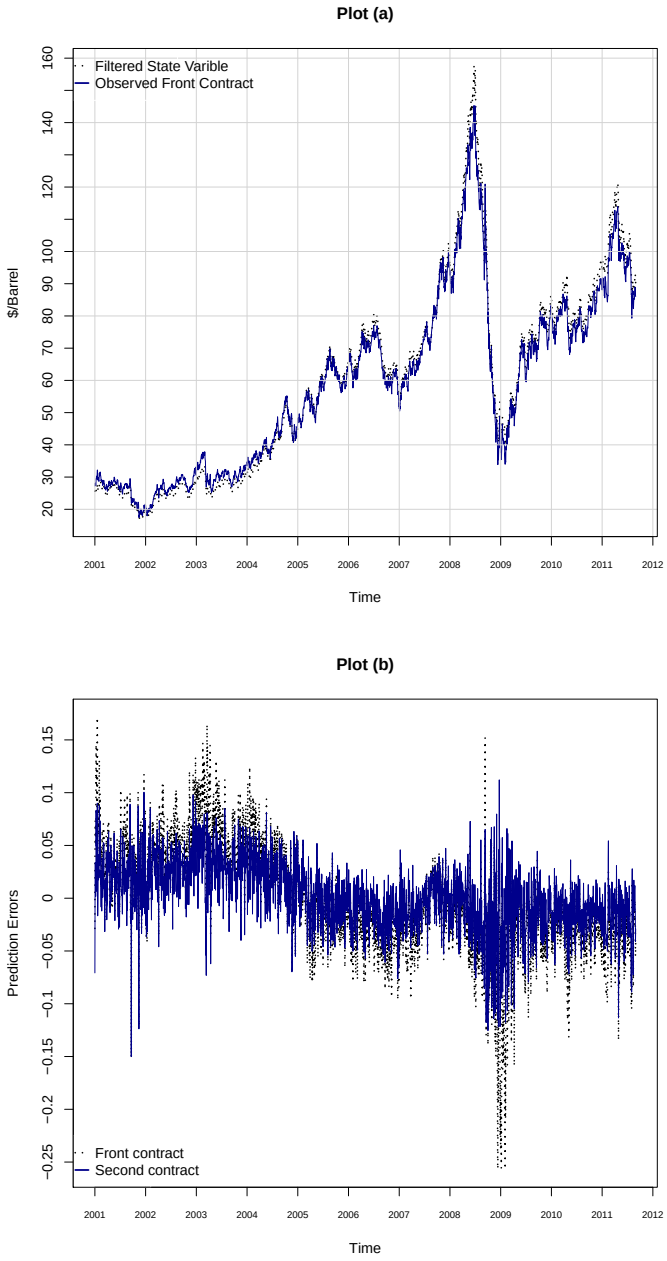

Figure 3.4: Time plot of the daily front futures prices for WTI crude oil from January 2, 2001 to September 6, 2011 using one-factor method: Plot (a) shows the estimated state variable (filtered spot price) and the front contract futures prices. Plot (b) shows the daily futures prediction errors corresponding to front and second futures contracts (the futures contracts of 1,2,3,4 and 5 month maturities are used in the estimation).

3.6 Two-Factor model for Commodity Prices:

Many empirical and theoretical studies including Gibson and Schwartz (1990) show that the convenience yield plays an important role in connecting futures prices to spot prices. Also Carmona and Ludkovski (2004) show that forward curves for commodities such as crude oil have backwardation, contango or more complex shapes over time which conflicts with the assumption of constant convenience yield. Gibson and Schwartz (1990) introduced a two-factor model to price commodity (specifically crude oil) contingent claims. In this model, two state variables are the commodity spot price S_t which follows a geometric Brownian motion with a stochastic drift and the instantaneous convenience yield denoted by δ_t. δ_t is commonly used in commodities pricing and assumed to follow a mean-reverting Vasicek process. Since this model has two factors, the model is flexible and can explain the reality of commodities forward curves and volatility structures in comparison to a one-factor model. The joint stochastic process of this model is specified as:

$$dS_t = (\mu - \delta_t) S_t \, dt + \sigma_1 S_t \, dW_{1t}, \tag{3.35}$$

$$d\delta_t = \kappa (\alpha - \delta_t) \, dt + \sigma_2 \, dW_{2t}. \tag{3.36}$$

where

the standard Brownian motion increments dW_{1t} and dW_{2t} are correlated with:

$$dW_{1t} \, dW_{2t} = \rho \, dt,$$

μ is the spot price drift,
σ_1 and σ_2 are the volatility of spot price and convenience yield respectively,
$\kappa > 0$ is the mean-reverting rate of convenience yield,
α is the long-run mean level of convenience yield.

In this model, it can be observed that if convenience yield is considered as a deterministic function of spot price S_t: $\delta(S_t) = \kappa \ln S_t$ instead of stochastic process, this model reduces to the one-factor model explained in section 3.5. Since convenience yield is the positive gain of holding a storable commodity minus its storage cost, it can take either negative or positive values. The dynamics of convenience yield in equation 3.36 can allow convenience yield be negative. From equation 3.35, we can observe that this model considers commodity's spot price like an asset with stochastic dividend yield δ_t. The joint dynamics of this model under a risk-neutral measure Q is specified as:

$$dS_t = (r - \delta_t) S_t \, dt + \sigma_1 S_t \, dW^*_{1t}, \tag{3.37}$$

$$d\delta_t = [\kappa (\alpha - \delta_t) - \lambda] \, dt + \sigma_2 \, dW^*_{2t}, \tag{3.38}$$

$$dW^*_{1t} \, dW^*_{2t} = \rho \, dt.$$

where λ is the market price of risk (assumed constant) associated with convenience yield δ_t, r is risk free interest rate which is assumed constant, and dW^*_{1t} and dW^*_{2t} are the standard Brownian motion increments under the equivalent risk-neutral measure Q.
Let $X_t = \ln S_t$ and let $Y_t = e^{\kappa t}\delta_t$. By applying Itô's lemma to equations 3.37 and 3.38, we derive the solution for the joint stochastic differential equation 3.37, 3.38 for any $s\,(0 < t \leq s)$ as follows:

$$X_s = X_t + (r - \frac{\sigma_1^2}{2})(s-t) - \int_t^s \delta_u du + \sigma_1 \int_t^s \mathrm{d}W^*_{1u}, \tag{3.39}$$

$$\delta_s = e^{-\kappa(s-t)}\delta_t + \hat{\alpha}(1 - e^{-\kappa(s-t)}) + \sigma_2 e^{-\kappa s}\int_t^s e^{\kappa v}\mathrm{d}W^*_{2v}. \tag{3.40}$$

where $(\hat{\alpha} = \alpha - \frac{\lambda}{\kappa})$,

To explicitly evaluate the integral $\int_t^s \delta_u du$, we plug equation 3.40 into δ_u, so we obtain:

$$\int_t^s \delta_u du = \int_t^s e^{-\kappa(u-t)}\delta_t du + \int_t^s \hat{\alpha}(1 - e^{-\kappa(u-t)})\,du + \sigma_2 \int_t^s e^{-\kappa u}\int_t^u e^{\kappa v}\mathrm{d}W^*_{2v}du. \tag{3.41}$$

By changing the order of integration in $\int_t^s e^{-\kappa u}\int_t^u e^{\kappa v}\mathrm{d}W^*_{2v}du$, we obtain:

$$\begin{aligned}\int_t^s \delta_u du &= \frac{\delta_t}{\kappa}(1 - e^{-\kappa(s-t)}) + \hat{\alpha}(s-t) - \frac{\hat{\alpha}}{\kappa}(1 - e^{-\kappa(s-t)}) \\ &\quad + \frac{\sigma_2}{\kappa}\int_t^s \left(1 - e^{-\kappa(s-v)}\right)\mathrm{d}W^*_{2v}.\end{aligned} \tag{3.42}$$

therefore, we substitute equation 3.42 into equation 3.39 to solve the joint stochastic process in equations 3.37 and 3.38 as follows:

$$\begin{aligned}X_s &= X_t + (r - \hat{\alpha} - \frac{\sigma_1^2}{2})(s-t) + \frac{\hat{\alpha} - \delta_t}{\kappa}(1 - e^{-\kappa(s-t)}) \\ &\quad - \frac{\sigma_2}{\kappa}\int_t^s \left(1 - e^{-\kappa(s-v)}\right)\mathrm{d}W^*_{2v} + \sigma_1 \int_t^s \mathrm{d}W^*_{1u},\end{aligned} \tag{3.43}$$

$$\delta_s = e^{-\kappa(s-t)}\delta_t + \hat{\alpha}(1 - e^{-\kappa(s-t)}) + \sigma_2 e^{-\kappa s}\int_t^s e^{\kappa v}\mathrm{d}W^*_{2v}. \tag{3.44}$$

Notice that to derive the solution for the joint stochastic process in equations 3.35 and 3.36, we can exchange r with μ and $\hat{\alpha}$ with α in equations 3.43 and 3.44.
Equations 3.43 and 3.44 show that the joint distribution of X_s and δ_s is bivariate normal

distribution where their expectations are:

$$\begin{aligned} E_Q[\,X_s \mid X_t,\delta_t\,] &= \mu_X = X_t + (\,r - \hat{\alpha} - \frac{\sigma_1^2}{2}\,)\,(s-t) \\ &\quad - \frac{1-e^{-\kappa(s-t)}}{\kappa}\delta_t + \frac{\hat{\alpha}}{\kappa}(1-e^{-\kappa(s-t)}), \qquad (3.45) \\ E_Q[\,\delta_s \mid \delta_t\,] &= \mu_\delta = e^{-\kappa\,(s-t)}\,\delta_t + \alpha^*\,(\,1 - e^{-\kappa\,(s-t)}\,), \qquad (3.46) \end{aligned}$$

We deploy Ito's isometry in stochastic integrals to calculate variances and covariance of X_s and δ_s as follows:

$$\begin{aligned} Var_Q[\,X_s \mid X_t,\delta_t\,] &= \sigma_X^2 = \frac{\sigma_2^2}{\kappa^2}\,E_Q\left[\left(\int_t^s \left(1-e^{-\kappa(s-u)}\right)\mathrm{d}W_{2u}^*\right)^2\right] \\ &\quad - \frac{2\sigma_1\sigma_2}{\kappa}\,E_Q\left[\int_t^s \left(1-e^{-\kappa(s-u)}\right)\mathrm{d}W_{2u}^*\int_t^s \mathrm{d}W_{1u}^*\right] \\ &\quad + \sigma_1^2\,E_Q\left[\left(\int_t^s \mathrm{d}W_{1u}^*\right)^2\right] \\ &= \frac{\sigma_2^2}{\kappa^2}\left\{(s-t) - \frac{2}{\kappa}(1-e^{-\kappa(s-t)}) + \frac{1}{2\kappa}(1-e^{-2\kappa(s-t)})\right\} \\ &\quad - \frac{2\sigma_1\sigma_2\rho}{\kappa}\left\{(s-t) - \frac{1}{\kappa}(1-e^{-\kappa(s-t)})\right\} + \sigma_1^2\,(s-t). \qquad (3.47) \end{aligned}$$

$$\begin{aligned} Var_Q[\,\delta_s \mid \delta_t\,] = \sigma_\delta^2 &= \frac{\sigma_2^2}{2\kappa}\,(\,1 - e^{-2\kappa\,(s-t)}\,), \qquad (3.48) \\ Cov_Q[\,(X_s,\delta_s) \mid (X_t,\delta_t)\,] &= \sigma_{X\delta} = \frac{\sigma_2^2}{2\kappa^2}\left(1-e^{-2\kappa(s-t)}\right) \\ &\quad + \frac{1}{\kappa}\left(\sigma_1\sigma_2\rho - \frac{\sigma_2^2}{\kappa}\right)\left(1-e^{-\kappa(s-t)}\right). \qquad (3.49) \end{aligned}$$

3.6.1 The Futures Price for Two-Factor Model:

In section 3.5, we explained that under a risk neutral measure Q, the futures price at time t with maturity T must be equal to $F(t\,,T\,,S_t,\delta_t) = E_Q[\,S_T \mid X_t,\delta_t\,]$. Since $X_T = \ln(S_T)$, the conditional distribution of S_T given $S_t\,,\ \delta_t$ under the risk-neutral measure Q is log-normal with the same parameters. Therefore, we use equations 3.45 and 3.47 to derive futures price, as follows:

$$\begin{aligned} F(t\,,\,T\,,\,S_t,\delta_t) &= E_Q[\,S_T \mid S_t,\delta_t\,] = \exp\left\{\,\mu_X + \frac{1}{2}\,\sigma_X^2\,\right\} \\ &= S_t\,\exp\left\{-\frac{1-e^{-\kappa(T-t)}}{\kappa}\delta_t + A(T,t)\right\}, \end{aligned} \tag{3.50}$$

where

$$\begin{aligned} A(T,t) &= \left\{\,r - \hat{\alpha} + \frac{\sigma_2^2}{2\kappa^2} - \frac{\sigma_1\sigma_2\rho}{\kappa}\,\right\}(T-t) + \frac{\sigma_2^2(1-e^{-2\kappa(T-t)})}{4\kappa^3} \\ &\quad + \left\{\hat{\alpha}\kappa + \sigma_1\sigma_2\rho - \frac{\sigma_2^2}{\kappa}\right\}\frac{1-e^{-\kappa(T-t)}}{\kappa^2}. \end{aligned} \tag{3.51}$$

We take logarithms in equation 3.50, and we obtain:

$$\ln F(t\,,\,T\,,\,S_t,\delta_t) = \ln S_t - \frac{1-e^{-\kappa(T-t)}}{\kappa}\delta_t + A(T,t). \tag{3.52}$$

We apply Ito's lemma to equation 3.52, in the risk neutral measure Q, and we see that $d\ln F(t\,,\,T\,,\,X_t,\delta_t)$ where $X_t = \ln S_t$ satisfies:

$$\begin{aligned} d\ln F(t\,,\,T\,,\,X_t,\delta_t) &= \frac{\partial \ln F}{\partial t}\,dt + \frac{\partial \ln F}{\partial X}\,dX_t + \frac{\partial \ln F}{\partial \delta}\,d\delta_t \\ &\quad + \frac{1}{2}\frac{\partial^2 \ln F}{\partial X^2}\,dX_t^2 + \frac{\partial^2 \ln F}{\partial X\partial\delta}\,dX_t\,d\delta_t + \frac{1}{2}\frac{\partial^2 \ln F}{\partial \delta^2}\,d\delta_t^2 \\ &= \left\{-\frac{\sigma_1^2}{2} - \frac{\sigma_2^2}{2\kappa^2} + \frac{\sigma_1\sigma_2\rho}{\kappa}\right\}dt - \frac{\sigma_2^2}{2\kappa^2}\,e^{-2\kappa(T-t)}\,dt \\ &\quad + \left\{\frac{\sigma_2^2}{\kappa} - \sigma_1\sigma_2\rho\right\}\frac{e^{-\kappa(T-t)}}{\kappa}dt \\ &\quad + \sigma_1\,\mathrm{d}W_{1t}^* - \frac{1-e^{-\kappa(T-t)}}{\kappa}\,\mathrm{d}W_{2t}^*. \end{aligned} \tag{3.53}$$

The result in equation 3.53 shows that in this model, the volatility of the log futures returns is given by:

$$\sigma_{lnF}^2 = \sigma_1^2 + \frac{\sigma_2^2}{\kappa^2}\left(1-e^{-\kappa(T-t)}\right)^2 - 2\sigma_1\sigma_2\rho\,\frac{\left(1-e^{-\kappa(T-t)}\right)}{\kappa}. \tag{3.54}$$

In equation 3.54 when the maturity approaches to infinity, the volatility converges to constant value as follows:

$$\lim_{T\to\infty} \sigma^2_{lnF} = \sigma_1^2 + \frac{\sigma_2^2}{\kappa^2} - \frac{2\sigma_1\sigma_2\rho}{\kappa}. \tag{3.55}$$

3.6.2 Parameter Estimation:

Both state variables spot price S_t and convenience yield δ_t are unobserved; therefore, we deploy the Kalman filter algorithm to calibrate this two-factor model. We construct the transition and observation equations for this model as follows:

3.6.2.1 Transition Equation:

We derive the exact transition equation by discretizing equations 3.43 and 3.44 in which we replace r with μ and $\hat{\alpha}$ with α in the following format:

$$[\boldsymbol{X}_{t+1}, \boldsymbol{\delta}_{t+1}]' = \boldsymbol{d}_t + \boldsymbol{T}_t [\boldsymbol{X}_t, \boldsymbol{\delta}_t]' + \boldsymbol{\eta}_t, \;\; t = 1, \ldots, NT, \tag{3.56}$$

where

$$\begin{aligned} \boldsymbol{d}_t &= \begin{bmatrix} (\mu - \alpha - \frac{\sigma_1^2}{2})\Delta t + \frac{\alpha}{\kappa}(1 - e^{-\kappa\Delta t}) \\ \alpha(1 - e^{-\kappa\Delta t}) \end{bmatrix}, \\ \boldsymbol{T}_t &= \begin{bmatrix} 1 & \frac{e^{-\kappa\Delta t} - 1}{\kappa} \\ 0 & e^{-\kappa\Delta t} \end{bmatrix}. \end{aligned} \tag{3.57}$$

Here $\boldsymbol{\eta}_t$ is serially uncorrelated Gaussian white noise with:

$$E(\boldsymbol{\eta}_t) = \boldsymbol{0} \quad \text{and} \quad Var(\boldsymbol{\eta}_t) = \begin{bmatrix} \sigma_X^2 & \sigma_{X\delta} \\ \sigma_{X\delta} & \sigma_\delta^2 \end{bmatrix}, \tag{3.58}$$

with

$$\begin{aligned}
\sigma_X^2 &= \frac{\sigma_2^2}{\kappa^2}\left\{\Delta t - \frac{2}{\kappa}(1-e^{-\kappa\Delta t}) + \frac{1}{2\kappa}(1-e^{-2\kappa\Delta t})\right\} \\
&\quad -\frac{2\sigma_1\sigma_2\rho}{\kappa}\left\{\Delta t - \frac{1}{\kappa}(1-e^{-\kappa\Delta t})\right\} + \sigma_1^2\,\Delta t, \\
\sigma_\delta^2 &= \frac{\sigma_2^2}{2\kappa}(\,1\,-\,e^{-2\kappa\,\Delta t}), \\
\sigma_{X\delta} &= \frac{\sigma_2^2}{2\kappa^2}\left(1-e^{-2\kappa\Delta t}\right) + \frac{1}{\kappa}\left(\sigma_1\sigma_2\rho - \frac{\sigma_2^2}{\kappa}\right)\left(1-e^{-\kappa\Delta t}\right).
\end{aligned} \tag{3.59}$$

3.6.2.2 Measurement Equation:

We derive the measurement equation be using equation 3.52 and adding Gaussian white noise as follows:

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t\left[\boldsymbol{X}_t, \boldsymbol{\delta}_t\right]' + \boldsymbol{\varepsilon}_t \quad t = 1, \ldots, NT, \tag{3.60}$$

where

$$\begin{aligned}
\boldsymbol{y}_t &\;\; is \;\; an \;\; N\times 1 \text{ vector of observations}, \;\; \boldsymbol{c}_t \text{ is } N\times 1 \text{ vector and } \boldsymbol{Z}_t \; is \; N\times 2 \text{ matrix}, \\
\boldsymbol{y}_t &= \left[\ln F(t\,,\,T_i)\right], \quad i = 1, \ldots, N, \\
\boldsymbol{c}_t &= \left[A(T_i, t)\right], \quad i = 1, \ldots, N, \\
\boldsymbol{Z}_t &= \left[1\;,\; \frac{e^{-\kappa(T_i-t)}\,-\,1}{\kappa}\right], \quad i = 1, \ldots, N.
\end{aligned} \tag{3.61}$$

where $A(T_i, t)$ is given in equation 3.51 and $\boldsymbol{\varepsilon}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\varepsilon}_t) = 0$ and $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{H}$. Since the number of parameters to estimate is quite large, we consider the variance of the future disturbances with all maturities to be identical which means $\boldsymbol{H} = \zeta^2\boldsymbol{I}_N$.

3.6.3 Empirical Results

We consider the daily observations of WTI crude oil futures prices from January 2, 2001 to September 6, 2011. We use this daily empirical data to estimate this two-factor model's

parameters. We use future contracts with 1,2,3,4 and 5 months maturities. Table 3.2 summarizes the calibration results for this two-factor commodity pricing model. Figure 3.5, plot (a) depicts the comparison between the filtered spot price and the front contract futures prices realizations and plot (b) demonstrates the estimated filtered convenience yield in this model. As we can notice from figure 3.2, implied convenience yield derived using 2 and 3 months to maturity, and figure 3.5, estimated convenience yield using two factor model, are relatively close.

Table 3.2: Estimated parameters for this two-factor model using the futures contracts of 1,2,3,4 and 5 months to maturities (the daily prices of WTI crude oil staring from January 2, 2001 to September 6, 2011).

	μ	σ_1	κ	α	σ_2	ρ	λ	ζ
value	0.2040	0.2911	1.5087	0.0272	0.3096	0.7606	0.00	0.08
Sd.Err	0.0706	0.0393	0.2141	0.0352	0.1708	0.5423	0.00	0.00

Schwartz (1997) proposes a three-factor model for commodity derivatives. In this model, three state variables are the commodity spot price s_t, the instantaneous convenience yield δ_t and the interest rate r_t. Schwartz (1997) adds a third interest rate factor into the two-factor model, explained in this section, and the assumed interest rate follows a simple Vasicek mean-reverting process. Since the volatility of convenience yield dominates the volatility of risk-free interest rate (has higher order of magnitude volatility), considering interest rate as third factor (stochastic), will not significantly improve the model accuracy as shown by Carmona and Ludkovski (2004). Beside in the futures (forward) price models, what matters most is the interest rate and convenience yield difference. Therefore, by fixing the interest rate (constant), all volatility of interest rates will be captured by stochastic process of convenience yield.

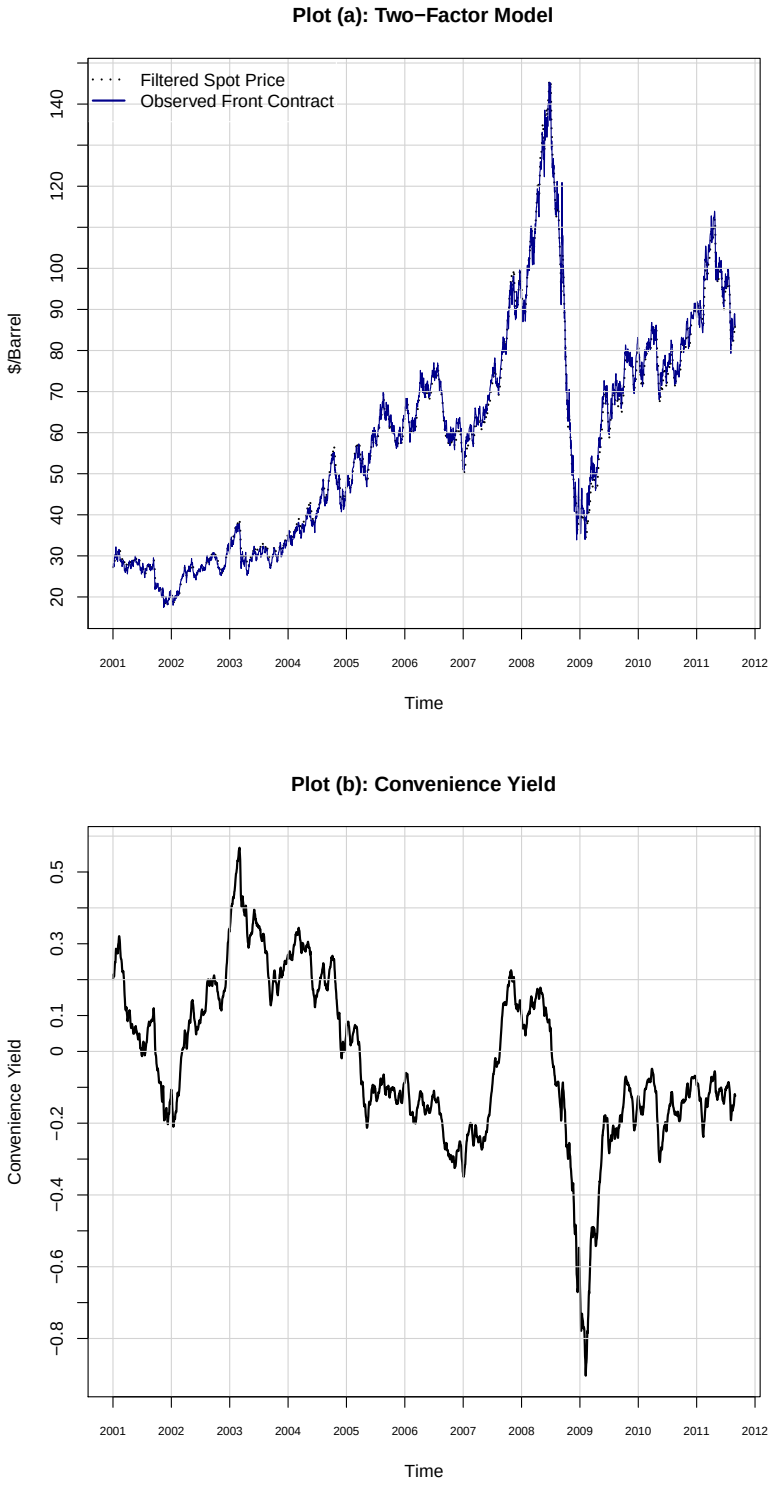

Figure 3.5: Time plot of the daily front futures prices for WTI crude oil from January 2, 2001 to September 6, 2011: Plot (a) shows the estimated filtered spot prices and the front contract futures prices derived by two-factor model. Plot (b) shows the estimated filtered convenience yield derived by two-factor model (the futures contracts of 1,2,3,4 and 5 month maturities are used in the estimation).

3.7 Generalized One-factor model for the commodity spot pricing

The mean-reverting property of commodity prices has been taken into account in several studies including Gibson and Schwartz (1990), Bessembinder *et al.* (1995) and Schwartz (1997). Schwartz (1997), reviewed in section 3.5, considered a one-factor mean-reverting model in which log-price follows a Vasicek process. The model has some disadvantages including an inability to fully explain either the reality of futures prices term structure and the high volatility of commodity prices, caused by remarkable uncertainty in the commodity markets. Here, we introduce a one-factor model that not only can capture the mean reverting nature of commodity prices but also can capture the undeniable high volatility of a commodity spot price by deploying a nonlinear drift term in the process. In this model, we assume that a commodity spot price, S_t, the only state variable, follows the dynamics:

$$dS_t = \kappa \left(\mu - \frac{(\ln S_t - \theta)}{\sqrt{(\ln S_t - \theta)^2 + 1}} \right) S_t dt + \sigma S_t dW_t, \tag{3.62}$$

where

S_t is the spot price at time t,
μ is a real number and $-1 < \mu < 1$,
θ is a real number, σ is the volatility of the process,
κ is the speed of mean reversion,
dW_t is the increment of a standard Brownian motion.

Notice that the condition $-1 < \mu < 1$ guarantees the log-price series to be a mean-reverting process. θ gives the model freedom to explore log-run mean for the log-process. Consider X_t as the logarithm of the spot price S_t at time $t\,(X_t = \ln(S_t))$. Applying Ito's lemma on 3.62, the log spot price follows the stochastic process:

$$dX_t = \kappa \left(\alpha - \frac{(X_t - \theta)}{\sqrt{(X_t - \theta)^2 + 1}} \right) dt + \sigma dW_t, \tag{3.63}$$

where $\alpha = \mu - \frac{\sigma^2}{2\kappa}$.
To simplify the model, the interest rate is assumed to be constant. Although convenience yield plays important role in commodity pricing models, the modeling of convenience yield dynamics is not an intention of this model. This model under a risk-neutral measure Q can be expressed as follows:

$$dX_t = \kappa \left(\alpha^* - \frac{(X_t - \theta)}{\sqrt{(X_t - \theta)^2 + 1}} \right) dt + \sigma dW_t^*, \tag{3.64}$$

where $\alpha^* = \alpha - \lambda$ (λ is the market price of risk and assumed to be constant), and dW_t^* denotes the increment of a standard Brownian motion under risk-neutral measure Q.

The volatility of the WTI crude oil price is very high, so the constant assumption of log-price in Schwartz (1997)'s one-factor is not quite appropriate. In this new model, although the diffusion term for log-price is still assumed to be constant, the model can not only capture the potential mean-reverting characteristics of the log-price evolution but it can also capture high volatility of the log-price of a commodity using its nonlinear drift term. Since Schwartz (1997)'s one-factor model assumed the commodity spot price transition density follows geometric Brownian motion, the model is incapable of capturing left skewness and it cannot also capture skewness in the log-price series. This new model, however, can capture this phenomena when the transition density is skewed to the left and it can also capture right skewness of log-price dynamics. To demonstrate this prominent property and compare with Schwartz (1997)'s model process, we simulate $10,000$ paths for both processes for five years and plot empirical densities depicted in figure 3.6. The plot 3.6 evidently shows that the empirical log simulated transition density for the new model is skewed to the left whereas the empirical log simulated transition density for the other model is symmetric. When μ is negative, the log-process in the new model is skewed to the left and vice versa.

3.7.1 The Stationary Solution

To show that a SDE has mean-reverting feature, we show that it has a stationary distribution (A SDE such as 3.63 has a stationary distribution if a unique solution, denoted by $P_{st}(x)$, exists such that $P_{st}(x) = \lim_{t\to+\infty} P(x\,,\,t)$ where $P(x\,,\,t)$ is the transition distribution). Here, we derive the stationary distribution for the new process's log-spot SDE given in equation 3.63 as follows:

$$P_{st}(x) = \frac{e^{-\frac{2\kappa\left(\sqrt{(x-\theta)^2+1}-\mu x\right)}{\sigma^2}}}{c}, \tag{3.65}$$

where c is the normalization factor and equals to:

$$c = \int_{-\infty}^{\infty} e^{-\frac{2\kappa\left(\sqrt{(x-\theta)^2+1}-\mu x\right)}{\sigma^2}} dx. \tag{3.66}$$

We could not solve this integral 3.66 analytically; however, we can simply use numerical methods to calculate c.

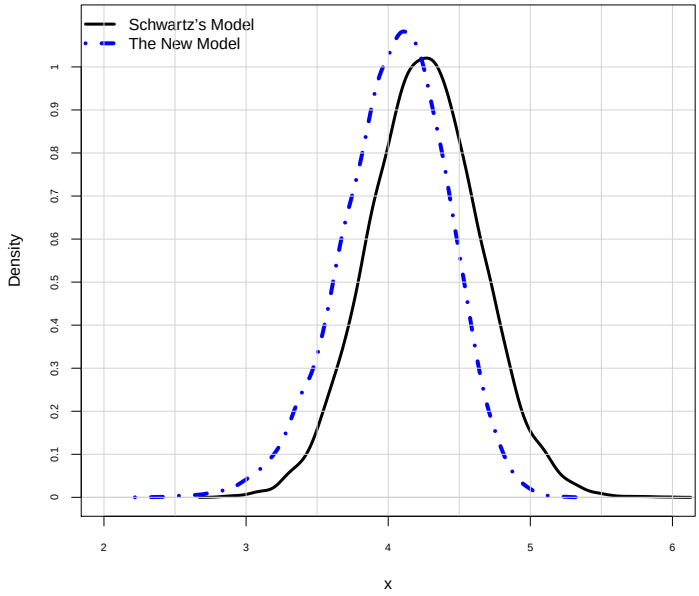

Figure 3.6: The plot depicts comparison between empirical transition densities for log-processes in the new model and Schwartz (1997)'s models after five years using $10,000$ simulated paths. Here it is assumed $\kappa = 3.2$, $\mu = -0.86$, $\sigma = 0.3$, and $\theta = 6$ for this new process 3.62 and we also assumed $\kappa = 0.27$, $\mu = 4.62$, $\sigma = 0.3$ for the Schwartz (1997)'s process 3.19. For both models, $\Delta t = \frac{1}{252}$, and $T = 5$ are assumed using identical random generated numbers.

3.7.2 Approximation of Transition Density

It can be shown that this new nonlinear SDE 3.62 has a unique transition distribution; however, all attempts to solve the SDE analytically failed. As a result, we deploy a numerical method to approximate the transition density. Moreover, we apply the *new local linearization method* (NLLM) developed by Shoji and Ozaki (1998) to estimate the model parameters. Since the method considers the stochastic behavior of the SDE's drift and diffusion terms in its approximation, the method has a high degree of accuracy. As reviewed in chapter 2.4.1, the nonlinear SDE is required to be in general form 2.49 with given conditions. Then the general stochastic process must to be transformed to a more tractable form as follows;

$$dX_t = h(X_t, t, \boldsymbol{\theta})\, dt + \sigma\, dW_t. \tag{3.67}$$

We derive this form 3.67 for our new dynamics 3.62 and give in equation 3.63. The log-price process's drift $h(x,t,\boldsymbol{\theta}) = \kappa\left(\alpha - \frac{(x-\theta)}{\sqrt{(x-\theta)^2+1}}\right)$ has all conditions required for this approximation method. When $t \in [s, s+\Delta s)$ and Δs is a small number, the SDE 3.67 can be linearized in the form as follows:

$$dX_t = (L_s X_t + M_s t + N_s)\,dt + \sigma\,dW_t, \tag{3.68}$$

where L_s, M_s, and N_s are constant and equal:

$$\begin{aligned} L_s &= \frac{\partial h(X_s,s,\boldsymbol{\theta})}{\partial x}, \\ M_s &= \frac{\sigma^2}{2}\frac{\partial^2 h(X_s,s,\boldsymbol{\theta})}{\partial x^2} + \frac{\partial h(X_s,s,\boldsymbol{\theta})}{\partial t}, \\ N_s &= h(X_s,s,\boldsymbol{\theta}) - X_s\frac{\partial h(X_s,s,\boldsymbol{\theta})}{\partial x} - s\,M_s. \end{aligned}$$

Here, the first derivatives with respect to x and t and second derivatives with respect to x of $h(x,t,\boldsymbol{\theta})$ are derived as follows:

$$\begin{aligned} \frac{\partial h(X_s,s,\boldsymbol{\theta})}{\partial x} &= \frac{-\kappa}{\left((X_s-\theta)^2+1\right)^{\frac{3}{2}}}, \\ \frac{\partial h(X_s,s,\boldsymbol{\theta})}{\partial t} &= 0, \\ \frac{\partial^2 h(X_s,s,\boldsymbol{\theta})}{\partial x^2} &= \frac{3\kappa(X_s-\theta)}{\left((X_s-\theta)^2+1\right)^{\frac{5}{2}}}. \end{aligned}$$

The dynamics 3.68 is in an Ornstein-Uhlenbeck stochastic process and using Itô' lemma, it can be easily shown that this SDE has an explicit solution as follows:

$$X_t = X_s + \frac{h(X_s,s,\boldsymbol{\theta})(e^{L_s(t-s)}-1)}{L_s} + \frac{M_s(e^{L_s(t-s)}-1-L_s(t-s))}{L_s^2} + \sigma\int_s^t e^{L_s(t-u)}\mathrm{d}W_u. \tag{3.69}$$

Equation 3.69 shows that the conditional distribution of X_T given X_s $(s<T)$ is approximately normally distributed with mean and variance as follows:

$$\begin{aligned} E(T,s) = E[X_T \,|\, X_s] &= X_s + \frac{h(X_s,s,\boldsymbol{\theta})(e^{L_s(T-s)}-1)}{L_s} \\ &\quad + \frac{M_s(e^{L_s(T-s)}-1-L_s(t-s))}{L_s^2}, \end{aligned} \tag{3.70}$$

$$V(T,s) = Var[X_T \,|\, X_s] = \frac{\sigma^2}{2L_s}\left(e^{2L_s(T-s)} - 1\right). \tag{3.71}$$

Notice that if we replace α by α^* in all involved terms in equations 3.70 and 3.71, we will derive the equivalent expected value and the equivalent variance under the risk-neutral measure Q.

3.7.3 The Futures Price for Generalized One-Factor Model:

Since the transition density for this new model is unknown, the futures price is approximated. The conditional distribution of S_T given S_t under the risk-neutral measure Q or real measure is approximated by log-normal distribution with the parameters given in equations 3.70 and 3.71. Therefore, the close approximation for futures price in this new model equals to:

$$\begin{aligned} F(t,T) &= E_Q[S_T \mid \mathfrak{F}_t] = e^{\{E_Q[X_T \mid \mathfrak{F}_t] + \frac{1}{2} Var_Q[X_T \mid \mathfrak{F}_t]\}} \\ &= e^{E(T,t) + \frac{1}{2}V(T,t)}. \end{aligned} \tag{3.72}$$

where α is substituted by α^* in all involved terms in equations 3.70 and 3.71.
By taking logarithm in equation 3.72, we derive:

$$\ln F(t,T) = E(T,t) + \frac{1}{2}V(T,t). \tag{3.73}$$

3.7.4 Parameter Estimation

Since the futures front contracts is extensively traded in certain exchanges, it is considered as proxy to the spot prices . The Kalman filter algorithm is mostly deployed to estimate unobserved state variable (here the spot prices). However, since transition and measurement equations that can be derived for this method are not linear, the linear Kalman filter algorithm cannot be applied to estimate the model's parameters. Instead, we apply the NLLM, explained in 3.7.2 algorithm to the front contracts as for this new model. To compare the new one-factor model and Schwartz's one-factor model, we also apply the NLLM to the same observations. It worth mentioning that the NLLM algorithm will generate exact maximum likelihood function for Schwartz's one-factor model. This is because the Schwartz's one-factor model is a linear stochastic process.

3.7.5 Empirical Results

Here, the front month futures prices are used as close approximation to the spot prices for WTI crude oil. The daily observations of WTI for front futures prices from October 1, 2004 to December 31, 2014, illustrated in left plot of figure 3.7, is chosen to study. For highlighting the difference between these two model more clearly, it can be seen that the

dataset used here is different form the one is used in the earlier sections. The right plot in figure 3.7 demonstrates the empirical density in this observed dataset. The empirical density of log-spot price clearly shows that the density is left skewed and leptokurtotic. Comparison between the Schwartz's one-factor model and this new generalized one-factor model is our purpose in this section. As it is shown in section 3.5, the transition density of the commodity spot price in Schwartz's one-factor model is log-normally distributed. The Schwartz's model, therefore, cannot capture the left skewness and excess kurtosis in the observed density of log-spot prices, shown in figure 3.7. Our model will capture these lacking important properties of real data. To calibrate the parameters of the Schwartz's one-factor model and this generalized one-factor model, the new local linearization method, reviewed in section 3.7.2, is deployed. The estimation results fitted to the observed dataset in both two models: Schwartz one-factor and the new generalized one-factor are summarized in table 3.3. In both models, κ is the mean-reverting speed and positive but κ of model is bigger than Schwartz's model κ due to different approach to capture mean-reversion. In this new model, μ is the skewness parameter whereas in Schwartz's model, μ is the long-run mean, so it is expected to be different. The estimated diffusion terms in both models σ are relatively close.

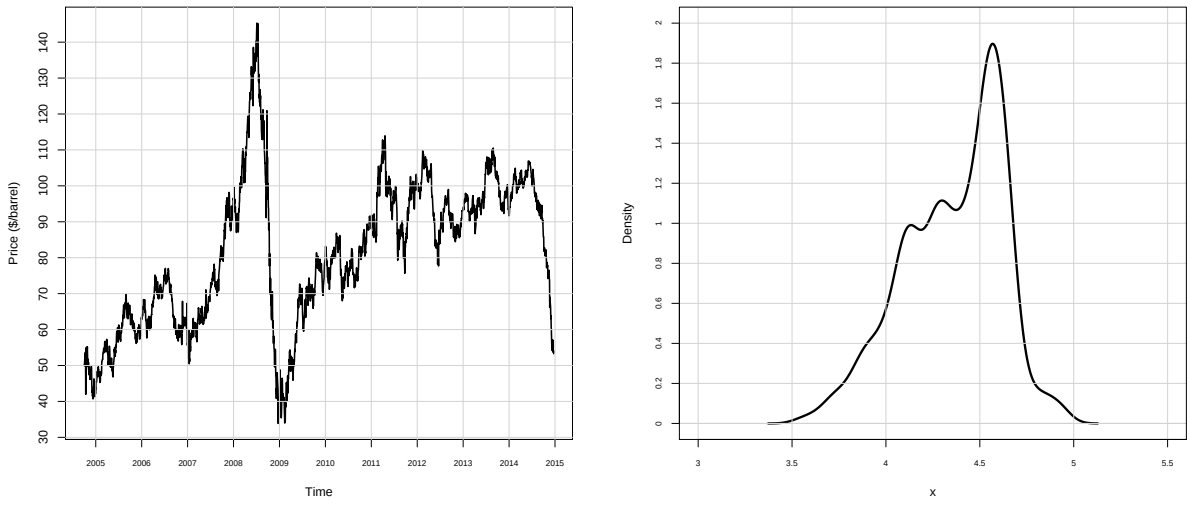

Figure 3.7: Left plot show time plot of the daily front futures prices for WTI crude oil from October 1, 2004 to December 31, 2014 (price in $/barrel). Right plot depicts the empirical density of the logarithm of the same dataset, given in the left plot.

Table 3.3: Estimated parameters for these two processes namely the new generalized one-factor and Schwartz one-factor models based on the observed daily WTI front futures prices from October 1, 2004 to December 31, 2014.

Model	κ	μ	σ	θ	-2 $logl$[1]	AIC[2]	BIC[3]
The New Process	3.182	-0.738	0.335	5.50	-12037.2	-12031.2	-12021.5
Schwartz Process	1.117	4.364	0.392	-	-11701.2	-11695.2	-11685.5

Note: 1-Log-likelihood, 2-Akaike information criterion, 2- Bayesian information criterion.

3.7.6 Goodness of Fit

Which of these two models namely the generalized one-factor model and Schwartz's one-factor model can describe the reality of the commodity pricing dynamics more clearly? This is the question of this section. The calibration results as well as the tests results for goodness of fit are given in table 3.3 for the observed daily WTI front futures prices from October 1, 2004 to December 31, 2014. Both AIC and BIC values assure that this generalized one-factor model entirely dominate Schwartz's one-factor model. Figure 3.8 illustrates the comparison between the empirical density for logarithm of the observed daily front futures prices of WTI with the fitted stationary distributions for both models derived using estimation results summarized in table 3.3. The plot depicts the superiority of the new one-factor model. It also shows that the new model is capable of capturing the most crucial properties of a commodity log-price dynamics such as skewness, heavy tails and kurtosis.

3.8 Conclusion

Pricing derivatives on commodities mainly rely on the assumed commodity pricing models. In this chapter, we first reviewed three important concepts in commodity markets namely forward contract, futures contract, and convenience yield. We then described three commodity pricing models namely preliminary, Schwartz (1997)' one-factor and Gibson and Schwartz (1990)' two-factor models in detail. In these models, we also derived futures and forward prices. This two-factor model is one of the most commonly applied for pricing by market participants. Later, we proposed a new generalized one-factor mean-reverting dynamics to model a commodity spot-price process. Since the new process has nonlinear dynamics with unknown transition density, the new local linearization algorithm was applied to estimate the model parameters. We argued that the new generalized SDE posses some key characteristics to capture the most essential properties including skewness, heavy tails and excess kurtosis, observed in a commodity

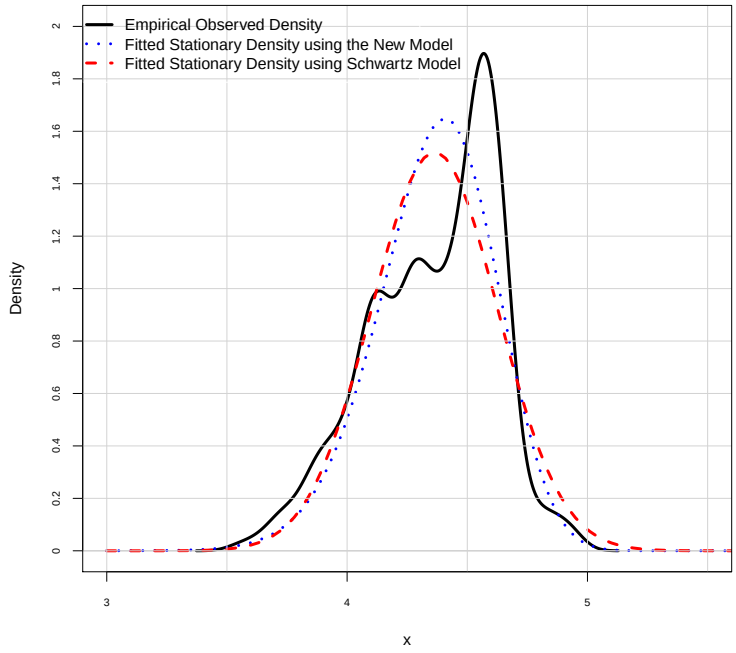

Figure 3.8: The plot shows comparison between the empirical density for the logarithm of observed daily WTI front futures prices staring from October 1, 2004 to December 31, 2014 with the fitted stationary distributions of these two models namely Schwartz one-factor and the new generalized one-factor for log-spot prices. The fitted parameters are summarized in table 3.3.

spot-price evolution. The new one-factor and Schwartz one-factor models were compared by fitting both models to daily front futures prices of WTI crude oil from October 1, 2004 to December 31, 2014. The calibration results confirmed the superiority of the new SDE. In chapters 4 and 6, we will directly or indirectly deploy these concepts and models in the spread models or trading strategies for the spread models with commodities as underlying.

Chapter 4

MODELING PRICES OF SPREADS

4.1 Introduction

Pairs trading is a "market neutral" trading strategy employed with the goal of making profit no matter whether the market has an upward or a downward trend. A market neutral strategy is a strategy the return of which is uncorrelated to the broader market return. Ideally, market neutral strategies also have a low volatility. Its return risk also remains largely unaffected by the market risk (see Vidyamurthy (2004)). The specific market neutral strategy of pairs trading is practiced by many hedge funds and money institutions. It was first introduced by Gerry Bamberger and later led by Nunzio Tartaglia of Morgan Stanley's quantitative department in the 1980s (Latte (2011)). Pairs trading, which is also called *spread trading*, has several categorizes namely *statistical arbitrage/relative value* trading and *risk arbitrage/merger arbitrage*. Relative pricing is the fundamental base for the statistical arbitrage pairs trading and risk arbitrage pairs trading can be applied in a merger agreement between two companies. Merger agreement has two main types: a *cash merger*, and a *stock merger*. In a cash merger agreement, the acquiring company offers to buy shares of the target company for an agreed cash price. Since there is uncertainty that the acquisition does not complete on time or the deal does not close at all, the target company's shares typically trade lower than expected value in which arbitragers step into buying target company's shares and simultaneously selling acquire company's shares. In a stock merger, the merger agreement imposes rigid parity relationship between underlying companies's stocks prices where the traders can take advantage and trigger their trades upon the specified parity relationship. In this study, we focus on statistical arbitrage. It is worth mentioning that despite the apparent contrast with risk arbitrage, a statistical arbitrage strategy is not a risk free strategy at all. In the equity markets, pairs trading is a investment strategy in which a portfolio is

simply two stocks with similar characteristics (for instance in the same market sector) that are out of their historical equilibrium level or *long-run mean*. The portfolio involves a long position (buy) in the underpriced stock, a short position (sell) in the overpriced stock with a predetermined ratio to maintain market neutrality. The overpriced stock and the underpriced stock are chosen based on relative pricing and do not depend on the actual prices of individual securities. Pairs trading can be deployed in commodity markets too. The price spread between gasoline and crude oil, called the *crack spread* and the location (geographical) spread of West Texas Intermediate (WTI) crude oil and Brent oil spot prices (front contracts) are two examples in this regard. The pairs in this strategy are selected based on historical long-term relationship between the underlying assets. Therefore, the spread process can be modeled as a mean-reverting dynamic process. Selecting trading pairs is a crucial and essential part of the strategy. Formerly, the pairs were chosen based on non-parametric methods or correlation. However, it is easy to show that these methods are no longer valid without careful investigating and testing data to see whether there is solid evidence to back up the long-run relationship both economically and statistically. Instead, the pairs should be chosen based on long-run relationship of underlying pairs known as mean-reversion. In other words, the trading pairs will be chosen if the pairs are *cointegrated* or a linear combination of price processes of underlying assets or commodities is a stationary process. When the spread is considerably away from its long-term equilibrium level, the strategy suggests taking a short position in the overvalued member of pair and long position in the undervalued pair simultaneously based on hedging ratio (linear combination). We expect that the deviation from the equilibrium level is temporary and the spread eventually reverts back to its long-run mean when the portfolio is in the money and profit can be realized by closing both positions.

This chapter is organized as follows: In section 4.2 we introduce cointegration and its application in pairs trading. Elliott *et al.* (2005) introduced a one factor mean-reverting Vasicek process to model the spread process that we review in section 4.3. In section 4.4, we review the two factor model proposed by Dempster *et al.* (2008) to model the spot spread process in energy markets to evaluate spread options. As we explained in chapter 3 , the spot prices for commodities such as crude and Brent oils are unobservable; as a result, we cannot use the two underlying commodities to test mean-reversion and find an appropriate *hedging ratio* for the commodities pairs. Instead, we can use the pair's future prices to model the spread process. In section 4.5, we apply these three models to our real empirical sample data and compare the results. Later, we analyze the recent behavioral change in the location spread between WTI crude oil and Brent oil. Finally, in sections 4.7, since highly important news can generate a shock in a spread process, we propose to implement a jump, which is compound Poisson process, in the one-factor and two-factor spread models introduced in sections 4.3 and 4.4 respectively. In these

models, jump sizes follow the double exponential distribution introduced by Kou (2002). Finally, in section 4.8, a new one-factor mean-reverting process is proposed to explain not only the mean-reverting property of the spread process, but also the skewness and the kurtosis characteristics of a spread process. The transition density of this nonlinear mean-reverting stochastic process is unknown so the new local linearization method, proposed by Shoji and Ozaki (1998) is used to estimate the model's parameters. Since the spread between WTI crude oil and Brent oil has recently experienced a structural break due to fundamental reasons, we deploy *Regime-Switching Models* (RSM) in this generalized one-factor mean-reverting dynamics to capture this phenomena in the spread process.

4.2 Cointegration Approach:

In recent times, much research has focused on the efficient markets hypothesis. The hypothesis argues that all existent information in the markets and economic realities have been incorporated into the asset prices (Serletis (2007)). In other words, the price changes are random and there is no arbitrage opportunity. In finance, an asset log return time series is commonly assumed to be *weakly stationary or covariance stationary*; however, asset price series, interest rates, or foreign exchange rates tend to be *non-stationary*. The standard classical estimation methods such as the ordinary least squares (OLS) method cannot be deployed to estimate the relationship between non-stationary variables because these methods are based on fundamental assumptions that the means and the variances of variables must be well-defined time-invariant constants. These assumptions fail to hold in the case of non-stationary variables. As such, applying these methods will give misleading inferences. This problem is referred as *spurious regression* in the statistical literature Phillips (1986). To analyze the long-run relationship between two non-stationary variables, we should take advantage of the well-known *cointegration* property. Now, it is time to define the terms and notions that we will frequently use in this research. A nonstationary process is a process whose moments (mainly mean and/or variance) change over time. A nonstationary time series, for instance, has no stable level in time: no constant long-run mean. A random walk as defined by x_t in equation 4.2 is an obvious example of a non-stationary process. A time series $\{r_t\}$ is called weakly stationary if (1) $E[r_t] = \mu$ (μ is a constant) and (2) $Cov(r_t\,,\, r_{t+\tau}) = \gamma_\tau$ (only depends on time difference τ not time t). When a time series such as $\{r_t\}$ is weakly stationary, any given observations, $\{r_{t_1}, r_{t_2}, \ldots, r_{t_n}\}$ will fluctuate with a fixed variance around a equilibrium level namely the long-run mean. If r_t is a stationary time series then a time series in the following form is called a *trend-stationary time series*:

$$y_t = \beta_0 + \beta_1 t + r_t \tag{4.1}$$

where β_1 is the linear growth rate. A simple check shows that y_t is autocorrelated. However, the autocorrelation of y_t is different from the autocorrelation of a random walk. This because by de-trending y_t will generate a stationary series, but a random walk is a non-stationary series ($E[W_t^2] = t$). In a trend-stationary series the mean $E[y_t] = \beta_0 + \beta_1 t$ is time dependent whereas the variance $Var[y_t] = Var(r_t)$, is constant. If in equation 4.1 r_t assumes as a simple random walk then both the mean $E[y_t] = \beta_0 + \beta_1 t$ and the variance $Var[y_t] = t\sigma^2$ are time dependent. A trend-stationary series can simply be transformed into a stationary process by excluding the time trend. The trend in many economic series cannot be justified by deterministic functions (deterministic trends). Indeed these series have stochastic trends. The idea of transforming a non-stationary process with stochastic trend known as *unit-root non-stationary* into a stationary process is called *differencing* (Hamilton (1994)). Assume y_t is a unit-root non-stationary process, y_t is *integrated of order* d denoted by $I(d)$ if after d times differencing, the transformed process has a stationary invertible $ARMA$ representation. In other words,

$$
\begin{aligned}
&y_t \sim I(d) \qquad \text{if } (1-B)^d y_t \text{ is a stationary process,}\\
&\text{where } B \text{ is back shift } By_t = y_{t-1} \text{ or:}\\
&(1-B)y_t = y_t - y_{t-1} = \Delta y_t, \qquad \text{and}\\
&(1-B)^2 y_t = (1-B)(1-B)y_t = (1-B)(y_t - y_{t-1})\\
&\qquad\qquad = (1-B)y_t - (1-B)y_{t-1} = \Delta y_t - \Delta y_{t-1} = \Delta^2 y_t \text{ and etc.}
\end{aligned}
$$

According to this definition, if y_t is already a stationary process, it is called $I(0)$ (integrated of order 0). One of the common errors in differencing is *over-differencing* and it should be avoided. For instance, differencing an $I(1)$ process twice is considered over-differencing. The lowest order of differencing that transforms a unit-root non-stationary process into a stationary process is the appropriate order of integration. In $ARMA(p,q)$, if we allow that the AR polynomial (characteristic equation of AR) has a root of unity as a characteristic root, then the extended $ARMA(p,q)$ is unit-root non-stationary and has a unit-root or, alternatively is integrated of order one ($I(1)$). This extended model is called the *Autoregressive integrated moving average (*$ARIMA(p, d = 1, q)$*) model*. Since the other roots of the characteristic equation are less than one in absolute value, the first difference of the time series will be weakly stationary. In this model, if one is a root of *multiplicity* order d, then the $ARIMA(p,d,q)$ is integrated of order d, denoted $I(d)$. To analyze relationship and interrelation between different markets and also to study jointly multiple assets, one can deploy *vector or multivariate time series*. Vector $ARMA(p,q)$ models ($VARMA(p,q)$) is the generalization of univariate $ARMA$ models. Unlike univariate $ARMA$, we may confront the issue of *identifiability*, the model may not be uniquely defined. To overcome this issue, *structural specification method* (Tsay (1991)) can be applied. In $VARMA(p,q)$, if MA polynomial has unit-roots, the $VARMA(p,q)$ models are said to be noninvertible. The spurious regression problem

occurs when we attempt to regress two unrelated non-stationary (for example two $I(1)$ variables) or trend stationary time series on each other and the ordinary least squares (OLS) method indicates existence of statistically significant relationship between them when such a relationship does not exist in reality.

Example 4.2.1. Consider regressing two independent random walks with drift on each other:

$$\begin{aligned} x_t &= \mu_1 + x_{t-1} + \varepsilon_{1,t} \\ y_t &= \mu_2 + y_{t-1} + \varepsilon_{2,t} \end{aligned} \tag{4.2}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are i.i.d Gaussian white noises.
We generate 500 samples from each random walks, x_t, y_t ($\mu_1 = 0.2$, $\mu_2 = 0.5$,) and then regress y_t on x_t. Figure (4.1) shows the simulated data. Although the Gaussian white noises are i.i.d, the correlation in simulated data is 0.9831.The regression results summarized in table (4.1) shows an incorrectly statistically significant relationship between them. However, the *Durbin Watson* (DW) statistic is 0.0434 with p-value $<$ 2.2e-16, which confirms that the residual series is autocorrelated, as a result, the mentioned relationship is no longer valid. In this model, when n (number of observations) gets large, the OLS estimator will approach to the ratio of two drifts ($\hat{\beta} \to \frac{\mu_2}{\mu_1}$); moreover, the OLS estimator will incorrectly indicate a statistically significant relationship. The spurious problem is not limited to trend nonstationary processes. In driftless nonstationary processes and trend stationary processes, we can also confront this issue. Differencing will resolve the problem, but when we use differencing, we lose some important information and the results typically have no economical or logical explanation.

Table 4.1: Regression results based on model (4.2) when $\mu_1 = 0.2$, $\mu_1 = 0.5$, and variances of innovations are 1. Number of simulations is 500 ($y \sim x$) and 56984 is used as seed to generate series.

	Estimate	Sd.Err	t value	Pr($>|t|$)
Intercept	-11.94971	1.22604	-9.747	<2e-16 ***
x	2.60172	0.02171	119.862	<2e-16 ***

Cointegration is deployed to estimate the long-run equilibrium levels between two unit-root variables such as two stocks in the same sector. In other words, two unit-root processes are called cointegrated if there exists a linear combination of them that is a stationary process. In general, let $\boldsymbol{X}_t$ be a vector of n integrated of order d ($I(d)$) time series. Cointegration exists if a vector $\boldsymbol{\beta}_t (\neq \mathbf{0})$ exists such that $\boldsymbol{Z}_t = \boldsymbol{\beta}^{'} \boldsymbol{X}_t$ and $\boldsymbol{Z}_t$ is

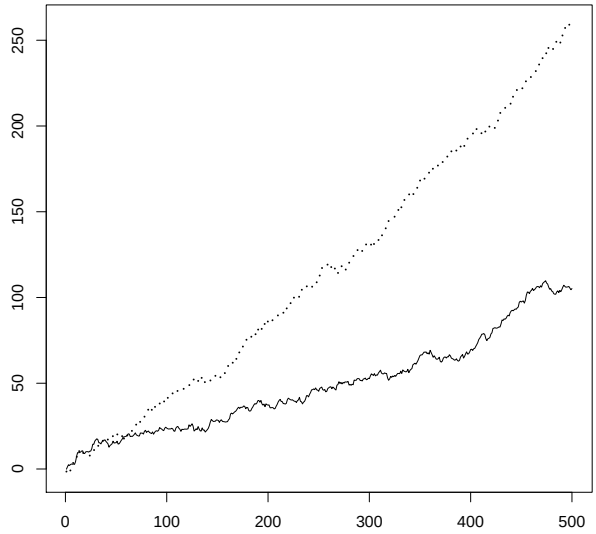

Figure 4.1: The time plots are simulated series for x_t (solid line) and y_t (dotted line) based on model 4.2 when $\mu_1 = 0.2,\ \mu_2 = 0.5,$ and variances of innovations are 1. Number of simulations is 500 and 56984 is used as seed to generate the series.

integrated of order $I(h)$ where $0 < h < d$. $\boldsymbol{\beta}$ is called the cointegrating vector. The $d - h$ is the number of common $I(1)$ stochastic trends. If cointegration exists between two financial markets and can be justified economically, then we have a possibility to deploy at least one of the markets to benchmark (forecast) the other one.

Example 4.2.2. Consider the bivariate time series as follows:

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + \mu_1 + \theta_1 z_{t-1} + \phi_1 \varepsilon_{1,t} \\ y_{2,t} &= y_{2,t-1} + \mu_2 + \theta_2 z_{t-1} + \phi_2 \varepsilon_{2,t} \\ z_t &= D_t + y_{1,t} + \beta\, y_{2,t} \end{aligned}$$

where D_t is a deterministic function of time and $(\varepsilon_{1,t}, \varepsilon_{2,t})$ is a Gaussian bivariate white noise with the positive definite covariance matrix Σ. In this setting, if z_t is a (trend) stationary time series, $y_{1,t}$ and $y_{2,t}$ are cointegrated. We derive the (trend) stationary condition for z_t as follows:

$$\begin{aligned} z_t &= D_t + y_{1,t} + \beta\, y_{2,t} \\ &= D_t + y_{1,t-1} + \mu_1 + \theta_1 z_{t-1} + \phi_1 \varepsilon_{1,t} + \beta y_{2,t-1} + \beta\mu_2 + \beta\theta_2 z_{t-1} + \beta\phi_2 \varepsilon_{2,t} \\ &= D_t - D_{t-1} + \mu_1 + \beta\mu_2 + (1 + \theta_1 + \beta\theta_2) z_{t-1} + \phi_1 \varepsilon_{1,t} + \beta\phi_2 \varepsilon_{2,t} \end{aligned} \tag{4.3}$$

Equation 4.3 shows that z_t is a $AR(1)$ process; therefore, the (trend) stationary condition for z_t is :

$$| 1 + \theta_1 + \beta\theta_2 | < 1 \tag{4.4}$$

It is also worth mentioning that both $y_{1,t}$ and $y_{2,t}$ are integrated of order one ($I(1)$). We simulate the following bivariate cointegrated time series to show how cointegration works:

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + 0.2z_{t-1} + \varepsilon_{1,t} \\ y_{2,t} &= y_{2,t-1} + z_{t-1} + \varepsilon_{2,t} \\ z_t &= y_{1,t} - 0.7\, y_{2,t} \end{aligned} \tag{4.5}$$

where the variance covariance matrix of the bivariate white noise, Σ is equal to I ($\Sigma = I$).

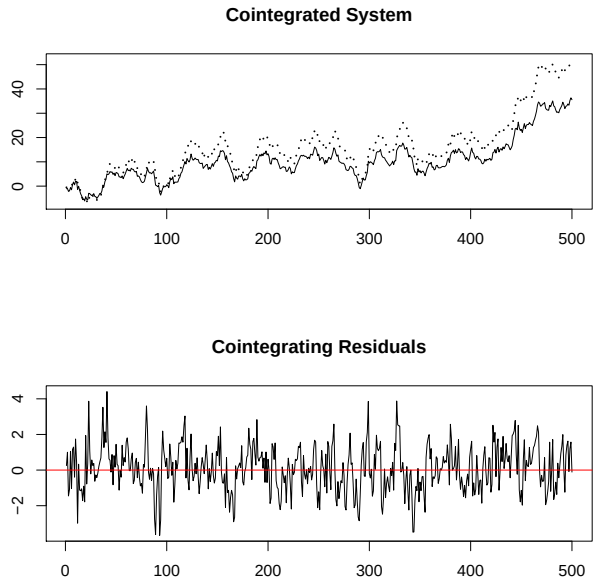

Figure 4.2: The time plots are simulated series for $y_{1,t}$ and $y_{2,t}$ and their cointegration residuals based on model (4.5) when variances of i.i.d white noises $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are 1. Number of simulations is 500 and 2987 is used as seed to generate series.

Four main steps are required to take when we attempt to apply cointegration technique. These steps are:

Step 1: Unit-Root Test: These tests indicate whether or not the involved time series individually have a unit-root. The *augmented Dickey-Fuller test (ADF)* is a well-known unit-root test. The ADF test assumes that the time series y_t has a general $ARMA(p,q)$ structure with unknown order. The null hypothesis and alternative hypothesis are defined as follows:

The null hypothesis (H_0): The time series y_t has a unit-root ($y_t \sim I(1)$).

The alternative hypothesis (H_1): y_t is a (trend) stationary series ($y_t \sim I(0)$).

The ADF test can be performed applying the following test regression:

$$y_t = D_t + \beta y_{t-1} + \sum_{i=1}^{p} \psi_i \Delta y_{t-i} + \varepsilon_t, \tag{4.6}$$

Or, equivalently:

$$\Delta y_t = D_t + \beta_c y_{t-1} + \sum_{i=1}^{p} \psi_i \Delta y_{t-i} + \varepsilon_t. \tag{4.7}$$

where D_t is a deterministic time function (constant, trend, or etc,) and chosen based on the assumed behavior of y_t, with the t ratio statistics respectively are:

$$\text{ADF-test:} \quad t_{(\hat{\beta}=1)} = \frac{\hat{\beta}-1}{sd(\hat{\beta})}, \tag{4.8}$$

$$\text{ADF-test:} \quad t_{(\hat{\beta}_c=0)} = \frac{\hat{\beta}_c}{sd(\hat{\beta}_c)}. \tag{4.9}$$

where $\hat{\beta}$ is the estimate of β using a least-squares method. Testing for a unit root is usually not an easy task and is statistically difficult for some series such as a stationary series contaminated by a structural shift. There are several tests for a unit root and the results, from one test to another, can be quite different. Consequently, we should use different tests and in particular more recent tests such as *Phillips & Perron test* (Phillips and Perron (1988)), *Elliot, Rothenberg & Stock test (ERS)* (Elliott *et al.* (1996)), or *Zivot & Andrews* (Zivot and Andrews (1992)) test.

Step 2: Cointegration Tests: Another crucial step is to test whether the study assets or commodities are cointegrated or not. If tests show the series are cointegrated, we can proceed and apply cointegrated $VECM$ model to estimate parameters including $\boldsymbol{\beta}$ (cointegrating vector). Cointegration tests may encounter hardships such as overlooking the scaling effects in actual applications Tsay (2010). Normally, we first choose, for instance, two stocks from the same sector that share the identical set

of fundamental properties (for example TD Canada Trust and Royal Bank) which we expect to be cointegrated. Then if cointegration tests such as the *Johansen test* Johansen (1991) confirm the existence of long-term equilibrium relationship, we can be confident that cointegration is not accidental and that the long-run relationship likely remain significant in the future. In this case, we can implement the cointegration trading strategies in the considered pairs.

Step 3: We estimate the short-term disequilibrium level (relative mispricing) based on the long-run equilibrium relationship that we estimate in the second step.

Step 4: We should check the robustness of short-term disequilibrium relationships, then we should define the pairs trading strategy if the short-term disequilibrium level reach certain levels.

The following example 4.2.3 is a cointegrated $VARMA(1,1)$ and depicts the cointegration system in vector form.

Example 4.2.3. Consider the following cointegrated $VARMA(1,1)$ model,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.25 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} - \begin{bmatrix} 0.1 & 0.2 \\ 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}, \tag{4.10}$$

where the covariance matrix of the bivariate white noise, and where $\Sigma = I$. The eigenvalues of the AR coefficient matrix are 0 and 1; therefore, the $VARMA(1,1)$ is not a stationary time series, Tsay (2010). First we should make sure that the two time series ($y_{1,t}$ and $y_{2,t}$) are integrated of order 1, so we reorganize the model as follows:

$$\begin{bmatrix} 1-0.5B & -0.25B \\ -B & 1-0.5B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1-0.1B & -0.2B \\ -0.2B & 1-0.4B \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \tag{4.11}$$

To diagonalize the AR matrix polynomial, we multiply equation 4.11 by:

$$\begin{bmatrix} 1-0.5B & 0.25B \\ B & 1-0.5B \end{bmatrix},$$

We derive,

$$\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1-0.6B & 0.05B \\ 0.8B & 1-0.9B \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}. \tag{4.12}$$

Figure (4.3) depicts that both component time series are highly autocorrelated. Also,

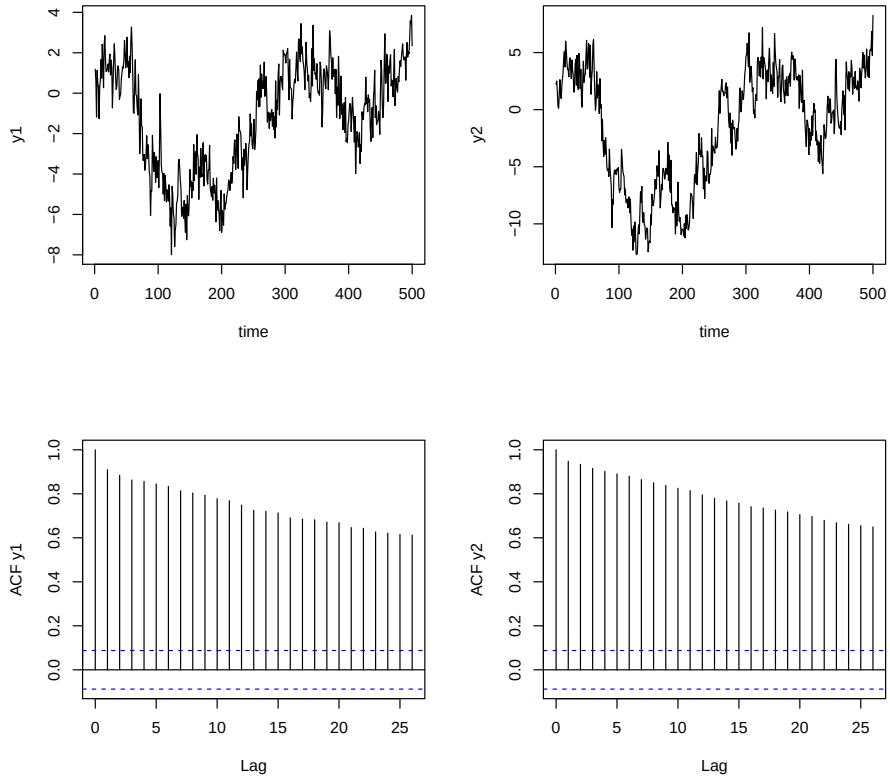

Figure 4.3: The time plots are simulated series for $y_{1,t}$ and $y_{2,t}$ and their autocorrelation functions based on model (4.10) when variances of i.i.d white noises $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are 1. Number of simulations is 500 and 6987 is used as seed to generate series.

the result in equation 4.12 shows that both marginal time series ($y_{1,t}$ and $y_{2,t}$) have a unit roots ($I(1)$). To show that in the model 4.10, $y_{1,t}$ and $y_{2,t}$ are cointegrated, we calculate the linear combination $z_t = y_{1,t} - 0.5y_{2,t}$ as follows,

$$z_t = \varepsilon_{1,t} - 0.5\varepsilon_{2,t}. \tag{4.13}$$

The result in equation 4.13 shows that z_t is a white noise series meaning that z_t is a stationary series; as a result, $y_{1,t}$ and $y_{2,t}$ are cointegrated.

As mentioned before, the vector time series may be noninvertible. To overcome the issue of estimating noninvertible cointegrated $VARMA(p,q)$ models, Engle and Granger (1987) proposed the *vector error correction model*($VECM$). We merely illustrate the bivariate cointegrated $VECM$ representations. The bivariate $VECM$ representations can be easily generalized. Interested readers can refer to Tsay (2010).

4.2.1 Bivariate Cointegrated $VECM(p,q)$

Consider $y_{1,t}$ and $y_{2,t}$ ($\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$) to be both integrated of order one ($I(1)$) and let $y_{1,t}$ and $y_{2,t}$ be cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, \beta)'$. Consequently, $\boldsymbol{\beta}'\boldsymbol{y}_t = y_{1,t} + \beta y_{2,t}$ is a stationary process ($I(0)$) , then a $VECM(p,q)$ representation exists in the form of:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} [1\,,\,\beta] \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \begin{bmatrix} \Delta y_{1,t-i} \\ \Delta y_{2,t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} - \sum_{i=1}^{q} \boldsymbol{\Theta}_i \begin{bmatrix} \Delta\varepsilon_{1,t-i} \\ \Delta\varepsilon_{2,t-i} \end{bmatrix}, \tag{4.14}$$

where the $\boldsymbol{\Phi}_i^*$ (AR polynomial coefficient matrices) can be derived using the $\boldsymbol{\Phi}_i$ of the original model as follows:

$$\begin{aligned} \boldsymbol{\Phi}_i^* &= -\sum_{j=i+1}^{p} \boldsymbol{\Phi}_j, \quad i = 1, \ldots, p-1, \\ \boldsymbol{\gamma}\boldsymbol{\beta}' &= -\sum_{i=1}^{p} \boldsymbol{\Phi}_i - \boldsymbol{I}, \end{aligned} \tag{4.15}$$

where

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

One can see that since $\Delta\boldsymbol{y}_t = (\Delta y_{1,t}, \Delta y_{2,t})'$ is stationary, but $\boldsymbol{y}_{t-1} = (y_{1,t-1}, y_{2,t-1})'$ is a nonstationary series, the $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ process must be a stationary in order to have a plausible relationship for the cointegrated $VECM$ representations. Consider the cointegrated $VARMA(1,1)$ in equation 4.10. Based on the prior explanations in this section, the equivalent $VECM$ representation of the equation will be:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} [1\,,\,-0.5] \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} - \begin{bmatrix} 0.1 & 0.2 \\ 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}. \tag{4.16}$$

It is worth mentioning that, in some studies, the cointegration tests are applied for log-prices. Vidyamurthy (2004), for example, adapts the cointegration test for a model in

which the log-price of stock A is regressed on the log-price of stock B. He then develops the pairs trading strategy using the following cointegration system:

$$\ln(p_t^A) = \mu + \beta \ln(p_t^B) + \varepsilon_t, \tag{4.17}$$

where β is the cointegration coefficient and μ is the long-run equilibrium level. In this system, if the two stocks are cointegrated, the estimated residuals must be a stationary series.

4.2.2 Empirical Results

Since as mentioned in chapter 3, spot WTI and Brent crude oils prices are unobservable, the front month future contracts can be considered as approximations of the spot prices for both crude oils. In order to compare our results with the results found in Dempster *et al.* (2008), we first test for cointegration in WTI crude oil and Brent oil monthly data. Chicago Mercantile Exchange (CME) monthly future prices (the front contract) of WTI crude oil and Intercontinental Exchange (ICE) monthly future prices (the front contract) of Brent oil are used from April 1994 to January 2005. Although it is generally accepted that WTI crude oil and Brent oil are unit root processes ($I(1)$), we use ADF and ERS tests to test for the unit root in WTI and Brent oils and their difference series to follow the cointegration algorithm's steps. The null hypothesis in both tests is that the series contains a unit root. The results of the tests are summarized in table 4.2. Both unit root tests for each Brent and WTI price series confirm that the null hypothesis should not be rejected in favor of the alternative hypothesis at the one percent level of significance. The tests also indicate that the null hypothesis for each differenced series of Brent and WTI should be rejected in favor of the alternative hypothesis at one percent level of significance. Therefore, the WTI crude oil series and Brent oil series are integrated of order 1 ($I(1)$). Figure 4.4 depicts the monthly front contracts future prices of WTI and Brent oils.

We applied the *Johansen test (JT)* to check whether WTI crude oil and Brent oil for our monthly data are cointegrated or not. The test results are summarized in table 4.3. The null hypothesis is no cointegration. The JT with maximal eigenvalue statistic rejects the null hypothesis in favor of the alternative hypothesis at 5 percent level of significance. Therefore, the WTI crude oil series and Brent oil series are cointegrated. The estimated $VECM$ representation is in equation 4.18.

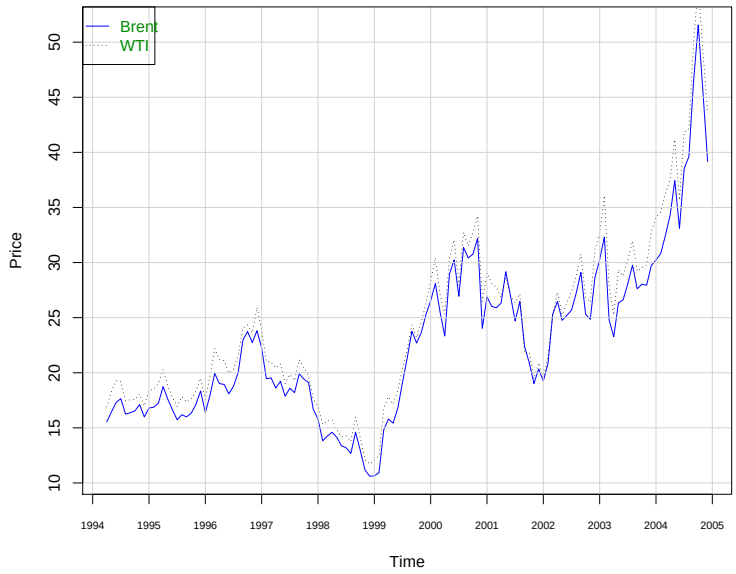

Figure 4.4: The monthly front contracts future prices for WTI crude oil and Brent oil from April 1994 to January 2005 (price in $/barrel).

$$\begin{bmatrix}\Delta y_{1,t}\\ \Delta y_{2,t}\end{bmatrix} = \begin{bmatrix}-0.5986\\ -0.3380\end{bmatrix}[\,1\,,\,-0.918]\begin{bmatrix}y_{1,t-1}\\ y_{2,t-1}\end{bmatrix} + \begin{bmatrix}1.0058 & -1.0118\\ 0.9921 & -1.0399\end{bmatrix}\begin{bmatrix}\Delta y_{1,t-1}\\ \Delta y_{2,t-1}\end{bmatrix} + \begin{bmatrix}0.4179 & -0.5751\\ 0.4251 & -0.5673\end{bmatrix}\begin{bmatrix}\Delta y_{1,t-2}\\ \Delta y_{2,t-2}\end{bmatrix} + \begin{bmatrix}0.7817 & -0.7723\\ 0.7833 & -0.7762\end{bmatrix}\begin{bmatrix}\Delta y_{1,t-3}\\ \Delta y_{2,t-3}\end{bmatrix} \quad (4.18)$$

4.3 One Factor Model for the Spread Process:

The main fundamental property of the spread process for a good pairs trading strategy is that they are mean-reverting; therefore, Elliott *et al.* (2005) proposed that the spread dynamic follows the Vasicek process. In this model, the spot spread, X_t is the only factor (state variable) and is assumed to follow the Vasicek stochastic process. The Vasicek model is a classical mean-reverting one-factor model which was initially intended to model short interest rates Vasicek (1977) as follows:

$$dX_t = a(\,b - X_t\,)\,dt + \sigma dW_t. \quad (4.19)$$

where the model assumes that the linear combination, X_t evolves as an Ornstein-Uhlenbeck process with constant coefficients, $a > 0$ is mean-reverting speed, b is the

Table 4.2: Unit root tests for WTI and Brent front monthly contracts future price series and their differenced series from April 1994 to January 2005.

Series	ADF test (CV)				ERS test (CV)			
	t-Stat	1pct	5pct	10pct	t-Stat	1pct	5pct	10pct
Brent	-1.1558	-3.46	-2.88	-2.57	22.0692	1.91	3.17	4.33
ΔBrent	-8.2466	-2.58	-1.95	-1.62	0.5905	1.91	3.17	4.33
WTI	-0.7803	-3.46	-2.88	-2.57	28.4303	1.91	3.17	4.33
ΔWTI	-8.4363	-2.58	-1.95	-1.62	0.9181	1.91	3.17	4.33

Table 4.3: The Johansen test results to check whether WTI crude oil and Brent oil prices for the specified data (monthly front contract prices from April 1994 to January 2005) are cointegrated or not.

#Coint.Rel (r)	Johansen test (CV)				Eigenvalues	
	t-Stat	10pct	5pct	1pct	λ_1	λ_2
$r \leq 1$	0.06	6.50	8.18	11.65	0.119	0.0005
$r = 0$	15.86	12.91	14.90	19.19	$\boldsymbol{\beta}' = (\ 1$	, -0.918)

long-run mean of the spread, σ is the volatility of the process, and dW_t is the increment of a standard Brownian motion. Under the risk neutral measure Q, the Vasicek dynamic can be expressed as follows:

$$dX_t = a(\, b^* - X_t\,)\, dt + \sigma dW_t^*. \tag{4.20}$$

where the $b^* = b - \lambda$, λ is the market price of risk and dW_t^* is the equivalent increment of a standard Brownian motion under Q.

By applying Ito's lemma, it is easy to show that the stochastic differential equation (SDE), 4.20 for any $s\,(0 < t \leq s)$ has a solution as follows:

$$X_s \;=\; e^{-a\,(s-t)}\, X_t \;+\; b^*\,(\,1 \;-\; e^{-a\,(s-t)}\,) \;+\; \sigma\, e^{-a\,s} \int_t^s e^{a\,u} \mathrm{d}W_u^*. \tag{4.21}$$

Equation 4.21 shows that the conditional distribution of X_T given X_t is normally distributed with mean and variance :

$$E_Q[\,X_T \mid X_t\,] \;=\; \mu_s = e^{-a\,(T-t)}\, X_t \;+\; b^*\,(\,1 \;-\; e^{-a\,(T-t)}\,), \tag{4.22}$$

$$Var_Q[\,X_T \mid X_t\,] \;=\; \sigma_s^2 = \frac{\sigma^2}{2a}\,(\,1 \;-\; e^{-2a\,(T-t)}\,). \tag{4.23}$$

4.3.1 The Future Spread Pricing:

Since for most commodities, the spot prices are not directly observable, we must instead consider the future spread. Let $DF(t\,,\,T\,,\,X_t)$ to be the future price for the spread with maturity T at time t. we assume that interest rate and market price of risk are constants. As a result, the future (forward) spread is equal to the expected value of the prices spread at time T under the corresponding risk neutral measure. We have:

$$DF(t\,,\,T\,,\,X_t) \;=\; E_Q[\,X_T \mid X_t\,] \;=\; e^{-a\,(T-t)}\,X_t \;+\; b^*\,(\,1 \;-\; e^{-a\,(T-t)}\,), \tag{4.24}$$

Applying Ito's lemma to equation 4.24, we derive:

$$dDF(t\,,\,T\,,\,X_t) \;=\; \frac{\partial DF}{\partial t}\,dt + \frac{\partial DF}{\partial X_t}\,dX_t \;+\; \frac{1}{2}\,\frac{\partial^2 DF}{\partial X^2}\,dX_t^2 \;=\; \sigma\,e^{-a\,(T-t)}\,dW_u^*. \tag{4.25}$$

The result in equation 4.25 proves that in this setting, the future spread process is a martingale and its corresponding volatility decreases exponentially in time to maturity T.

4.3.2 Parameter Estimation:

One of the implications in commodities pricing is that the state variables are mostly unobservable. To estimate latent state variables, the Kalman filter as discussed in chapter 2 is generally deployed. Therefore, we construct the transition and observation equations for this model as follows:

4.3.2.1 Transition Equation:

In this model, the transition equation can be obtained by exact discretization of equation 4.21 in the following form:

$$\boldsymbol{X}_{t+1} = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{X}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t,\;\; t = 1, \ldots, NT, \tag{4.26}$$

where

$$\begin{aligned} \boldsymbol{d}_t \;&=\; b\,(\,1 \;-\; e^{-a\,\Delta t}\,), \quad \boldsymbol{T}_t \;=\; e^{-a\,\Delta t}, \\ \boldsymbol{R}_t \;&=\; \sigma\,\sqrt{\frac{1 \;-\; e^{-2a\,\Delta t}}{2a}}, \end{aligned} \tag{4.27}$$

Here $\boldsymbol{\eta}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\eta}_t) \;=\; 0$ and $Var(\boldsymbol{\eta}_t) \;=\; 1$.

4.3.2.2 Measurement Equation:

There are several reasons why the future spread prices in commodity market can be noisy including bid-ask spreads, differing maturity dates for underlying commodities, and observing the underlying prices non-simultaneously. Therefore, the measurement equation can be derived using equation, 4.24 plus Gaussian white noise as follow:

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, NT, \tag{4.28}$$

where

$$\begin{aligned} \boldsymbol{y}_t \;\; & is \;\; an \;\; N \times 1 \text{ vector of observations, and } \boldsymbol{c}_t \,\&\, \boldsymbol{Z}_t \; are \, N \times 1 \text{ vectors,} \\ \boldsymbol{y}_t \;\; &= \left[DF(t\,,\, T_i\,) \right], \qquad\qquad i = 1, \ldots, N, \\ \boldsymbol{c}_t \;\; &= \left[b^* \left(1 \,-\, e^{-a\,(T_i - t)} \right) \right], \quad i = 1, \ldots, N, \\ \boldsymbol{Z}_t \;\; &= \left[e^{-a\,(T_i - t)} \right], \qquad\qquad i = 1, \ldots, N, \end{aligned} \tag{4.29}$$

Here $\boldsymbol{\varepsilon}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\varepsilon}_t) = 0$ and $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{H}$.

Elliott *et al.* (2005) pointed out that the method has advantages including capturing mean-reversion, the fundamental property of pairs trading strategy, considering time continuous process for the spread process and tractability of the method (the parameters can be simply estimated using the Kalman filter algorithm). However, there are some convincing reasons why this method cannot be applied to most commodities spread processes. First, there is now enough evidence to support the statement that the behavior of most commodities spread process cannot be summarized in one factor. Secondly, the underlying commodities (future) prices in the spread process rapidly respond to eventful market news. These events sometimes highly impact on one of the underlying prices and have less impact on the other one causing the volatility of the spread process increases.

4.3.3 Empirical Results

The same data sample for WTI and Brent crude oils that is used in section 4.2.2 is considered. Moreover, monthly future location spread prices (the front contract) between WTI crude oil and Brent oil are used from April 1994 to January 2005. We attempt to fit the model for location spreads between WTI and Brent oils. We assume the location spread to be difference between WTI crude oil price (C) and Brent oil price (B). We should mention that the cointegration coefficient between WTI and Brent is $\beta = 0.918$ in the study data set and $C - \beta B$ is natural choice; however, since the location spread is such a common trade that even options are written on the location

spreads and traded in practice in various exchange including CME group, we consider to fit the model for location spreads. As in previous example 4.2.2, we use monthly data to estimate the model's parameters. The monthly futures location spread are calculated with five different maturities from April 1994 to January 2005. The involved future contracts are the 1,3,6,9 and 12 months to maturities. Figure 4.5 depicts the monthly front contract future prices' location spread between WTI and Brent oils. Table 4.4 lists the estimation results for Elliott's one factor model. Figure 4.6 depicts how estimates of parameters evolve in the model maximum likelihood optimization function for our sample data. Also, figure 4.7 compares the filtered spot spread with the front contract realizations which usually assumed as spot spread.

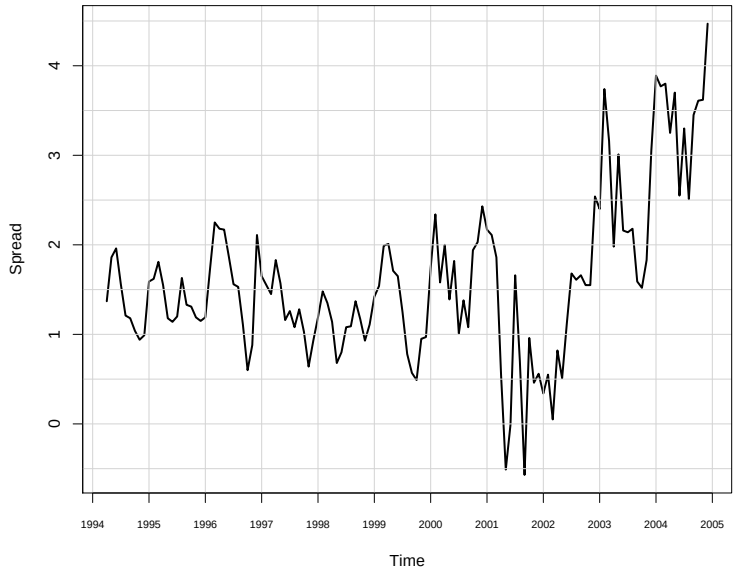

Figure 4.5: The monthly front contract location spread between WTI crude oil and brent oil from April 1994 to January 2005 (spread in $/barrel).

4.4 Two-Factor model for the Spread Process:

The uncertainly on the commodities spread models is high and requires more than one factor to have enough flexibility. Dempster *et al.* (2008) introduced a two factor model to price the spread options in energy markets. In this model, the spot spread, X_t and deviation from the long-run mean of the spot spread, δ_t are two unobserved state variables.

Table 4.4: Estimated parameters for Elliott's one factor model using monthly location spread between WTI and Brent oils for future contracts of 1,3,6,9 and 12 months to maturities from April 1994 to January 2005.

	a	σ	b	λ	ζ
value	0.954	1.499	2.057	0.971	0.233
Sd.Err	0.062	0.129	0.482	0.486	0.056

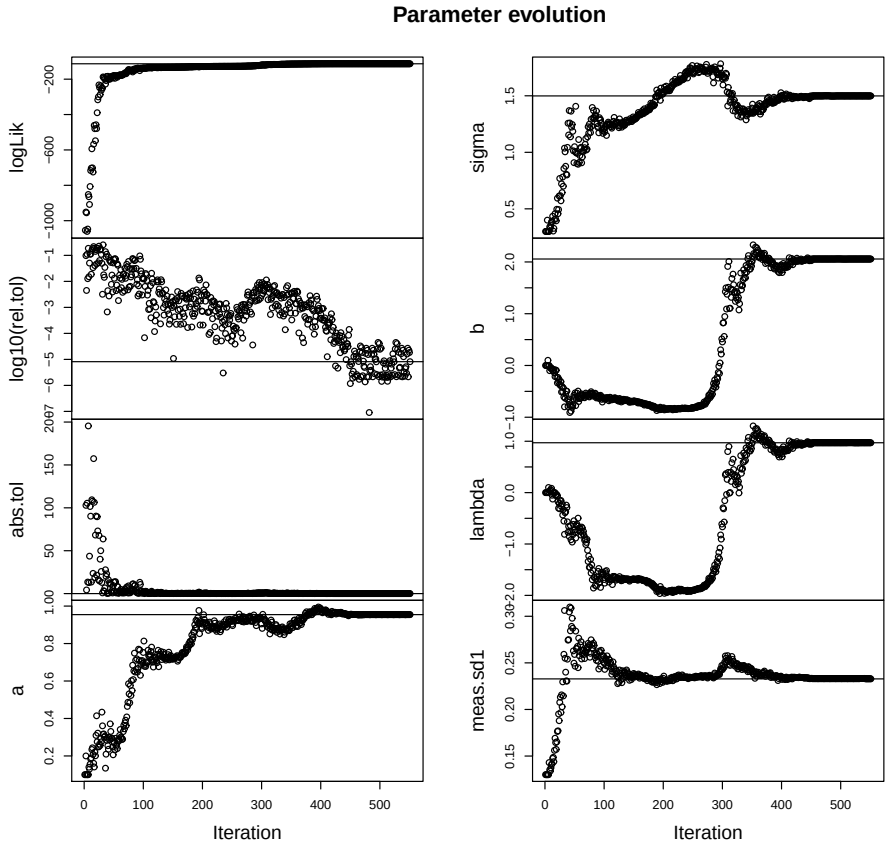

Figure 4.6: Elliot's model parameters estimation showing how parameters evolve in optimization's function generated by Kalman filter. As we can see, after some iterations, all parameters along with maximum likelihood function converge and stabilize. The data set is monthly location spread between WTI and Brent oils for future contracts of 1,3,6,9 and 12 months to maturities from April 1994 to January 2005.

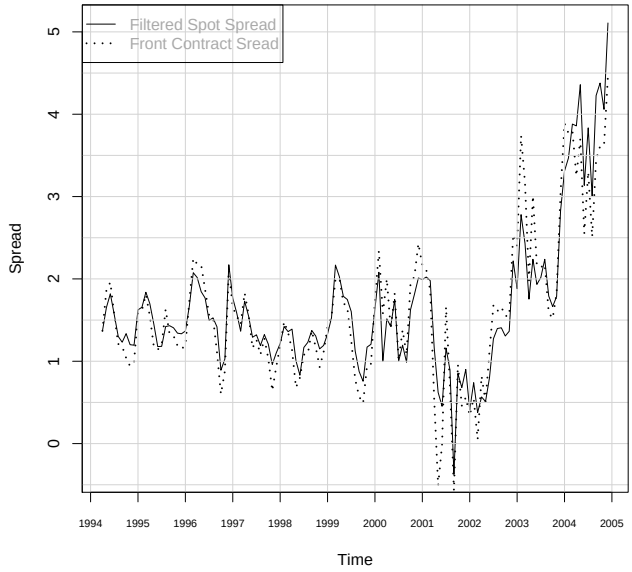

Figure 4.7: Compare Elliot's model filtered spot spread series with the front contract monthly realization. The data set that is used to estimate filter spot spread is monthly location spread between WTI and Brent oils for future contracts of 1,3,6,9 and 12 months to maturities from April 1994 to January 2005

Both factors are considered to follow mean-reverting Ornstein-Uhlenbeck processes and so the joint stochastic process is:

$$dX_t = \kappa_1(\alpha + \varphi(t) + \delta_t - X_t)\,dt + \sigma_1\,dW_{1t}, \tag{4.30}$$

$$d\delta_t = -\kappa_2\,\delta_t\,dt + \sigma_2\,dW_{2t}. \tag{4.31}$$

where the standard Brownian motion increments, dW_{1t} and dW_{2t}, are correlated with:

$$dW_{1t}\,dW_{2t} = \rho\,dt, \tag{4.32}$$

α and 0 are the long-run means of the two factors, X_t and δ_t respectively,
$\kappa_1 > 0$ and $\kappa_2 > 0$ are the mean reverting rates (speeds),
σ_1 and σ_2 are the volatility of the processes,
$\varphi(t)$ is the seasonality component of the spread process.

The equation 4.30 shows that the spot process X_t is a mean-reverting process with a stochastic long-term mean process, $\alpha + \delta_t$. The seasonality component, $\varphi(t)$ is inherited from the underlying commodities prices. For instance, the crack spread between gasoline

and WTI crude oil shows clear evidence of seasonality and is inherited from the gasoline price. In this model, the applied Fourier-series type seasonality function $\varphi(t)$ is used. The seasonality function $\varphi(t)$ was proposed by Durbin and Koopman (1997) and defined as:

$$\varphi(t) \;=\; \sum_{i=0}^{k} \{a_i \cos(2\pi i t) \;+\; b_i \sin(2\pi i t)\}. \tag{4.33}$$

where a_i and b_i are constant. The one-factor model 4.3 and cointegration method such as $VECM$ assume that long-term mean relationship between the underlying commodities prices is constant. On the other hand, Dempster *et al.* (2008) correctly claims that several reasons including high frequency of events, consumers' behaviorial change and economical crises require this long-term relationship not to be constant. As a result, the deviation from the long-run mean δ_t was considered to follow another mean-reverting stochastic process with enough flexibility to confront the reality of the commodities spread model. The joint dynamics of the model under a risk-neutral measure Q can be written as:

$$\begin{aligned} dX_t &= \kappa_1 (\alpha - \lambda_1 + \varphi(t) + \delta_t - X_t)\, dt + \sigma_1\, dW^*_{1t}, && (4.34)\\ d\delta_t &= \kappa_2 (-\lambda_2 - \delta_t)\, dt + \sigma_2\, dW^*_{2t}, && (4.35)\\ & dW^*_{1t}\, dW^*_{2t} = \rho\, dt. \end{aligned}$$

where λ_1, λ_2 are the market price of risks (assumed constants) for the state variables X_t, δ_t respectively and dW^*_{1t} and dW^*_{2t} are the standard Brownian motion increments under the equivalent risk-neutral measure Q.
By applying Ito's lemma,we derive the solution for the joint stochastic differential equation (SDE) 4.34, 4.35 for any $s\,(0 < t \leq s)$ as follows:

$$\begin{aligned} X_s &= e^{-\kappa_1 (s-t)} X_t + (\alpha - \lambda_1)(1 - e^{-\kappa_1 (s-t)}) + \kappa_1 e^{-\kappa_1 s} \int_t^s \delta_u e^{\kappa_1 u} du \\ &\quad + \kappa_1 e^{-\kappa_1 s} \int_t^s \varphi(u) e^{\kappa_1 u} du + \sigma_1 e^{-\kappa_1 s} \int_t^s e^{\kappa_1 u} \mathrm{d}W^*_{1u}, && (4.36)\\ \delta_s &= e^{-\kappa_2 (s-t)} \delta_t - \lambda_2 (1 - e^{-\kappa_2 (s-t)}) + \sigma_2 e^{-\kappa_2 s} \int_t^s e^{\kappa_2 v} \mathrm{d}W^*_{2v}. && (4.37) \end{aligned}$$

We plug 4.37 into 4.36 to derive:

$$\begin{aligned} X_s &= e^{-\kappa_1(s-t)}\,X_t + (\,\alpha - \lambda_1\,)\,(\,1 - e^{-\kappa_1(s-t)}\,) \\ &\quad + \frac{\kappa_1\delta_t}{(\kappa_1-\kappa_2)}\left\{\,e^{-\kappa_2(s-t)} - e^{-\kappa_1(s-t)}\right\} - \lambda_2\,(\,1 - e^{-\kappa_1(s-t)}\,) \\ &\quad + \frac{\kappa_1\lambda_2}{(\kappa_1-\kappa_2)}\left\{\,e^{-\kappa_2(s-t)} - e^{-\kappa_1(s-t)}\right\} + G(t,s) \\ &\quad + \frac{\kappa_1\sigma_2}{(\kappa_1-\kappa_2)}\int_t^s \left\{\,e^{-\kappa_2(s-v)} - e^{-\kappa_1(s-v)}\right\} \mathrm{d}W^*_{2v} \\ &\quad + \sigma_1\,e^{-\kappa_1\,s}\int_t^s e^{\kappa_1\,u}\mathrm{d}W^*_{1u}, \end{aligned} \tag{4.38}$$

$$\delta_s = e^{-\kappa_2\,(s-t)}\,\delta_t - \lambda_2\,(\,1 - e^{-\kappa_2\,(s-t)}\,) + \sigma_2\,e^{-\kappa_2\,s}\int_t^s e^{\kappa_2\,v}\mathrm{d}W^*_{2v}. \tag{4.39}$$

where $G(t,s)$ is the seasonality function and assumed to be:

$$G(t,s) = \kappa_1\,e^{-\kappa_1\,s}\int_t^s \varphi(u)\,e^{\kappa_1 u}du. \tag{4.40}$$

Equations 4.38 and 4.39 show that the conditional distribution of X_T given X_t and δ_t is normal and has mean and variance (using Ito's isometry in stochastic integrals) as follows:

$$\begin{aligned} E_Q[\,X_T \mid X_t,\delta_t\,] &= \mu_X = e^{-\kappa_1\,(T-t)}\,X_t + (\,\alpha - \lambda_1\,)\,(\,1 - e^{-\kappa_1(T-t)}\,) \\ &\quad + \frac{\kappa_1\delta_t}{(\kappa_1-\kappa_2)}\left\{e^{-\kappa_2(T-t)} - e^{-\kappa_1(T-t)}\right\} - \lambda_2\,(1 - e^{-\kappa_1(T-t)}) \\ &\quad + \frac{\kappa_1\lambda_2}{(\kappa_1-\kappa_2)}\left\{\,e^{-\kappa_2(T-t)} - e^{-\kappa_1(T-t)}\right\} + G(t,T), \end{aligned} \tag{4.41}$$

$$Var_Q[\,X_T \mid X_t,\delta_t\,] = \sigma_X^2 = A_1 + A_2 + 2\rho A_3. \tag{4.42}$$

where

$$
\begin{aligned}
A_1 &= \frac{\sigma_1^2}{2\kappa_1}(1 - e^{-2\kappa_1 (T-t)}),\\
A_2 &= \frac{\kappa_1^2\sigma_2^2}{(\kappa_1-\kappa_2)^2}\left\{\frac{1}{2\kappa_1}(1 - e^{-2\kappa_1 (T-t)}) + \frac{1}{2\kappa_2}(1 - e^{-2\kappa_2 (T-t)})\right.\\
&\quad \left. -\frac{2}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)(T-t)})\right\},\\
A_3 &= \frac{\kappa_1\sigma_1\sigma_2}{(\kappa_1-\kappa_2)}\left\{\frac{1}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)(T-t)})\right.\\
&\quad \left. -\frac{1}{2\kappa_1}(1 - e^{-2\kappa_1 (T-t)})\right\}.
\end{aligned}
$$

In equation 4.42, if we take the limit as $T \to \infty$, we will find constant variance:

$$
\lim_{T\to\infty} Var_Q[\, X_T \mid X_t, \delta_t\,] = \frac{\sigma_1^2}{2\kappa_1} + \frac{\kappa_1\sigma_2^2}{2\kappa_2(\kappa_1+\kappa_2)} + \frac{\sigma_1\sigma_2\rho}{\kappa_1+\kappa_2}. \tag{4.43}
$$

4.4.1 The Futures Price of the Spot Spread:

As we explained in section 4.3.1, under the risk neutral measure Q the futures price for the spot spread with maturity T at time t must be equal to $DF(t\,,T\,,X_t) = E_Q[\,X_T \mid X_t, \delta_t\,]$, so we have:

$$
\begin{aligned}
DF(t\,,T\,,X_t) &= E_Q[\,X_T \mid X_t, \delta_t\,] = e^{-\kappa_1 (T-t)}\, X_t + (\,\alpha - \lambda_1\,)\,(1 - e^{-\kappa_1 (T-t)}\,)\\
&\quad + \frac{\kappa_1\delta_t}{(\kappa_1-\kappa_2)}\left\{e^{-\kappa_2 (T-t)} - e^{-\kappa_1 (T-t)}\right\} - \lambda_2\,(1 - e^{-\kappa_1 (T-t)})\\
&\quad + \frac{\kappa_1\lambda_2}{(\kappa_1-\kappa_2)}\left\{e^{-\kappa_2 (T-t)} - e^{-\kappa_1 (T-t)}\right\} + G(t,T).
\end{aligned} \tag{4.44}
$$

Applying Ito's lemma to equation 4.44, in the risk neutral measure Q, shows that $dDF(t\,,\,T\,,\,X_t)$ satisfies:

$$\begin{aligned} dDF(t\,,\,T\,,\,X_t) &= \frac{\partial DF}{\partial t}\,dt + \frac{\partial DF}{\partial X}\,dX_t + \frac{\partial DF}{\partial \delta}\,d\delta_t \\ &\quad + \frac{1}{2}\,\frac{\partial^2 DF}{\partial X^2}\,dX_t^2 + \frac{\partial^2 DF}{\partial X \partial \delta}\,dX_t\,d\delta_t + \frac{1}{2}\,\frac{\partial^2 DF}{\partial \delta^2}\,d\delta_t^2 \\ &= \frac{\kappa_1 \sigma_2}{(\kappa_1 - \kappa_2)}\left\{ e^{-\kappa_2(T-t)} - e^{-\kappa_1(T-t)} \right\} \mathrm{d}W_{2t}^* \\ &\quad + \sigma_1 e^{-\kappa_1(T-t)}\,\mathrm{d}W_{1t}^*. \end{aligned} \tag{4.45}$$

The result in equation 4.45 shows that in this model, the future spread process is a Q martingale.

4.4.2 Parameter Estimation:

Since the state variables spot spreads, X_t and deviation from long-term mean, δ_t are unobserved, to calibrate this two factor model, we apply the Kalman filter algorithm. The transition and observation equations for this model can, therefore, be constructed as follows:

4.4.2.1 Transition Equation:

We obtain the transition equation using exact discretization of equations 4.38 and 4.39 in the following format:

$$[\boldsymbol{X}_{t+1}, \boldsymbol{\delta}_{t+1}]' = \boldsymbol{d}_t + \boldsymbol{T}_t\,[\boldsymbol{X}_t, \boldsymbol{\delta}_t]' + \boldsymbol{\eta}_t, \quad t = 1, \ldots, NT, \tag{4.46}$$

where

$$\begin{aligned} \boldsymbol{d}_t &= \begin{bmatrix} \alpha\,(\,1 - e^{-\kappa_1 \Delta t}\,) + G(t, t+\Delta t) \\ 0 \end{bmatrix}, \\ \boldsymbol{T}_t &= \begin{bmatrix} e^{-\kappa_1 \Delta t} & \frac{\kappa_1}{(\kappa_1 - \kappa_2)}\left\{ e^{-\kappa_2 \Delta t} - e^{-\kappa_1 \Delta t} \right\} \\ 0 & e^{-\kappa_2 \Delta t} \end{bmatrix}, \end{aligned} \tag{4.47}$$

and $\boldsymbol{\eta}_t$ is serially uncorrelated Gaussian white noise with:

$$E(\boldsymbol{\eta}_t) = \mathbf{0} \quad \text{and} \quad Var(\boldsymbol{\eta}_t) = \begin{bmatrix} \sigma_X^2 & \sigma_{X\delta} \\ \sigma_{X\delta} & \sigma_\delta^2 \end{bmatrix}. \tag{4.48}$$

where

$$
\begin{aligned}
\sigma_X^2 &= \frac{\sigma_1^2}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t}) \\
&\quad + \frac{\kappa_1^2\sigma_2^2}{(\kappa_1-\kappa_2)^2}\left\{\frac{1}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t}) + \frac{1}{2\kappa_2}(1 - e^{-2\kappa_2 \Delta t})\right. \\
&\quad \left. - \frac{2}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)\Delta t})\right\} \\
&\quad + \frac{2\rho\kappa_1\sigma_1\sigma_2}{(\kappa_1-\kappa_2)}\left\{\frac{1}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)\Delta t})\right. \\
&\quad \left. - \frac{1}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t})\right\}, \\
\sigma_\delta^2 &= \frac{\sigma_2^2}{2\kappa_2}(1 - e^{-2\kappa_2 \Delta t}), \\
\sigma_{X\delta} &= \frac{\kappa_1\sigma_2}{(\kappa_1-\kappa_2)}\left\{-\frac{1}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)\Delta t})\right. \\
&\quad \left. + \frac{1}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t})\right\} + \frac{\rho}{(\kappa_1+\kappa_2)}(1 - e^{-(\kappa_1+\kappa_2)\Delta t}).
\end{aligned}
\tag{4.49}
$$

Dempster *et al.* (2008) shows that the correlation ρ between short and long end fluctuations approaches zero; therefore, ρ is assumed to be zero in the model estimation.

4.4.2.2 Measurement Equation:

In section 4.3.2.2, we argue that the future spread prices in commodity market are noisy. Therefore, we derive the measurement equation using equation 4.44 and add Gaussian white noise as follows:

$$
\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t \left[\boldsymbol{X}_t, \boldsymbol{\delta}_t\right]' + \boldsymbol{\varepsilon}_t \quad t = 1, \ldots, NT \tag{4.50}
$$

where

$$\begin{aligned}
\boldsymbol{y}_t &\text{ is an } N\times 1 \text{ vector of observations, } \boldsymbol{c}_t \text{ is } N\times 1 \text{ vector and } \boldsymbol{Z}_t \text{ is } N\times 2 \text{ matrix,}\\
\boldsymbol{y}_t &= \left[\, DF(t\,,\,T_i\,)\,\right], \quad i=1,\ldots,N,\\
\boldsymbol{c}_t &= \Big[\,(\,\alpha\,-\lambda_1\,)\,(\,1\,-\,e^{-\kappa_1(T_i-t)}\,)\,-\,\lambda_2\,(1-e^{-\kappa_1(T_i-t)})\\
&\quad +\frac{\kappa_1\lambda_2}{(\kappa_1-\kappa_2)}\left\{e^{-\kappa_2(T_i-t)}\,-\,e^{-\kappa_1(T_i-t)}\right\}+\,G(t,T_i)\Big], \quad i=1,\ldots,N,\\
\boldsymbol{Z}_t &= \left[e^{-\kappa_1\,(T_i-t)}\,,\ \frac{\kappa_1\left\{e^{-\kappa_2(T_i-t)}-e^{-\kappa_1(T_i-t)}\right\}}{(\kappa_1-\kappa_2)}\right].\ \ i=1,\ldots,N. \qquad (4.51)
\end{aligned}$$

In these expressions, $\boldsymbol{\varepsilon}_t$ is serially uncorrelated Gaussian white noise with $E(\boldsymbol{\varepsilon}_t) = 0$ and $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{H}$. Since the number of parameters to estimate is quite large, we consider the variance of the future spreads disturbances with all maturities to be identical which means $\boldsymbol{H} = \zeta^2\boldsymbol{I}_N$.

4.4.3 Empirical Results

The same data set that we describe in section 4.3.3 was used to compare the results with other models explained in this chapter. Here we apply Dempster *et al.* (2008)'s model to fit to our sample set to location spread between WTI and Brent crude oils. Since there is not enough evidence to support seasonality in the WTI and Brent oils' location spread, the seasonality component is considered to be zero ($\varphi(t) = 0$). Table 4.5 summarizes the calibration results for Dempster's two factor model and comparison with Dempster's estimated results. As we see in table 4.5, the results are slightly different which arises from using slightly different data samples. Figure 4.9 depicts how estimates of parameters evolve in the model's maximum likelihood optimization function for our sample data. Also figure 4.8 compares the filtered spot spread with the front contract realizations which usually assumed as approximation for spot spread.

4.5 Analysis of Recent Behavioral Change in Spread between WTI and Brent Crude oils:

In recent years, the behavior of the Brent and WTI crude oils' spread has experienced an unprecedented change. Historically, WTI crude oil has mostly traded at higher price

Table 4.5: Estimated parameters for Dempster's two factor model using monthly future location spread between WTI and Brent oils from April 1994 to January 2005 and comparison with Dempster's estimated parameters.

Estimated by		K_1	α	σ_1	λ_1	K_2	σ_2	λ_2	ζ
Moosavi	value	4.86	0.00	3.736	0.195	0.137	0.965	0.482	0.104
	Sd.Err	0.265	0.00	0.313	1.19	0.019	0.070	0.298	0.0285
Dempster	value	5.04	0.00	3.93	0.50	0.14	1.45	0.21	0.12
	Sd.Err	0.28	0.00	0.28	1.18	0.02	0.10	0.28	0.03

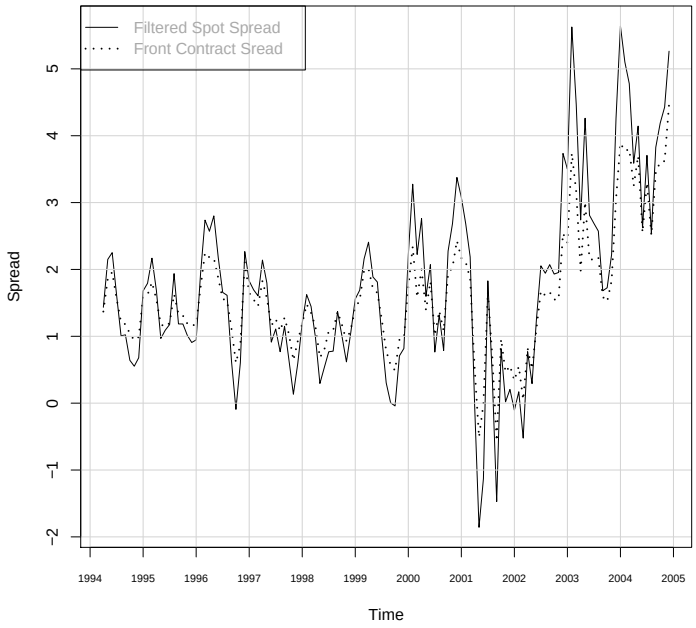

Figure 4.8: Compare Dempster's model filtered spot spread series with the front contract monthly realization. The data set that is used to estimate filter spot spread is monthly location spread between WTI and Brent oils for future contracts of 1,3,6,9 and 12 months to maturities from April 1994 to January 2005.

over Brent oil. Recently[1], Brent oil has traded at dramatically higher price than the WTI crude oil. Figure 4.10 illustrates the daily spread process between WTI crude oil and Brent oil from April 1994 to November 2013 which clearly depicts a significant

[1]Early 2011.

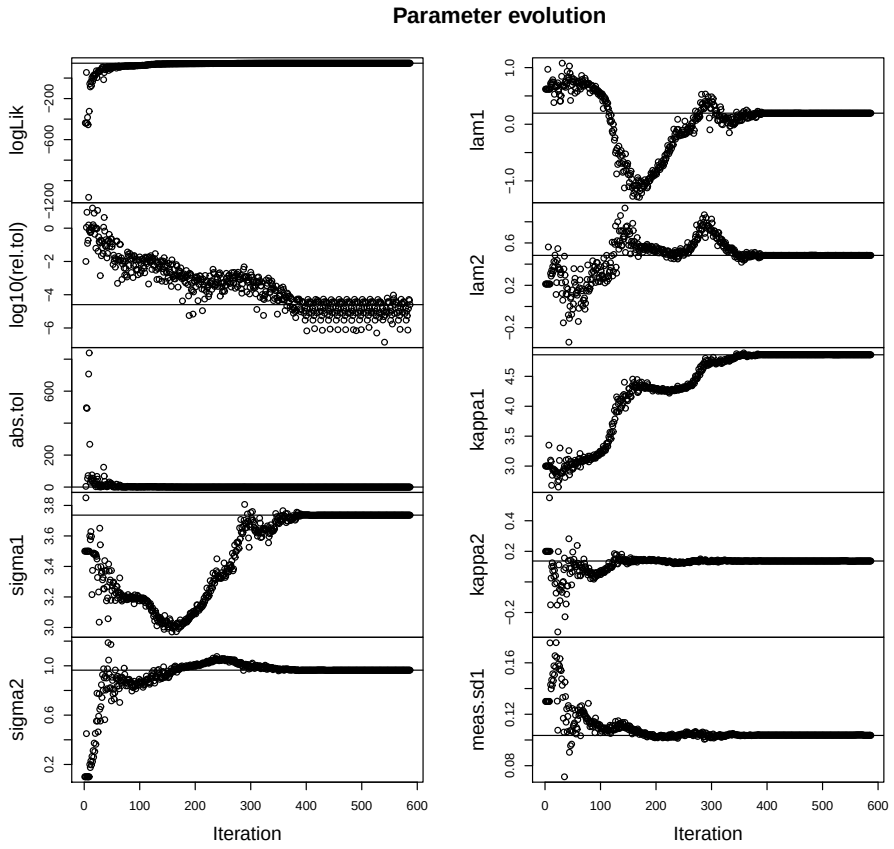

Figure 4.9: Dempster's model parameters estimation showing how estimates of parameters evolve in optimization's function for Kalman filter. As we can see, after some iterations, all parameters along with maximum likelihood function converge and stabilize. The data set is monthly location spread between WTI and Brent oils for future contracts of 1,3,6,9 and 12 months to maturities from April 1994 to January 2005.

behavioral change (reversal) starting from early 2011. Although WTI and Brent are two major global crude oil benchmarks, the WTI price describes the price for most U.S. oil consumers pay and the Brent price mostly represents the price for European and international consumption. In spite of the fact that the quality of WTI crude oil is slightly higher than that of Brent oil and theoretically the WTI should be more or less the same, the spread between WTI and Brent oils had greatly differed. The most common reasons for this significant widening in the spread are: the unexpected recent hike in production in the United States has built up the inventories at Cushing, Oklahoma where the WTI crude oil is delivered and priced, and has logistical; economic constraints

(Borenstein and Kellogg (2012)). This excessive supply has caused imbalance in supply and demand, leading to a lower equilibrium price for WTI crude oil compare to the Brent oil at the hub. Furthermore, the tension in oil producing countries, particularly Middle East countries and Venezuela mainly affects the Brent oil rather than WTI crude oil. It worth mentioning that when the WTI has traded lower then the Brent, the U.S. oil producers earned relatively less money compare to the other international producers. With the advent of contemporary technologies and improvement in oil's extraction and process, the United States and Canada are expecting booms in oil production. This resulting increased production will make them self-sufficient and they will no longer need to import crude oil in the near future (Ingrid (2013)). These kind of events are imposing a significant level of volatility in spread between WTI and Brent oils. We attempt to

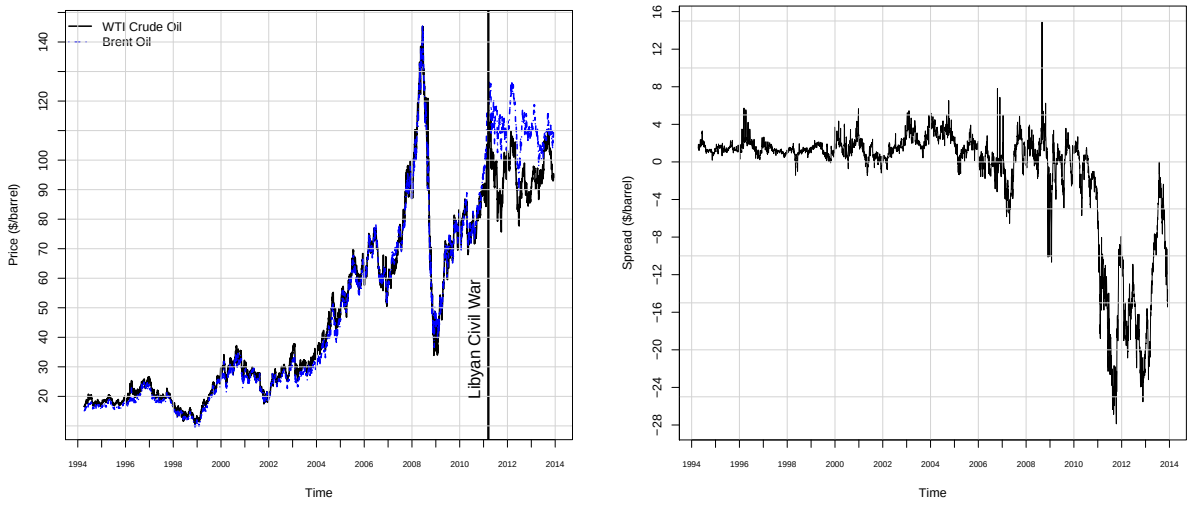

Figure 4.10: The daily front contract of WTI crude oil prices and Brent crude oil prices and their location spread series from April 1994 to November 2013 (price in $/barrel).

explore the validity of all three approaches, which we explained and implemented for monthly spread from April 1994 to January 2005 in this chapter to the recent daily spread from April 1994 to November 2013. First, we applied the *Johansen test (JT)* to check whether WTI crude oil and Brent oil for our daily data are cointegrated or not. The test results are summarized in table 4.6. As we expected that there is not enough evidence to reject the null hypothesis, that they are not cointegrated, using the *JT* with maximal eigenvalue statistic at any percent level of significance. Therefore, the daily WTI crude oil series and Brent oil series are no longer cointegrated. Although the cointegration test fail to support the comovement and cointegration between the WTI and Brent oils, still many industry members and economists believe that the spread

will inevitably approach parity in the near future. We can also argue that the spread series has some merits of mean-reverting which logically, economically and fundamentally make sense. Elliott's and Dempster's models can still be deployed to model the spread series; however, these models are not flexible enough to capture all recent behavioral changes in the daily spread process. As we can see in figure 4.10, the volatility in the spread process is significantly high, and is eventful. Sometimes, events have very different impact on WTI crude oil vs Brent oil. For instance, the Libyan Civil war in 2011 had higher impact on Brent oil than it did on WTI crude oil, which makes sense since Brent is European vs North American WTI (see figure 4.10). These shocks may persist for quite long time, so Elliott's and Dempster's models cannot fully capture all these type of empirical phenomena. All of this empirical evidence has economic and physical interpretation. To capture all empirical phenomena in the spread process, we extend both Elliott's and Dempster's models by adding a *compound Poisson process where jump sizes follow a double exponential distribution*. Intuitively, the jump component of the models may capture the market reaction to the highly important outside events. In other words, when there is no outstanding news, the spread process follow the normal Elliott's and Dempster's models and when the highly influential news arrives based on a Poisson process, the spread process changes based on jump size distribution. In the following sections, we will introduce this new method in detail.

Table 4.6: Johansen test results to check whether WTI crude oil , Brent oil for the specified daily data are cointegrated or not.

#Coint.Rel (r)	Johansen test(CV)			
	t-Stat	10pct	5pct	1pct
$r \leq 1$	5.32	6.50	8.18	11.65
$r = 0$	12.49	12.91	14.90	19.19

4.6 One-factor Mean-reverting Model with Jump

The Merton (1976) jump diffusion (MJD) model represents one attempt to modify the Black-Scholes model. The model tries to capture the skewness and excess kurtosis of the log-returns density of risky financial assets. Merton (1976) showed that the change in a risky asset price has two components: the *"normal"* and *"abnormal"*. The normal variations in the price explain the small daily changes caused by some regular characteristics of market such as supply and demand, and the abnormal variations in the price explain the sudden shocks (jumps). In MJD model, normal continuous changes in

the log-returns is expressed by a Brownian motion with drift process similar to the Black-Scholes model, and the discontinuous jumps are expressed by a compound Poisson process with jump sizes that follow a normal distribution. Later, Kou (2002) extend the MJD method by employing double exponential distribution for jump sizes. Here, we extend the one-factor spread model that we explained in section 4.3 by adding a compound Poisson process where jump sizes follow double exponential distribution. Under the real measure P, the spread process X_t evolves as an Ornstein-Uhlenbeck process plus jump as follows:

$$dX_t = a(\, b - X_t \,)\, dt + \sigma dW_t \, + \, Y_t dN_t. \tag{4.52}$$

where $a > 0$ is mean-reverting speed, b is the long-run mean of the spread, σ is the volatility of the process, dW_t is the increment of a standard Brownian motion, N_t is the Poisson process governing the jumps of X_t and has a jump intensity parameter λ (the average number of jumps per year; therefore, the probability of a jump in time Δt is $\lambda \Delta t$). The jump size Y_t follows an asymmetric double exponential distribution. The double exponential distribution is given as:

$$f_Y(y) = \begin{cases} p\eta_1 e^{-\eta_1 y}, & \text{if } y \geq 0 \\ q\eta_2 e^{\eta_2 y}, & \text{if } y < 0 \end{cases} \tag{4.53}$$

where the upward and downward jumps follow the two independent exponential distributions with rates $\eta_1\,,\, \eta_2 > 0$ respectively, and $p, q \geq 0\,,\, p + q = 1$ are assumed to be the probabilities of upside and downside spikes respectively. Notice that jumps are a sequence of independent identically distributed (i.i.d.) random variables. In the model, all sources of randomness W_t,N_t, and Y_t are assumed to be independent. It is easy to show that:

$$E(Y) = \frac{p}{\eta_1} \, - \, \frac{q}{\eta_2}, \quad Var(Y) = pq(\frac{1}{\eta_1} \, + \, \frac{1}{\eta_2})^2 + (\frac{p}{\eta_1^2} \, + \, \frac{q}{\eta_2^2}). \tag{4.54}$$

In the compound Poisson process, in a short time interval, Δt, either there is only one jump or there is not a jump. The probability of one jump in the time interval, Δt is:

$$P\,(N_{t+\Delta t} - N_t = 1) = \lambda \Delta t e^{-\lambda \Delta t} \simeq \lambda \Delta t.$$

Therefore, the Euler discretization of the model will be:

$$X_{t+\Delta t} \, = \begin{cases} X_t \, + \, a\,(\, b - X_t\,)\,\Delta t \, + \, \sigma\,\sqrt{\Delta t}\,N(0,1), & \text{with probability } 1 - \lambda \Delta t, \\ X_t \, + \, a\,(\, b - X_t\,)\,\Delta t \, + \, \sigma\,\sqrt{\Delta t}\,N(0,1) + Y_t, & \text{with probability } \lambda \Delta t. \end{cases} \tag{4.55}$$

4.7 Two-Factor Model with Jumps for the Spread Process:

As we mentioned in section 4.4, the uncertainly on the commodity spread models is very high and may require more than one factor to have enough flexibility. The commodities spread process is also quite eventful. When highly important news arrives in the markets, underlying commodities may be affected in remarkably different ways (for instance, in the WTI-Brent spread process, some events may have higher impact on WTI crude oil rather than Brent oil and vice versa). The rate and magnitude of these types of events also different causing jumps upwards and downwards with different rates and size in the spread process. Therefore, we are motivated to extend the Dempster *et al.* (2008) two factor model by adding a jump component where jump sizes follow an asymmetric double exponential distribution to price the spread options in energy markets. In this model, the spot spread process X_t and the deviation from the long-run mean of the spot spread Y_t are two unobserved state variables. Both factors are considered to follow mean-reverting Ornstein-Uhlenbeck processes and the joint stochastic process plus jump is:

$$dX_t = \kappa_1(\alpha + \varphi(t) + \delta_t - X_t)\,dt + \sigma_1\,dW_{1t} + Y_t dN_t, \tag{4.56}$$

$$d\delta_t = -\kappa_2\,\delta_t\,dt + \sigma_2\,dW_{2t}, \tag{4.57}$$

where:

The standard Brownian motion increments, dW_{1t} and dW_{2t} are correlated with:

$$dW_{1t}\,dW_{2t} = \rho\,dt, \tag{4.58}$$

α and 0 are the long-run means of the two factors, X_t and δ_t respectively,
$\kappa_1 > 0$ and $\kappa_2 > 0$ are the mean reverting rates (speeds),
σ_1 and σ_2 are the volatility of the processes,
$\varphi(t)$ is the seasonality component of the spread process,
N_t is the Poisson process governing the jumps of X_t and has a jump intensity parameter λ (the average number of jumps per year),
The jump size Y_t follows the asymmetric double exponential distribution given in equation 4.53,
$\varphi(t)$ is the seasonality component, given in equation 4.33 and is inherited from the underlying commodities prices.

Notes that jumps are a sequence of independent identically distributed (i.i.d.) random variables. Except W_{1t} and W_{2t}, all other sources of randomness including W_{1t}, W_{2t}, N_t, and Y_t are assumed to be independent.

4.8 The Generalized One-factor Mean-reverting Model

Figure 4.10 confirms that the behavior of the Brent and WTI crude oils' spread has recently experienced an unprecedented change. Both Elliott *et al.* (2005)'s one factor mean-reverting model, reviewed in section 4.3, and the two factor model proposed by Dempster *et al.* (2008) and reviewed in section 4.4 to model the spot spread process have failed to capture these structural changes in the spread process between the Brent and WTI crude oils. The skewness, heavy tails, kurtosis and volatility variant are some fundamental unquestionable factors of the spread process that cannot be captured by the above mentioned models. Here, we attempt to capture all these crucial properties in a single factor model. In this generalized one-factor mean-reverting model (GOFMRM), the only factor is the spot spread X_t and its dynamics is considered to follow:

$$dX_t = \kappa\,(\mu\ -\ X_t)\ dt + \sigma\,\sqrt{\nu^2(X_t-\theta)^2+1}\,dW_t, \tag{4.59}$$

where

μ is the long-run mean,
$\kappa \geq 0$ is the mean reversion rate,
$\sigma > 0$, $\nu \geq 0$ are the volatility parameters of the process and are considered constant,
θ is the skewness parameter and constant,
dW_t is the increment of a standard Brownian motion.

As we can see the volatility is no longer constant. The total volatility can be decomposed into two parts: The first part σ is constant and does not depend on the spot spread X_t location and the second part ν is imposed based on the location of the spread from the long-run mean μ with different weights (forces) in two sides based on θ. In other words, when $\mu = \theta$, the process is symmetric, when $\theta < \mu$, the process is skewed to the right and when $\theta > \mu$, the process is skewed to the left. Moreover, when ν increases the process will have heavier tail or tails and higher peak (kurtosis). It is worth mentioning that when $\nu = 0$, the model reduces to the Vasicek model. Figure 4.11 illustrates these claims based on the simulation results using different parameter settings.

Let $y = \varphi(x) = \int \frac{dx}{\sqrt{\nu^2(x-\theta)^2+1}} = \frac{\sinh^{-1}(\nu(x-\theta))}{\nu}$ and applying Itô's lemma, we derive:

$$dY_t = \left\{\kappa(\mu-\theta)\mathrm{sech}(\nu Y_t)\ -\ \left(\frac{\kappa}{\nu}+\sigma^2\nu\right)\tanh(\nu Y_t)\right\}\,dt + \sigma\,dW_t, \tag{4.60}$$

If $\theta = \mu$, the SDE will be symmetric and will be simplified in the form as follows:

$$dY_t = -\left(\frac{\kappa}{\nu}+\sigma^2\nu\right)\frac{e^{2\nu Y_t}-1}{e^{2\nu Y_t}+1}\,dt + \sigma\,dW_t, \tag{4.61}$$

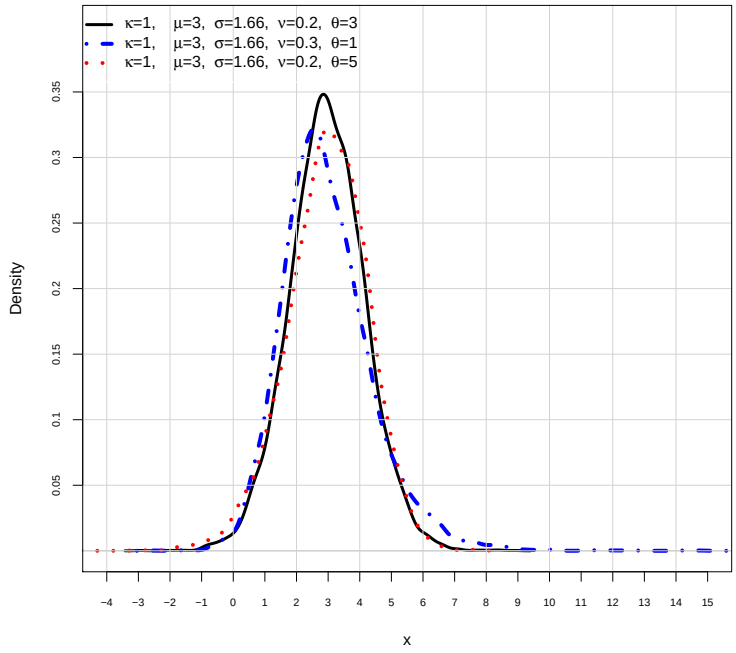

Figure 4.11: The empirical distributions using the three different parameter settings for the generalized on-factor mean-reverting process. We simulated 10,000 paths where the $\Delta t = \frac{1}{252}$, and $T = 5$ are considered using identical random generated numbers.

4.8.1 The Stationary Solution

Consider a stochastic differential equation as follows:

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t. \tag{4.62}$$

When this stochastic process 4.62 satisfies some relatively simple conditions, the process has a unique limiting distribution (stationary distribution) of X_T as T approaches infinity (see McLeish (2005) for more detail) where the distribution is in the following form:

$$P_{st}(x) = \frac{1}{c\,\sigma^2(z)} \exp\left(2 \int_0^x \frac{\mu(z)}{\sigma^2(z)} dz\right), \tag{4.63}$$

where c is the normalization constant equal to:

$$c = \int_{-\infty}^{\infty} \frac{1}{\sigma^2(z)} \exp\left(2 \int_0^x \frac{\mu(z)}{\sigma^2(z)} dz\right) dx < \infty, \tag{4.64}$$

Based on this definition, we derive the stationary solution for the mean-reverting process defined in equation 4.59 as follows:

$$P_{st}(x) = \frac{\exp\left\{\frac{2\kappa}{\sigma^2}\left(\frac{(\mu-\theta)\arctan(\nu(x-\theta))}{\nu} - \frac{\ln(\nu^2(x-\theta)^2+1)}{2\nu^2}\right)\right\}}{c\,\sigma^2(\nu^2(x-\theta)^2+1)}, \tag{4.65}$$

where c equals to:

$$c = \int_{-\infty}^{\infty} \frac{\exp\left\{\frac{2\kappa}{\sigma^2}\left(\frac{(\mu-\theta)\arctan(\nu(x-\theta))}{\nu} - \frac{\ln(\nu^2(x-\theta)^2+1)}{2\nu^2}\right)\right\}}{\sigma^2(\nu^2(x-\theta)^2+1)} dx, \tag{4.66}$$

To derive c, we first change variables ($u = \arctan(\nu(x-\theta))$) in the integral, then use *Mathematica* to obtain c as follows:

$$\begin{aligned} c \;=\; & \frac{\nu}{\kappa}\, i^{-\frac{2\kappa}{\sigma^2\nu^2}}\, 2^{-\frac{2\kappa}{\sigma^2\nu^2}-1} \left\{ i^{\frac{2\kappa}{\sigma^2\nu^2}} \left(\frac{{}_2F_1\left(-\frac{2\kappa}{\sigma^2\nu^2}, \frac{i\kappa(\mu\nu-\theta\nu+i)}{\sigma^2\nu^2}, \frac{\sigma^2\nu^2+i\kappa(\mu\nu-\theta\nu+i)}{\sigma^2\nu^2}, -1\right)}{\mu\nu-\nu\theta+i} - \right.\right. \\ & \left.\left. \frac{{}_2F_1\left(-\frac{2\kappa}{\sigma^2\nu^2}, \frac{\kappa(-i\mu\nu+i\theta\nu-1)}{\sigma^2\nu^2}, \frac{\sigma^2\nu^2+\kappa(-i\mu\nu+i\theta\nu-1)}{\sigma^2\nu^2}, -1\right)}{\mu\nu-\nu\theta-i} \right) + \right. \\ & \Gamma\left(\frac{2\kappa}{\sigma^2\nu^2}+1\right) \left(\frac{\Gamma\left(\frac{\sigma^2\nu^2+i\kappa(\mu\nu-\theta\nu+i)}{\sigma^2\nu^2}\right)\left(\sinh\left(\frac{\pi\kappa(\mu-\theta)}{\sigma^2\nu}\right) - \cosh\left(\frac{\pi\kappa(\mu-\theta)}{\sigma^2\nu}\right)\right)}{(\mu\nu-\nu\theta+i)\Gamma\left(\frac{\sigma^2\nu^2+\kappa(1+i\mu\nu-i\theta\nu)}{\sigma^2\nu^2}\right)} + \right. \\ & \left.\left. \frac{\Gamma\left(\frac{\sigma^2\nu^2+\kappa(-i\mu\nu+i\theta\nu-1)}{\sigma^2\nu^2}\right)\left(\sinh\left(\frac{\pi\kappa(\mu-\theta)}{\sigma^2\nu}\right) + \cosh\left(\frac{\pi\kappa(\mu-\theta)}{\sigma^2\nu}\right)\right)}{(\mu\nu-\nu\theta-i)\Gamma\left(\frac{\sigma^2\nu^2+\kappa(1-i\mu\nu+i\theta\nu)}{\sigma^2\nu^2}\right)} \right) \right\}. \end{aligned} \tag{4.67}$$

where
${}_2F_1(a,b,c,z)$ is the *the hypergeometric function* and defined as:

$${}_2F_1(a,b,c,z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k z^k}{(c)_k\, k!}$$

$(a)_0 = 1,\ (a)_k = a(a+1)(a+2)\ldots(a+k-1)$,
Notice that the definite integration of this real-valued function is a complex answer,

which does not sound quite right. However, when the result is numerically evaluated, the complex component is infinitesimally zero and such a tiny imaginary component is caused by round-off error. This is because Mathematica deploys a set of symbolic methods in the more general complex domain. Since calculation of this analytical result is more complicated than calculating the original integral 4.66 numerically, we prefer numerical computation of the integral over evaluating the expression 4.67. Figure 4.65 shows the stationary distribution for this mean-reverting process given in equation 4.59 in which its derived stationary solution is in equation 4.65. By looking at these graphs, we can see that the stationary distribution in this example is skewed to the left. The long-run mean, the variance, the skewness $\boldsymbol{\gamma_1} = \frac{E[X^3]-3E[x]Var(X)-(E[X])^3}{(Var(X))^{\frac{3}{2}}}$, and the excess kurtosis $\boldsymbol{\gamma_2} = \frac{E\left[(X-E[X])^4\right]}{\sigma^4} - 3$ of the stationary distribution 4.65 can be easily calculated numerically.

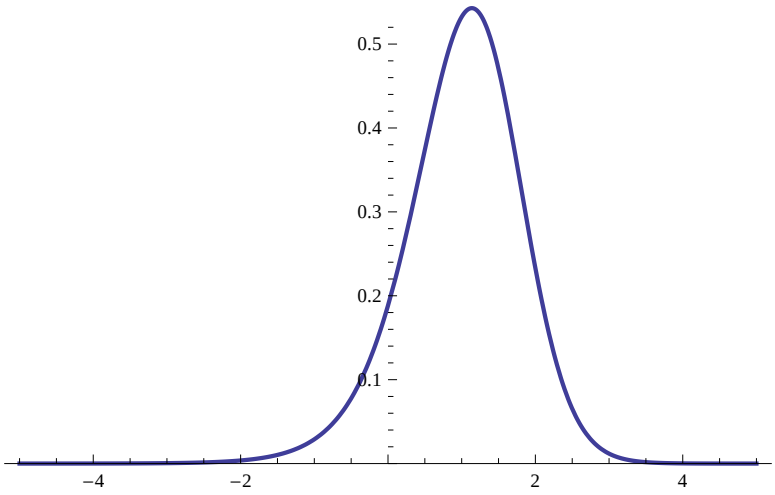

Figure 4.12: The plot depicts the stationary solutions (distributions) for this mean-reverting process given in the SDE form in equation 4.59 in which its derived stationary solution is in equation 4.65. Here, we have set $\kappa = 1$, $\mu = 1$, $\sigma = 1$, $\nu = 0.4$ and $\theta = 2$.

4.8.2 Parameter Estimation

This new mean-reverting SDE 4.59 is a nonlinear diffusion with a unique but unknown transition density. As a result, we are restricted to apply numerical or approximation methods. As we explained in 2.4.1, the new local linearization method introduced by Shoji and Ozaki (1998) is a strong approximation method that can be applied to the fairly general one-dimensional dynamics in the form 2.49 with given conditions. Therefore, since this new mean-reverting process 4.59 is in the required form and meets all conditions,

we can deploy this method to estimate the model's parameters. As it was reviewed this method in 2.4.1, the SDE should be transferred to the more tractable process in the form as follows:

$$dY_t = h(Y_t, t, \boldsymbol{\theta})\, dt + \sigma\, dW_t. \tag{4.68}$$

This form was derived for this new mean-reverting model 4.59 in equation 4.60. Here $h(y,t,\boldsymbol{\theta}) = \kappa(\mu-\theta)\text{sech}(\nu y) - \left(\frac{\kappa}{\nu} + \sigma^2\nu\right)\tanh(\nu y)$ and $y = \varphi(x) = \frac{\sinh^{-1}(\nu(x-\theta))}{\nu}$. Then the SDE 4.68 was linearized in the following bilinear SDE when $t \in [s, s+\Delta s)$:

$$dY_t = (L_s Y_t + M_s t + N_s)\, dt + \sigma\, dW_t, \tag{4.69}$$

where L_s, M_s, and N_s are constant and equal:

$$\begin{aligned} L_s &= \frac{\partial h(Y_s, s, \boldsymbol{\theta})}{\partial y}, \\ M_s &= \frac{\sigma^2}{2}\frac{\partial^2 h(Y_s, s, \boldsymbol{\theta})}{\partial y^2} + \frac{\partial h(Y_s, s, \boldsymbol{\theta})}{\partial t}, \\ N_s &= h(Y_s, s, \boldsymbol{\theta}) - Y_s \frac{\partial h(Y_s, s, \boldsymbol{\theta})}{\partial y} - s\, M_s. \end{aligned}$$

Here, the first derivatives with respect to y and t and second derivatives with respect to y of $h(y,t,\boldsymbol{\theta})$ are as follows:

$$\begin{aligned} \frac{\partial h(Y_s, s, \boldsymbol{\theta})}{\partial y} &= \frac{-\kappa(\mu-\theta)}{Y_s\sqrt{1-\nu^2 Y_s^2}} - \frac{\kappa + \nu^2\sigma^2}{1+\nu^2 Y_s^2}, \\ \frac{\partial h(Y_s, s, \boldsymbol{\theta})}{\partial t} &= 0, \\ \frac{\partial^2 h(Y_s, s, \boldsymbol{\theta})}{\partial y^2} &= \frac{\kappa(\mu-\theta)(1-2\nu^2 Y_s^2)}{Y_s^2(1-\nu^2 Y_s^2)^{\frac{3}{2}}} + \frac{2\nu^2(\kappa+\nu^2\sigma^2)Y_s}{(1+\nu^2 Y_s^2)^2}. \end{aligned}$$

In section 2.4.1, we show that the conditional distribution of Y_t given Y_s is normally distributed with mean and variance as follows:

$$E_s = E[Y_t \,|\, Y_s] = Y_s + \frac{h(Y_s, s, \boldsymbol{\theta})(e^{L_s(t-s)} - 1)}{L_s} + \frac{M_s(e^{L_s(t-s)} - 1 - L_s(t-s))}{L_s^2} \tag{4.70}$$

$$V_s = Var[Y_t \,|\, Y_s] = \frac{\sigma^2}{2L_s}\left(\, e^{2L_s\,(t-s)} \,-\, 1\,\right). \tag{4.71}$$

Thus, using transformation rule in probability density, we obtain the conditional distribution of X_t given X_s as follows:

$$f_{X_t|X_s}(x_t) = \frac{1}{\sqrt{2\Pi V_s}\sqrt{1+\nu^2(x_t-\theta)^2}} \exp\left\{ -\frac{\left(\frac{\sinh^{-1}(\nu(x_t-\theta))}{\nu} - E_s\right)^2}{2V_s} \right\}, \tag{4.72}$$

where E_s and V_s are given in equations 4.70 and 4.71 respectively and can be evaluated at x_s using $Y_s = \varphi(x_s) = \frac{\sinh^{-1}(\nu(x_s-\theta))}{\nu}$.
Notice that if we substitute $\mu^* = \mu - \lambda$, λ may be interpreted as the market price of risk. For μ in equation 4.59 and all related formulas, the results will be under the risk neutral measure Q.
Using the distribution of $X_t \mid X_s$ in this model for observations series $\{x_1, x_2, \cdots, x_n\}$, its log-likelihood function can be expressed as follows (see Shoji and Ozaki (1998) for more details):

$$\ln \mathcal{L}(\boldsymbol{\theta} \mid x_1\ ,\ x_2\ ,\ \cdots\ ,\ x_n) = -\frac{1}{2}\sum_{i=1}^{n-1}\left\{\frac{\left(\frac{\sinh^{-1}(\nu(x_{i+1}-\theta))}{\nu} - E_i\right)^2}{V_i} + \ln(2\Pi V_i)\right\} + \ln(f(x_1)) - \frac{1}{2}\sum_{i=1}^{n-1}\ln(1+\nu^2(x_i-\theta)^2). \tag{4.73}$$

4.8.3 The Futures Spread Price

Let $DF(s\,,T\,,X_s)$ to be the futures price for the spread with maturity T at time s. Consider the interest rate and the market price of risk to be constants. Therefore, the expected value of the spot spread price at time s under the risk neutral measure is the futures (forward) spread price. Unfortunately, the transition density of this model is unknown; therefore, the new local linearization method was used to approximate the transition density, given in equation 4.72. Using this transition density, the futures spread price can be approximated as follows:

$$\begin{aligned} DF(s\,,T\,,X_s) &= E_Q[X_T \mid X_s] = \int_{-\infty}^{\infty} x \frac{\exp\left\{-\frac{\left(\frac{\sinh^{-1}(\nu(x-\theta))}{\nu} - E_s^*\right)^2}{2V_s^*}\right\}}{\sqrt{2\Pi V_s^*}\sqrt{1+\nu^2(x-\theta)^2}}dx \\ &= \frac{e^{\frac{\nu(-2E_s+\nu V_s^*)}{2}}(e^{2\nu E_s^*}-1)}{2\nu} + \theta, \end{aligned} \tag{4.74}$$

where E_s^* and V_s^* are derived using equations 4.70 and 4.71 respectively when $\mu^* = \mu - \lambda$ is substituted for μ.

4.9 One-factor Regime Switching Model

Many financial and economic variables respond abruptly to some major events including financial crises (recession), major political decisions (quantitative easing), or natural

disaster (earthquake) causing a significant behavioral changes. Political instabilities or tensions in major crude oil producers and exporters will cause abrupt crude oil price increase, and rise its volatility. It is a commonly accepted fact that many economic and financial time series behave very differently in each business cycles namely *expansion (boom), medium, or contraction (recession)* (see Hamilton (2008)). In constructing a model for asset prices, each business cycle can be considered as a *regime (state). Hidden Markov Models (HMM)* (for more details see Elliott *et al.* (1995) and Cappe *et al.* (2005)) or *Regime-Switching Models* (RSM) are increasingly deployed in different subjects such as finance, economics, psychology, speech recognition, and genetics by scholars and can capture these abrupt behavioral changes in the term structures. Some applications of hidden Markov models can be found in studies such as Erlwein and Mamon (2009). Here, we will deploy the regime-switching model in which the spread between WTI and Brent crude oil follows our GOFMRM term structure in each regime. Figure 4.10 clearly illustrates that in early 2011, there was fundamentally abrupt structural changes in the observed spread (Section 4.5 provides some reasons for these changes). One can see that prior to 2011, the constant assumption of volatility in both one Elliott *et al.* (2005) and two Dempster *et al.* (2008) factor models is not quite an appropriate assumption; however, this new one factor model can capture this phenomena. Assume $\boldsymbol{z}_k$ to be a homogenous Markov chain with a finite state $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ in discrete time ($k = 0, 1, 2, \ldots$) and has dynamics as follows:

$$\boldsymbol{z}_{k+1} = \boldsymbol{\Pi}\boldsymbol{z}_k + \boldsymbol{\zeta}_k, \tag{4.75}$$

where $\boldsymbol{\zeta}_k$ is the martingale increment ($E[\boldsymbol{\zeta}_k|\mathfrak{F}_{k-1}] = 0$) and $\boldsymbol{\Pi} = (\pi_{ij})$ is the transition probability matrix.
Each state in the Markov chain $\boldsymbol{s}_i$ can be presented with the canonical basis $\boldsymbol{e}_i \in \Re^n$, where $\boldsymbol{e}_i$ is the unit vector taking 1 in the i^{th} position and 0 elsewhere $i = 1, 2, \ldots, n$. We consider $< \boldsymbol{z}_k, \boldsymbol{e}_i >= \mathbf{1}_{\{\boldsymbol{z}_k = \boldsymbol{e}_i\}}$ to express each $\boldsymbol{z}_k$ in canonical basis, where $< ., . >$ presents the ordinary Euclidean scalar product in $\Re^n$. For instance, if $n = 2$, $< \boldsymbol{z}_k, \boldsymbol{e}_1 >$ and $< \boldsymbol{z}_k, \boldsymbol{e}_2 >$ represent two different states in the model. Let $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_n)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T, \boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)^T$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_n)^T$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^T$ to be the new mean-reverting model's parameters in regime switching form. Here, T represents matrix transposition. To simplify the notation, we also assume $\kappa(\boldsymbol{z}_k) =< \boldsymbol{\kappa}, \boldsymbol{z}_k >$, $\mu(\boldsymbol{z}_k) =< \boldsymbol{\mu}, \boldsymbol{z}_k >$, $\sigma(\boldsymbol{z}_k) =< \boldsymbol{\sigma}, \boldsymbol{z}_k >$, $\nu(\boldsymbol{z}_k) =< \boldsymbol{\nu}, \boldsymbol{z}_k >$, and $\theta(\boldsymbol{z}_k) =< \boldsymbol{\theta}, \boldsymbol{z}_k >$. This assumptions lead that this generalized mean-reverting dynamics to be expressed in regime switching form as follows:

$$dX_t = \kappa(\boldsymbol{z}_t)\left(\mu(\boldsymbol{z}_t) - X_t\right) dt + \sigma(\boldsymbol{z}_t)\sqrt{\nu(\boldsymbol{z}_t)^2(X_t - \theta(\boldsymbol{z}_t))^2 + 1}\, dW_t. \tag{4.76}$$

It follows from the explanations in this section and section 4.5 that the behavior of the spread between WTI and Brent crude oils made a fundamental structural change in early 2011. Consequently, the spread switched to a new regime and has stayed in this new

regime since. In a general regime-switching algorithm, at any given time, there is always the possibility to switch from one regime to another based on the transition probability matrix. However, based on the empirical observations, it is assumed that there are two regimes and it is exactly known to which regime each observation belongs. This observation considerably simplifies the calibration of the model. As a result, to calibrate the model, we merely first fit the daily spread observations between WTI crude oil and Brent oil from April 1994 to December 2010 to this new generalized mean-reverting process and it is considered as first regime and then separately fit the daily observations between WTI crude oil and Brent oil from January 2011 to November 2013 to the model and it is considered as a second regime. This exclusive property of this spread enable us to implement the regime switching model without using actual a regime switching algorithm. Therefore, it is not necessary to discuss the regime switching algorithm in further detail here and interested readers can be referred to Elliott *et al.* (1995) and Cappe *et al.* (2005). It is required to emphasize that although we explained the regime switching algorithm, the way the model was implemented, the *structural break* model was really deployed not regime switching algorithm.

4.9.1 Empirical Results

The spot spread (front contracts spread) between WTI and Brent crude oils from April 1994 to November 2013 is considered. Here we compare the two models, and Elliott *et al.* (2005)'s one-factor model, Vasicek, reviewed in section 4.3, and the new mean-revering process, introduced in this section using regime switching algorithm. It will be shown which one can capture this spread process properties. The new local linearization method (2.4.1) is deployed to estimate the parameters of the Vasicek and the generalized mean-revering models for the daily spread between WTI and Brent crude oils from April 1994 to December 2010 for the first regime and applied to the observed data from January 2011 to November 2013 to estimate the second regime parameters in both methods. The calibration results in both regimes for both models applied to our dataset are summarized in table 4.7.

4.9.2 Goodness of Fit

Whether this generalized one-factor mean-reverting process can capture the reality of the spread process better than the Vasicek process or not is the question of section. The estimation results with the tests results for goodness of fit is summarized in table 4.7 for the observed spread series between WTI and Brent oils in two regimes. It is seen that both AIC and BIC values in both regimes confirm the superiority of this generalized mean-reverting model over the Vasicek model and this new model especially leads the Vasicek in first regime. Figure 4.13 demonstrates the comparisons results for the fitted

Table 4.7: Estimated parameters for two processes: generalized mean-reverting and Vasicek one-factor models using daily spread between WTI and Brent oils in two regimes when first regime observations are from April 1994 to December 2010 and second regime observations are from January 2011 to November 2013.

Model	Regime	κ	μ	σ	ν	θ	-2 logl[2]	AIC[3]	BIC[4]
New MR	First	7.054	1.573	4.215	1.254	1.455	4564.6	4574.6	4581.2
Process	Second	3.444	-15.840	12.415	0.054	-2.444	2008.1	2018.1	2021.3
Vasicek	First	15.923	1.105	9.468	-	-	7250.3	7256.3	7267.0
Process	Second	3.970	-15.886	15.485	-	-	2035.6	2041.6	2048.8

Note: 1-Log-likelihood, 2-Akaike information criterion, 2- Bayesian information criterion.

models in both regimes for both models based on estimated parameters summarized in table 4.7: The left plot depicts comparison between the empirical density for the observed daily spread between WTI and Brent crude oils data (April 1994 to December 2010) in first regime with the fitted stationary distributions of these two models: Vasicek and generalized one-factor mean-reverting. The right plot shows the comparison between the empirical density for the observed daily spread between WTI and Brent crude oils data (January 2011 to November 2013) in second regime with the fitted stationary distributions of these two models. These subplots of figure 4.13 show how this new powerful dynamics captures the main attributes of the observed data in both regimes especially in the first regime.

4.10 Conclusion:

In this chapter, we described three main approaches to pricing a spread process namely cointegration, one-factor and two-factor models. We apply these three models to our real empirical sample data and compare the results. Later, we analyze the recent behavioral change in the location spread between WTI crude oil and Brent oil. Since these models are not flexible enough to capture all behavior changes, we extend the one-factor and two-factor spread models by adding a compound Poisson process where jump sizes follow a double exponential distribution. However, the implementation of these two processes with jumps are left for future studies. Later, the generalized one-factor mean-reverting dynamics is introduced to capture the major properties of spread processes, compared to Vasicek process and showed the new model's absolute superiority by fitting to the spread between WTI and Brent oils. Since the spread between WTI and Brent in observed data is experienced fundamental change in early 2011, the generalized model and Vasicek process are deployed in regime switching framework by considering two regimes. The

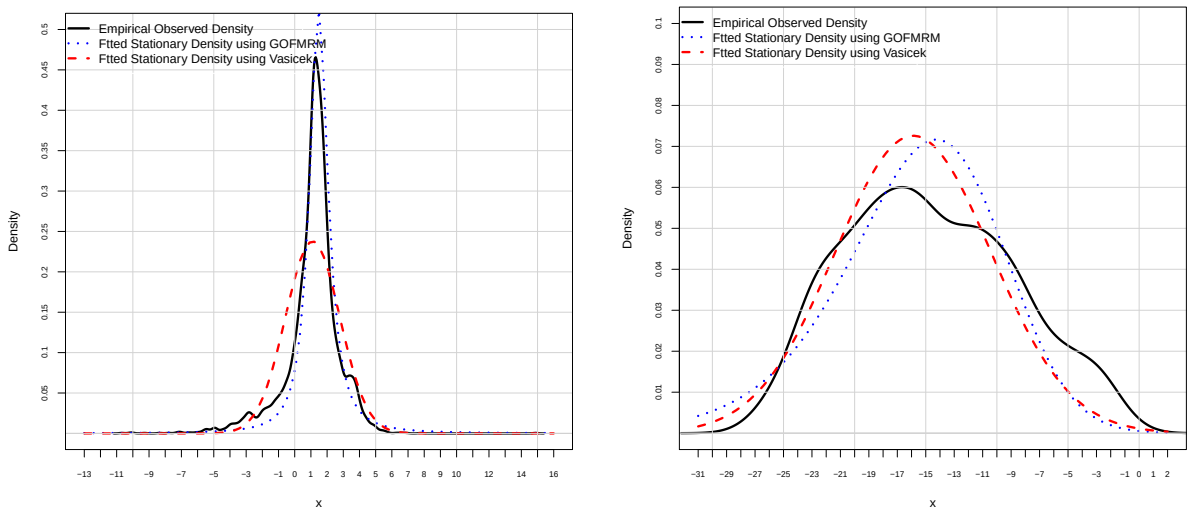

Figure 4.13: The left plot depicts comparison between the empirical density for the observed daily spread between WTI and Brent crude oils data (April 1994 to December 2010) in first regime with the fitted stationary distributions of these two models: Vasicek and generalized one-factor mean-reverting. The right plot shows the comparison between the empirical density for the observed daily spread between WTI and Brent crude oils data (January 2011 to November 2013) in second regime with the fitted stationary distributions of these two models. The fitted parameters are summarized in table 4.7 in both regimes.

estimation results are fully justified applying the regime switching algorithm. The AIC and BIC criteria in both regimes choose this new generalized model. In chapter 5, we will introduce a novel mean-reverting random walk, obtain its continuous stochastic process, generalized the continuous dynamics, and apply to empirical data to show the flexibility of the model and its advantages over the Elliott *et al.* (2005)'s one-factor model.

Chapter 5

MODELING ENERGY SPREADS WITH A NOVEL MEAN-REVERTING STOCHASTIC PROCESS

5.1 Introduction

There are well developed futures markets for crude oil, in various grades, delivered at various locations, and delivered at various dates. Market participants often trade spreads between these futures because of the fundamental relationships between them. The main concept behind the spread trading strategy relies on fluctuation about the equilibrium level (long-run mean) of the spread process. In other words, *mean-reversion* is the most important property of the spread process that makes market participants consider taking advantage of it. In spread trading, we construct a portfolio of two commodities that are closely related to each other and have similar characteristics. In some cases as, for instance, the spread between West Texas Intermediate (WTI), low-sulfur, a low-density crude oil priced in Cushing, Oklahoma and West Texas Sour (WTS), high-sulfur, a medium-density crude oil priced in Midland, Texas, the underlying commodities fulfill nearly identical needs. The spread trading belongs to the *market neutral* class of trading strategies. This is because its return is largely independent of the market return (for more details see Vidyamurthy (2004)). The fundamental base for the spread trading, known as *pairs trading* in stock markets, is *relative pricing*. The choice of assets used in pairs trading is based on the philosophy "if it walks like a duck and quacks like a duck, it must be a duck". In other words, pick assets with similar characteristics. In fact, a comprising the difference portfolio of two similar assets the price of which devi-

ate from their historical equilibrium level known as the long-run mean is constructed. The portfolio involves a short position in the overpriced and a long position in the underpriced security with a predetermined ratio or *hedging ratio*. In spread trading, the hedging ratio is sometimes known and need not be estimated. The location spread (e.g. spread between WTI and Brent crude oils) and calendar spread (e.g. spread between two futures prices with different expiries for WTI) are other examples in the energy markets. Spread trading can also be used for hedging purposes. For instance, refineries use various *crack spreads* (spread between crude oil and refined products) to hedge against exposure to the price risk between their input; output commodities. One simple way to model a spread portfolio is to consider each underlying commodity's spot prices separately, for instance, to follow the classical two-factor model introduced by Gibson and Schwartz (1990) where their returns are correlated. However, due to high volatility of energy markets, Alexander (1999) shows that a constant correlation assumption is not quite appropriate. Other classical methods to model the spread are based on long-term relationship between underlying equities or commodities known as *cointegration*. Two unit-root processes (integrated of order 1) are called cointegrated if a linear combination of them that is a stationary process exists. If two securities or commodities are cointegrated, their spread can be modeled directly. However, the cointegration based models may fail to capture all properties of some spread processes (e.g. spread between WTI and Brent crude oils) in energy markets. The Vasicek process is another trivial stochastic model to capture an essential property of the spread process, mean-reversion. Elliott *et al.* (2005) introduced the Vasicek process to model the spread process in equity markets. However, in section 5.3, we will show that the Vasicek dynamics cannot fully describe the reality of the spread process. Dempster *et al.* (2008) also proposed a two factor model to model the spot spread process in energy markets to evaluate spread options. In this study, we introduce a one-factor mean-reverting process to capture not only the mean-reverting property of the spread process, but also the skewness, the heavy tails and the kurtosis features of the spread process. This chapter is organized as follows:

In section 5.2, we empirically analyze the observed spread between WTI and WTS crude oils. We review Elliott *et al.* (2005)'s one-factor Vasicek-like model and list its advantages and disadvantages in section 5.3. In section 5.4, we define the contemporary mean-reverting random walk. In section 5.5, we derive the *scaling limit* of this mean-reverting RW and find its stochastic process. The new mean-reverting process is compared to Elliott *et al.* (2005)'s one factor model and its advantages are listed in section 5.6. Section 5.7 reviews the new local linearization method introduced by Shoji and Ozaki (1998) to estimate parameters of this nonlinear stochastic process. In section 5.8, we deploy both models: the new one factor mean-reverting model and the Vasicek process, to price the spread between WTI and WTS crude oils. Using both observed and estimated results, we discuss which process can better describe the reality of the spread process.

5.2 Empirical Analysis of The Spot Spread Process

The West Texas Intermediate (WTI) crude spot prices are unobservable, but the front month futures prices are commonly considered as approximation of the spot prices for WTI crude oil. In this paper, Chicago Mercantile Exchange (CME) daily futures prices (the front contract) of WTI crude oil and CME spot prices of West Texas Sour (WTS) oil are used from January 2000 to January 2013 (Bloomberg (2014)). Figure 5.1 depicts the daily front contract of WTI crude oil prices and spot WTS crude oil prices and their spread series. The spread dataset is summarized in table 5.1 and the spread's empirical density plotted in figure 5.2. Both the summarized data and the empirical density evidently show that the spread series is right skewed and leptokurtic. Although it is generally accepted that WTI crude oil and WTS crude oil are unit root processes ($I(1)$) (see e.g. Dempster *et al.* (2008)), we apply the *augmented Dickey-Fuller test (ADF)* and *Elliot, Rothenberg & Stock test (ERS)* tests to test for the unit root in WTI and WTS oils price series. The null hypothesis in both tests is that the series contains a unit root. The results of the tests are summarized in table 5.2. Both unit root tests for each WTI and WTS price series confirm that the null hypothesis should not be rejected in favor of the alternative hypothesis at one percent level of significance. Figure 5.3 illustrates the mean-reversion property of the observed spread dataset using two subplots: the difference of standardized spread vs standardized spread and average of partitioned difference spread vs average of partitioned standardized spread. The spread series also seems to show that when the WTI and WTS prices increase, their spread becomes more volatile although this conclusion is drawn from the 2008 to 2010 financial crisis must be interpreted with caution. Their spread also starts widening in favor of WTI due to higher quality (lighter and lower in sulfur) of WTI. This phenomena can generate right skewness.

Table 5.1: The summarized information of the observed dataset (the daily spread between WTI and WTS from January 2000 to January 2013.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skew.	E. Kurtosis
-0.540	1.840	2.700	3.146	4.300	10.750	0.873	0.663

5.3 One Factor Model for the Spread Process:

The Vasicek process was initially proposed by Vasicek (1977) to model the evolution of short interest rates so as to capture one crucial feature of such rates, their mean-reversion. Under the Vasicek process, interest rates can become negative in principle, which at one time was considered to be a major disadvantages when the process is used to

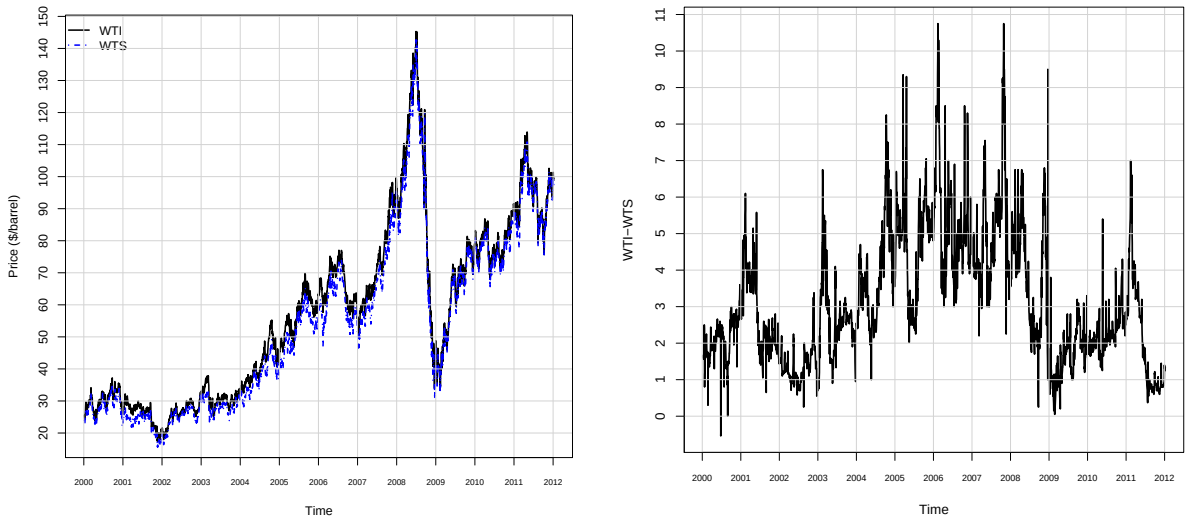

Figure 5.1: The daily front contract of WTI crude oil prices and spot WTS crude oil prices and their spread series from January 2000 to January 2013 (price in $/barrel).

Table 5.2: Unit root tests for daily WTS spot prices and WTI front daily contracts future price series and their differenced series from January 2000 to January 2013.

Series	ADF test (CV)				ERS test (CV)			
	t-Stat	1pct	5pct	10pct	t-Stat	1pct	5pct	10pct
WTS	-1.2566	-3.46	-2.88	-2.57	17.4293	1.99	3.26	4.48
ΔWTS	-40.2462	-2.58	-1.95	-1.62	0.015	1.99	3.26	4.48
WTI	-1.2424	-3.46	-2.88	-2.57	19.1049	1.99	3.26	4.48
ΔWTI	-41.0425	-2.58	-1.95	-1.62	0.0179	1.99	3.26	4.48

model interest rates. However, this is not a disadvantage, but rather an essential feature to model spread processes. Because of this and because of its simplicity and analytical tractability, Elliott *et al.* (2005) proposed that the spread dynamics follows the Vasicek process. In this model, the spot spread, X_t is the only factor (state variable) and is assumed to follow the Vasicek stochastic process as follows:

$$dX_t = \kappa\,(\mu\ -\ X_t)\ dt + \sigma\, dW_t, \tag{5.1}$$

where the model assumes that the linear combination, X_t evolves as an Ornstein-Uhlenbeck process with constant coefficients, $\kappa > 0$ is the speed of mean-reversion, μ is the long-run spread mean, σ is the volatility of the process, and dW_t is the increment

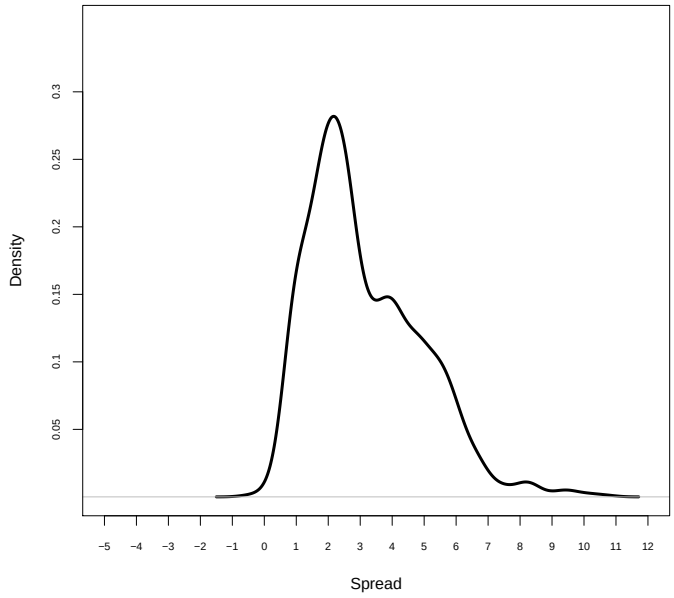

Figure 5.2: The plot depicts the empirical density of the observed daily spread between WTI and WTS crude oil from January 2000 to January 2013.

of a standard Brownian motion. By applying Ito's lemma, it is easy to show that the stochastic differential equation (SDE), 5.1 for any $s\,(0 < t \leq s)$ has a solution as follows:

$$X_s \;=\; e^{-\kappa\,(s-t)}\,X_t \;+\; \mu\,(\,1 \;-\; e^{-\kappa\,(s-t)}\,) \;+\; \sigma\,e^{-\kappa\,s}\int_t^s e^{\kappa\,u}\mathrm{d}W_u. \tag{5.2}$$

Equation 5.2 shows that the conditional distribution of X_T given X_t is normally distributed with mean and variance:

$$E[\,X_T \mid X_t\,] \;=\; e^{-\kappa\,(T-t)}\,X_t \;+\; \mu\,(\,1 \;-\; e^{-\kappa\,(T-t)}\,), \tag{5.3}$$

$$Var[\,X_T \mid X_t\,] \;=\; \frac{\sigma^2}{2\kappa}\,(\,1 \;-\; e^{-2\kappa\,(T-t)}\,). \tag{5.4}$$

Elliott *et al.* (2005) pointed out that the method has advantages including capturing mean-reversion, the fundamental property of pairs trading strategy, considering time continuous process for the spread process and tractability of the method (the parameters can easily be estimated using the Kalman filter algorithm). However, there are some convincing reasons why this method cannot be applied to most commodities spread processes. The underlying commodities (future) prices in the spread process rapidly respond to eventful market news. These events sometimes highly impact one of the

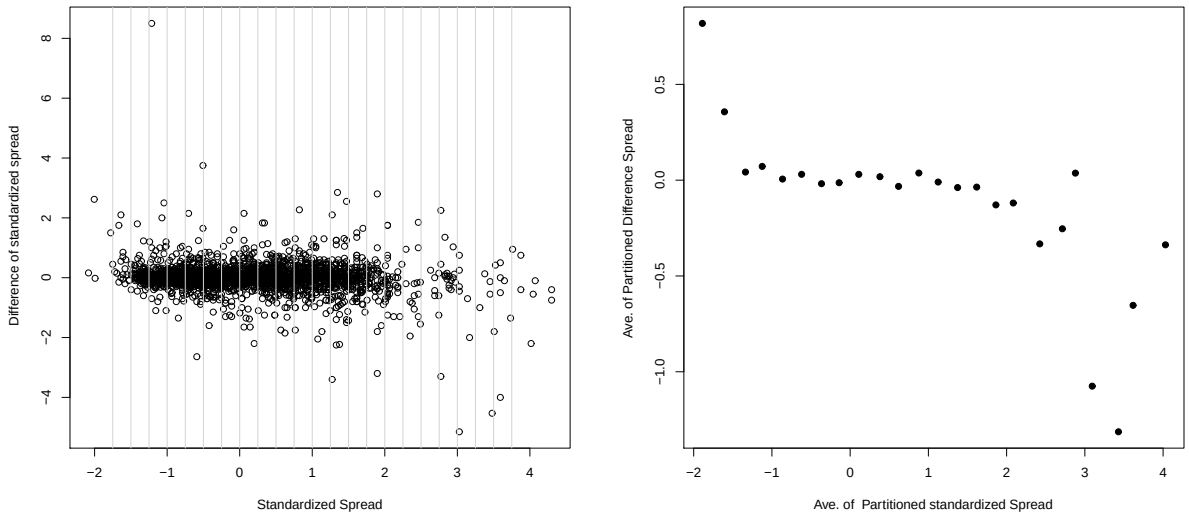

Figure 5.3: Left panel is difference of standardized spread ($\frac{\text{spread}-\mu(\text{spread})}{\sigma(\text{spread})}$) vs standardized spread, showing slight evidence of mean-reversion. To show the pattern more clearly, we binned the horizontal axis into 25 bins of equal width; calculate the average of the points in the bins in right panel. This shows mean-reversion more clearly.

underlying prices and have less impact on the other one causing the volatility of the spread process to increase. This phenomenon can be tracked in the form of skewness, kurtosis, heavy tails in the transition density of the price spread process, and the empirical results discussed in section 5.2 confirm this claim. To model the spread dynamics, we propose a new mean-reverting process to resolve these issues with of Elliott *et al.* (2005)'s model. Without sacrificing simplicity; its small number of parameters, we will, as shown below have to sacrifice some analytical tractability. We first construct its random walk model in section 5.4. Then we derive its continuous stochastic form and its generalized dynamics in sections 5.5 and 5.6 respectively.

5.4 Mean-reverting Random Walk:

A random walk (RW) is a mathematical mechanism to model a path based on a succession of random steps. A random walk can be applied to trace the behavior of various paths including the evolution of stock prices, the financial status of a gambler, a drunkard walking, and a molecule traveling in liquid. Random walks are deployed in many sciences such as finance, physics, economics, and computer science to capture behaviors of various processes. Random walks have various forms and are usually considered as

Markov chains, but sometimes can define more complicated walks. Random walks can occur in one, two, or many dimensions. Moreover, random walks may change based on the time parameters. For example, simple discrete time walks are indexed by the natural numbers while some sophisticated walks can be assumed to take steps at random times. Here, we construct a walk that is mean-reverting. In this mean-reverting RW, we define the probabilities of taking each step either forward or backward to depend on the current location of the walker. When the walker diverges from the mean, by changing the probabilities of traveling forward and backward, the walker will tend to revert back to the mean, which means that the force will be stronger as the walker deviates further away from the mean. Let S_i be the position at i^{th} step and let P_i be the probability of moving right when we take i^{th} step. We update the probabilities of moving to right or left based on following function:

$$P_i(\text{right}) = \begin{cases} \dfrac{1}{2 + aS_{i-1}}, & \text{if } S_{i-1} \geq 0 \\ 1 - \dfrac{1}{2 - aS_{i-1}}, & \text{elsewhere} \end{cases} = \begin{cases} \dfrac{1}{2} - \dfrac{aS_{i-1}}{2(2 + aS_{i-1})}, & \text{if } S_{i-1} \geq 0, \\ \dfrac{1}{2} - \dfrac{aS_{i-1}}{2(2 - aS_{i-1})}, & \text{elsewhere,} \end{cases} \tag{5.5}$$

where
$S_{i-1} \in \mathbb{Z}$ is the current location,
$a \geq 0$ is the mean-reversion speed.

We can see as S_{i-1} increases the probability of going right decreases and vice versa. When S_{i-1} approaches infinity, P_i approaches zero. As a decreases, positive P_i decreases or negative P_i increases with lower speed and approaches to a simple random walk as a approaches to zero. Consider n be the number of steps taken and let each step to be $+1$ or -1. Then the number of possible different paths that can be traveled will be 2^n. The number of walks that satisfy $S_n = k$ where $k > 0$ equal to the number of ways of choosing $(n + k)/2$ elements from an n element set (for this to be non-zero, it is necessary that $n + k$ be an even number), $\binom{n}{(n+k)/2}$. Note that for simple random walk the $P(S_n = k)$ is equal to $2^{-n}\binom{n}{(n+k)/2}$. Here, in this new setting, the probability changes according to the location; as a result, each path to reach k from origin in n walks has its own probability and may be different from another path. One evident but impractical way to calculate this probability is to calculate the probability for every possible path and sum them up to come up with the $P(S_n = k)$.

Lemma 5.4.1. The mean-reverting random walk generated by the transition probability function 5.5 is symmetric:

We need to show that $P(S_n = k) = P(S_n = -k)$ *where* $k > 0$. *To show this it is sufficient*

to prove that for any arbitrary path, $R = \{0 \to k_1 \to k_2 \cdots \to k_{n-1} = k-1 \to k\}$ that reaches k starting from origin, there exists a corresponding path, $R^- = \{0 \to -k_1 \to -k_2 \cdots \to k_{n-1} = -k+1 \to -k\}$ to reach $-k$ starting from origin with identical probability. To build path R^- from path R, we start from origin and in each step of path R, we take a step for R^- in opposite direction so by continuing this method at any given stage of building the paths, the locations in the paths have identical distances but opposite direction from origin. Clearly, based on the definition of the transition probability function 5.5, the probabilities of path R and R^- are equal.

5.5 Mathematical Derivation of Continuous-time Form

We attempt to derive probability of being in location k after taking n steps, $P(k\,,\,n)$. We obtain the difference equation to calculate $P(k\,,\,n)$ for $k \in \{-n, -n+2, \cdots, n-2, n\}$ as follows:

$$P(k\,,\,n) = \begin{cases} \left(\frac{1}{2} + \frac{a(k+1)}{2(2\,+\,a(k\,+\,1))}\right) P(k+1\,,\,n-1) + \\ \quad \left(\frac{1}{2} - \frac{a(k-1)}{2(2\,+\,a(k\,-\,1))}\right) P(k-1\,,\,n-1), & \text{if } k \geq 1, \\ \left(\frac{1\,+\,a}{2\,+\,a}\right) P(1\,,\,n-1) + \\ \quad \left(\frac{1\,+\,a}{2\,+\,a}\right) P(-1\,,\,n-1), & \text{if } k = 0, \\ \left(\frac{1}{2} + \frac{a(k+1)}{2(2\,-\,a(k\,+\,1))}\right) P(k+1\,,\,n-1) + \\ \quad \left(\frac{1}{2} - \frac{a(k-1)}{2(2\,-\,a(k\,-\,1))}\right) P(k-1\,,\,n-1), & \text{if } k \leq -1, \end{cases} \tag{5.6}$$

where

$$\begin{cases} P(0\,,\,0) \;=\; 1, \\ P(k\,,\,0) \;=\; 0 \quad \text{for } k \neq 0. \end{cases}$$

It might be easier to solve this difference equation, (5.6) by transforming to continuous form and attempting to solve its corresponding partial differential equation (PDE). To do so, we define step size as Δx taking each Δt time. But before we proceed, we must modify the probabilities of moving right or left in equation 5.5 such that to be applicable

in continuous form as follows:

$$P(\text{right}) = \begin{cases} \frac{1}{2} - \frac{ax}{2(2\ +\ ax)}\frac{\Delta t}{\Delta x}, & \text{if } x \geq 0, \\ \frac{1}{2} - \frac{ax}{2(2\ -\ ax)}\frac{\Delta t}{\Delta x}, & \text{elsewhere,} \end{cases} \tag{5.7}$$

Note that at given arbitrary location x, the probability of moving right, $P_r = \frac{1}{2} - \frac{ax}{2(2+a|x|)}$ and the probability of moving left, $P_l = \frac{1}{2} + \frac{ax}{2(2+a|x|)}$ lead to an average drift of $P_r - P_l = \frac{-ax}{2+a|x|}$ space steps per time step. In other words, a speed of $\left\{\frac{-ax}{2+a|x|}\frac{\Delta x}{\Delta t}\right\}$. That is all fine as long as we keep the time and space steps the same. But it leads us into problems when we refine the grid by taking the limit as $\Delta x \to 0$ and $\Delta t \to 0$. We observe that we want to make the mean speed the same no matter what time step we pick. To get that we write $P_r = \frac{1}{2} - \frac{ax\Delta t}{2\Delta x(2+a|x|)}$ and similarly for $P_l = \frac{1}{2} + \frac{ax\Delta t}{2\Delta x(2+a|x|)}$. This will give us a net speed of $\frac{-ax\Delta t}{\Delta x(2+a|x|)}\frac{\Delta x}{\Delta t} = \frac{-ax}{2+a|x|}$, independent of the discretization. In the original integer lattice setting, we do not see this because $\Delta t = \Delta x = 1$, so it does not make any difference. As we can see, this mean-reverting random walk resembles a single random walk with drift. The probability that a particle is at location $x = k\Delta x$ at time $t = n\Delta t$, where $k \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$ is:

$$P(x\,,\,t) = \begin{cases} \left(\frac{1}{2} + \frac{a(x\ +\ \Delta x)}{2(2\ +\ a(x\ +\ \Delta x))}\frac{\Delta t}{\Delta x}\right) P(x+\Delta x\,,\,t-\Delta t) + \\ \quad \left(\frac{1}{2} - \frac{a(x\ -\ \Delta x)}{2(2\ +\ a(x\ -\ \Delta x))}\frac{\Delta t}{\Delta x}\right) P(x-\Delta x\,,\,t-\Delta t), \quad \text{if } x > 0, \\ \left(\frac{1}{2} + \frac{a(x\ +\ \Delta x)}{2(2\ -\ a(x\ +\ \Delta x))}\frac{\Delta t}{\Delta x}\right) P(x+\Delta x\,,\,t-\Delta t) + \\ \quad \left(\frac{1}{2} - \frac{a(x\ -\ \Delta x)}{2(2\ -\ a(x\ -\ \Delta x))}\frac{\Delta t}{\Delta x}\right) P(x-\Delta x\,,\,t-\Delta t), \quad \text{if } x < 0, \end{cases} \tag{5.8}$$

Equivalently, equation 5.8 can be written in the following general form for arbitrary $x \in \Re$:

$$\begin{aligned} P(x\,,\,t) \;\; = \;\; & \left(\frac{1}{2} + \frac{a(x\ +\ \Delta x)}{2(2\ +\ a\mid x\ +\ \Delta x\mid)}\frac{\Delta t}{\Delta x}\right) P(x+\Delta x\,,\,t-\Delta t)\ + \\ & \left(\frac{1}{2} - \frac{a(x\ -\ \Delta x)}{2(2\ +\ a\mid x\ -\ \Delta x\mid)}\frac{\Delta t}{\Delta x}\right) P(x-\Delta x\,,\,t-\Delta t), \end{aligned} \tag{5.9}$$

To calculate $P(x\,,\,t+\Delta t)$ for arbitrary $(x \in \Re)$, we deploy equation 5.9 as follows:

$$\begin{aligned} P(x\,,\,t+\Delta t) \;\; = \;\; & \left(\frac{1}{2} \,+\, \frac{a(x \,+\, \Delta x)}{2(2 \,+\, a \mid x \,+\, \Delta x \mid)} \frac{\Delta t}{\Delta x}\right) P(x+\Delta x\,,\,t) \,+ \\ & \left(\frac{1}{2} \,-\, \frac{a(x \,-\, \Delta x)}{2(2 \,+\, a \mid x \,-\, \Delta x \mid)} \frac{\Delta t}{\Delta x}\right) P(x-\Delta x\,,\,t) \end{aligned} \tag{5.10}$$

or equation 5.10 can be written in the form as follows:

$$\begin{aligned} P(x\,,\,t+\Delta t) \;\; = \;\; & \frac{1}{2}\left[\, P(x+\Delta x\,,\,t) \,+\, P(x-\Delta x\,,\,t)\right] \,+ \\ & \frac{\Delta t}{2\,\Delta x}\left[\, f(x+\Delta x\,,\,t) \,-\, f(x-\Delta x\,,\,t)\right], \end{aligned} \tag{5.11}$$

where $f(x\,,\,t) = \left(\frac{ax}{2+a|x|}\right) P(x\,,\,t)$.

Now, we expand all terms on both side of equation 5.11, in a Taylor series as follows:

$$\begin{aligned} & P(x\,,\,t+\Delta t) = P(x,t) \,+\, \frac{\partial P(x,t)}{\partial t}\Delta t \,+\, \frac{1}{2}\,\frac{\partial^2 P(x,t)}{\partial t^2}\Delta t^2 \,+\, O(\Delta t^3), \\ & \frac{1}{2}\left[\, P(x+\Delta x\,,\,t) \,+\, P(x-\Delta x\,,\,t)\right] \,=\, P(x,t) + \frac{1}{2}\,\frac{\partial^2 P(x,t)}{\partial x^2}\Delta x^2 \,+\, O(\Delta x^4), \\ & \frac{\Delta t}{2\,\Delta x}\left[\, f(x+\Delta x\,,\,t) \,-\, f(x-\Delta x\,,\,t)\right] = \frac{\Delta t}{\Delta x}\left\{\frac{\partial f(x,t)}{\partial x}\Delta x \,+\, O(\Delta x^3)\right\}, \end{aligned} \tag{5.12}$$

we plug the results in equation 5.12 into equation 5.11 and after simplifying and dividing both sides by Δt , we have:

$$\begin{aligned} & \frac{\partial P}{\partial t} \;\; + \;\; \frac{\partial^2 P}{2\partial t}\Delta t \,+\, \cdots \,= \\ & \qquad \frac{\partial}{\partial x}\left\{\frac{ax}{2 \,+\, a \mid x \mid} P\right\} \,+\, \frac{1}{2}\frac{\partial^2}{\partial x^2}\left\{P\frac{\Delta x^2}{\Delta t}\right\} \,+\, \cdots \end{aligned} \tag{5.13}$$

For this approach to work, we must take the limit in equation (5.13) as $\Delta x \to 0$ and as

$\Delta t \to 0$ in a particular way such that:

$$D \;=\; \lim_{\Delta x, \Delta t \to 0} \frac{\Delta x^2}{\Delta t}, \quad \text{some number } D.$$

Therefore, the resulting PDE for arbitrary $x \in \Re$ is:

$$\frac{\partial P(x\,,\,t)}{\partial t} \;=\; -\frac{\partial}{\partial x}\left(\frac{-ax}{2 \;+\; a \mid x \mid} P(x\,,\,t)\right) \;+\; \frac{D}{2}\,\frac{\partial^2 P(x\,,\,t)}{\partial x^2}. \tag{5.14}$$

It is now easy to generalize the PDE in equation 5.14 by including another parameter $\kappa > 0$, to have another control for mean-reverting speed for arbitrary $x \in \Re$, as follows:

$$\frac{\partial P(x\,,\,t)}{\partial t} \;=\; -\frac{\partial}{\partial x}\left(\frac{-\kappa a x}{2 \;+\; a \mid x \mid} P(x\,,\,t)\right) \;+\; \frac{D}{2}\,\frac{\partial^2 P(x\,,\,t)}{\partial x^2}, \tag{5.15}$$

where the boundary conditions and the initial condition are as follows:

$$\begin{aligned} \lim_{x \to \pm\infty} P(x\,,\,t) \;&=\; \lim_{x \to \pm\infty} \frac{\partial P(x\,,\,t)}{\partial x} \;=\; 0, \\ P(x\,,\,t=0) \;&=\; \delta(x), \end{aligned} \tag{5.16}$$

and $\delta(x)$ is the *Dirac delta* function $\left(\delta(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\epsilon^2}\right)\right)$.

5.5.1 The Fokker Planck Equation

The *Fokker Planck equation (FPE)* provides a practical methods for stochastic modeling in wide range of studies including finance, physics, and biology Risken (1989). The FPE describes the probability density function that evolves in time (the continuous stochastic process) as a partial differential equation. The general one-dimension FPE for the probability density function, $P(x\,,\,t)$ is in the following generic form:

$$\frac{\partial P(x\,,\,t)}{\partial t} \;=\; -\frac{\partial}{\partial x}\left[\mu(x,t) P(x\,,\,t)\right] \;+\; \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\left[D(x,t) P(x\,,\,t)\right], \tag{5.17}$$

where $\mu(x,t)$ is the drift or force and $D(x,t)$ is the diffusion coefficient.
The stochastic differential equation (SDE) for this $P(x\,,\,t)$ is in the following form:

$$dX_t = \mu(X_t, t)\, dt + \sqrt{D(X_t, t)}\, dW_t. \tag{5.18}$$

Therefore, the equivalent SDE for our FPE in equation 5.15 is given:

$$dX_t = \frac{-\kappa a X_t}{2 \;+\; a \mid X_t \mid}\, dt + \sigma\, dW_t, \tag{5.19}$$

where κ, a and $\sigma > 0$, and dW_t is the increment of a standard Brownian motion.
The SDE in equation 5.19 is clearly mean-reverting since the drift is always the opposite sign to the location of the particle. The drift also increases with the distance of the particle to the origin, although it reaches a limit. Figure 5.4 depicts comparison between simulated probability density functions of X_t for the SDE in equation 5.18 for given parameters. By looking these graphs, we can see that as we decrease a, the distribution approaches to normal and as we increase a , the distribution has higher peak and thinner tails comparing to normal , which is what we would expect.

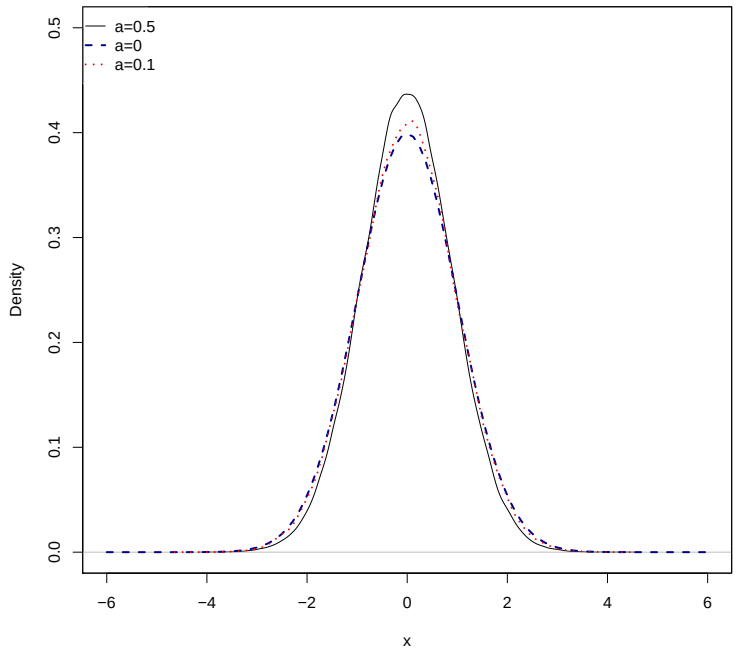

Figure 5.4: Comparison between simulated probability density functions of X_t for the SDE in equation 5.18 for given parameters. Here, for all these three random walks, we assumed $\kappa = 1$, $\sigma = 1$, $\Delta t = 1/252$, $T = 1$ and $100,000$ simulated paths.

Definition 5.5.1. Stationary Solution (Distribution): A unique solution to the FPE 5.17 or its equivalent SDE 5.18 is called a stationary distribution and is denoted by $P_{st}(x)$ if the limiting distribution of X_T as $T \to +\infty$ exists:

$$\lim_{t \to +\infty} P(x\,,\, t) \quad = \quad P_{st}(x). \tag{5.20}$$

We should note that in FPE 5.17 or its equivalent SDE 5.18, if the process has a stationary solution (distribution), $\mu(x,t)$ and $D(x,t)$ are independent of t but not necessarily vice versa.

5.5.2 The Stationary Solution

We attempt to derive a stationary solution for the mean-reverting process defined in equation 5.15 (or equivalently in equation 5.19). Based on the stationary solution definition and from equation 5.15, we have:

$$-\frac{d}{dx}\left[\frac{-\kappa a x}{2+a\mid x\mid}P_{st}(x)\right]+\frac{\sigma^2}{2}\frac{d^2P_{st}(x)}{dx^2}=0. \tag{5.21}$$

Or

$$-\frac{d}{dx}\left[\frac{-\kappa a x}{2+a\mid x\mid}P_{st}(x)-\frac{\sigma^2}{2}\frac{dP_{st}(x)}{dx}\right]=0. \tag{5.22}$$

Based on the boundary conditions, we will have zero flux:

$$\frac{-\kappa a x}{2+a\mid x\mid}P_{st}(x)-\frac{\sigma^2}{2}\frac{dP_{st}(x)}{dx}=c,\qquad \text{with}\quad c=0\,(\text{by applying BC (5.16)}). \tag{5.23}$$

Therefore, we have:

$$\frac{dP_{st}(x)}{P_{st}(x)}=\frac{-2\kappa a x}{\sigma^2(2+a\mid x\mid)}dx. \tag{5.24}$$

Integrating equation 5.24 yields the solution as follows:

$$P_{st}(x)=\begin{cases} c^{-1}\exp\left(\frac{4\kappa\ln(2+ax)}{a\,\sigma^2}-\frac{2\kappa x}{\sigma^2}\right), & \text{if } x\geq 0,\\ c^{-1}\exp\left(\frac{4\kappa\ln(2-ax)}{a\,\sigma^2}+\frac{2\kappa x}{\sigma^2}\right), & \text{otherwise,}\end{cases} \tag{5.25}$$

where c is the normalization constant and equals to:

$$c=2\int_0^\infty \exp\left(\frac{4\kappa\ln(2+ax)}{a\,\sigma^2}-\frac{2\kappa x}{\sigma^2}\right)dx,$$

Using *Mathematica*, we derive c as follows:

$$c=\frac{2^{2+\frac{4\kappa}{a\,\sigma^2}}\,e^{\frac{4\kappa}{a\,\sigma^2}}}{a}E_{\frac{-4\kappa}{a\,\sigma^2}}\left(\frac{4\kappa}{a\,\sigma^2}\right), \tag{5.26}$$

where $E_n(z)$ is the exponential integral function, which is defined as $E_n(z) = \int_1^\infty \frac{e^{-zt}}{t^n} dt$.

Figure 5.5 shows the stationary solutions (distributions) for the MRW given in the SDE form in equation 5.18 in which its derived stationary solution is in equation 5.25. By looking these graphs, we can see that as a increases, the stationary distribution has higher peak and thinner tails. It worth mentioning that the long-run mean and the skewness of the stationary solution 5.25 are both zero.

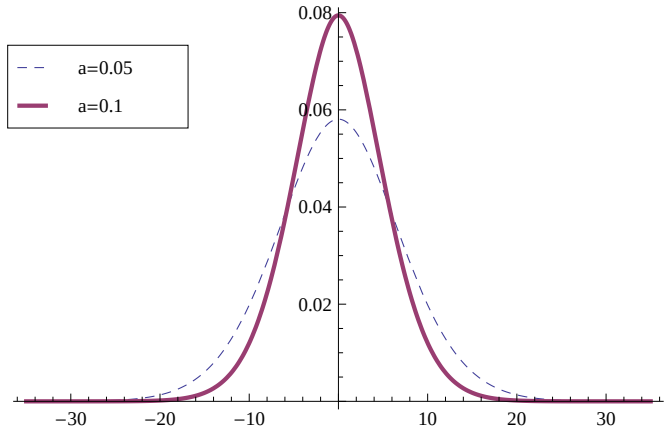

Figure 5.5: The plot depicts the stationary solutions (distributions) for the MRW given in the SDE form in equation 5.18 in which its derived stationary solution is in equation 5.25. By looking these graphs, we can see that as a increases, the stationary distribution has higher peak and thinner tails. Here, for these two stationary solutions, we assumed $\kappa = 1$, $\sigma = 1$.

5.5.3 Analytical Time Dependent Solution

To simplify and drop one of parameters in SDE 5.19, let $Y_t = aX_t$ and apply Ito's lemma to derive:

$$dY_t = \frac{-\alpha Y_t}{2 + \mid Y_t \mid} dt + \sigma^{'} dW_t, \tag{5.27}$$

where $\alpha = a\kappa$ and $\sigma^{'} = a\sigma$. Equivalently, its corresponding PDE is:

$$\frac{\partial P(x\,,\,t)}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{-\alpha x}{2 + \mid x \mid} P(x\,,\,t)\right) + \frac{D}{2}\frac{\partial^2 P(x\,,\,t)}{\partial x^2}, \tag{5.28}$$

where $D = \sigma^{'}$.
We apply the transformation in equation 5.29 to equation 5.28 to pave the way for deriving the time dependent solution as follows:

$$P(x\,,\,t) \;=\; \exp\left(\frac{-2}{D}\,\frac{V(x)}{2}\right)\,q(x\,,\,t), \tag{5.29}$$

where the $V(x)$ is obtained using the stationary solution in equation 5.25 as follows:

$$V(x) = \alpha(\mid x \mid -2\ln(2+\mid x \mid)). \tag{5.30}$$

By applying this transformation, the FPE 5.28 is converted to the Schrödinger type equation as follows:

$$\frac{\partial q(x\,,\,t)}{\partial t} \;=\; -V_s(x)q(x\,,\,t) \;+\; \frac{D}{2}\,\frac{\partial^2 q(x\,,\,t)}{\partial x^2}, \tag{5.31}$$

where the $V_s(x)$ is:

$$V_s(x) = \left(\frac{\alpha x^2}{2D} - 1\right)\frac{\alpha}{(2+\mid x \mid)^2}. \tag{5.32}$$

The PDE 5.31 might be solved using superposition method when the eigenvalues are discrete as follows:

$$q(x\,,\,t) = \sum_{n=0}^{\infty} a_n(0)e^{-\lambda_n t}\psi_n(x), \tag{5.33}$$

where $\lambda_n \geq 0$ and $\psi_n(x)$ are eigenvalues and eigenfunctions respectively, and can be derived solving the following eigenvalues problem:

$$\frac{D}{2}\,\frac{d^2\psi_n(x)}{dx^2} - V_s(x)\psi_n(x) = -\lambda_n\psi_n(x), \tag{5.34}$$

Since $\psi_n(x)$ must be symmetric, we assume $x > 0$. Let $z = \frac{x+2}{a}$ where $a > 0$ is a constant; therefore, we have:

$$\frac{d^2\psi_n(z)}{dz^2} - \left\{\left(-\frac{\alpha^2}{D^2} + \frac{2\lambda_n}{D}\right)a^2 + \frac{\frac{4\alpha^2 a}{D^2}}{z} - \frac{\frac{\alpha(4\alpha-2D)}{D^2}}{z^2}\right\}\psi_n(z) = 0, \tag{5.35}$$

Let $a = \frac{1}{\sqrt{\frac{4\alpha^2}{D^2} - \frac{8\lambda_n}{D}}}$ where $0 \leq \lambda_n < \frac{\alpha^2}{2D}$, $\kappa = \frac{4\alpha^2 a}{D^2}$ and $\mu = \mid \frac{1}{2} - \frac{2\alpha}{D} \mid$. These assumptions will lead the ordinary differential equation (ODE), 5.35 to the following format:

$$\frac{d^2\psi_n(z)}{dz^2} - \left\{-\frac{1}{4} + \frac{\kappa}{z} + \frac{\frac{1}{4} - \mu^2}{z^2}\right\}\psi_n(z) = 0, \tag{5.36}$$

The ODE, 5.36 is known as *Whittaker's equation* and has nontrivial general solutions as follows:

$$C_1 W_{\kappa,\mu}\left(\frac{x+2}{a}\right) + C_2 M_{\kappa,\mu}\left(\frac{x+2}{a}\right), \tag{5.37}$$

where $x >= 0$, and $W_{\kappa,\mu}(z)$ and $M_{\kappa,\mu}(z)$ are Whittaker functions.
For arbitrarily $x \in \Re$, it is easy to show that the general solutions is in the following form:

$$C_1 W_{\kappa,\mu}\left(\frac{|x|+2}{a}\right) + C_2 M_{\kappa,\mu}\left(\frac{|x|+2}{a}\right), \tag{5.38}$$

Based on our boundary conditions given in equations 5.16, only $W_{\kappa,\mu}(z)$ satisfies these boundary conditions all time. Since $W_{\kappa,\mu}(z) = W_{\kappa,-\mu}(z)$, without loss of generality, we can consider $\mu = \frac{2\alpha}{D} - \frac{1}{2}$.
In this section, we attempt to solve this mean-reverting SDE analytically to find transition density; however, we unfortunately failed to solve explicitly this particular stochastic process. The trouble is mostly caused by the absolute value term in the SDE.
Consider the stochastic process of the form as follows:

$$dX_t = f(X_t)\,dt + \sigma\,dW_t, \tag{5.39}$$

Craddock and Dooley (2001) deploy *Lie group symmetries* to classify the SDEs of the form 5.39 based on the drift function $f(x)$ (heat equations) that. They show that have a point symmetry that obtains a constant solution to the fundamental solution. Here, we state their theorem (without proof) as follows:

Theorem 5.5.1. *There exists a point symmetry for the equivalent heat equations of the SDEs in the form of 5.39 taking* **1** *to the fundamental solution if and only if* $f(x)$ *fulfills one of the following Ricatti equations:*

$$\begin{aligned}
f'(x) + \frac{1}{2}f^2(x) &= \frac{A}{x^2} - \frac{B}{2}, \\
f'(x) + \frac{1}{2}f^2(x) &= \frac{x^2}{8} + \frac{C}{x^2}, \\
f'(x) + \frac{1}{2}f^2(x) &= \frac{C}{(x+2)^2}, \\
f'(x) + \frac{1}{2}f^2(x) &= \frac{2}{3}Cx, \\
f'(x) + \frac{1}{2}f^2(x) &= \frac{Cx^2}{2} + D,
\end{aligned} \tag{5.40}$$

where A, B, C, *and* $D \in \Re$ *and arbitrary.*

Unfortunately, the drift term of this new mean-reverting dynamics 5.27 does not satisfy one of the Ricatti equations 5.40; therefore, we could not find explicit solution for this SDE.

5.6 The New Mean-reverting Process versus Vasicek Process

The new mean-reverting process, 5.27 and Vasicek, 5.1 both have the mean-reverting property that is crucial to model a spread process. We need to compare them to see which one is more appropriate to model a commodity spread process. Since the long-run mean in a spread process does not have to be zero, we generalize this mean-reverting dynamic, 5.27 as follows:

$$dX_t = \kappa \left(\mu - \frac{X_t}{2 + \mid X_t \mid} \right) dt + \sigma \, dW_t, \tag{5.41}$$

where

μ is a real number and $-1 < \mu < 1$,
σ is the volatility of the process,
κ is the mean reversion rate (speed),
dW_t is the increment of a standard Brownian motion.

One should notice that unlike the Vasicek process 5.1, in this new process 5.41, the long-run mean is not μ. We will derive the long-run mean later in this section. We should emphasize that when $|\mu| \geq 1$, the process is no longer mean-reverting. This nonlinear SDE does not have analytical transition density; therefore, we are limited to apply a numerical method to estimate parameters. Here, we will deploy the new local linearization method introduced by Shoji and Ozaki (1998), which we discuss in section 2.4.1. The equivalent Vasicek process for this process is given in equation 5.1. In these two stochastic processes, when the process tends to deviate away from their long-run means, their drift functions impose force to revert back to long-run mean quite differently. Figure 5.9 depicts their drift functions. By looking these graphs, we can see that in the Vasicek model as the process attempts to deviate from the long-term mean μ, the drift linearly increases the force to revert back the process to mean μ; however, in the MRW model as the process diverges away from its long-run mean, up to certain ranges of x, the amount of force approximately increases linear, then asymptotically becomes constant. This suggests that this new mean-reverting process has more chance to stay away from the long-run mean for a longer time; in other words, it has relatively heavier tails, which is usually a property of financial processes.

To comparison these two processes, we simulate $10,000$ paths for both processes with identical parameters and random values. We also simulate Brownian motion. In each

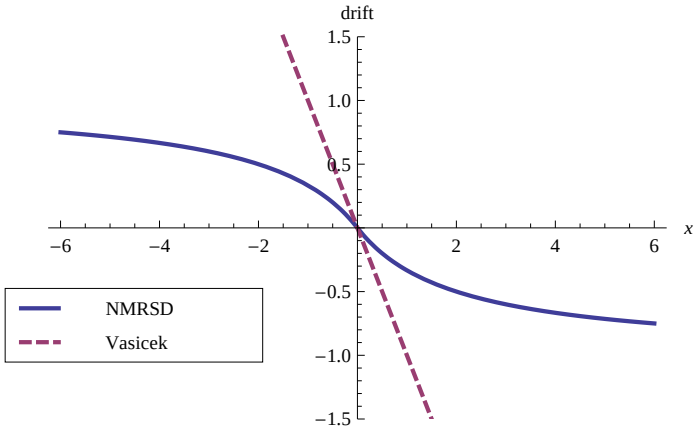

Figure 5.6: The plot depicts the drift functions for the new mean-reverting process given in equation 5.41 and the Vasicek process given in equation 5.1. We assumed $\kappa = 1$, $\mu = 0$ for both models.

step, we record the empirical variance for all three processes. The constructed variance paths are shown in figure 5.7, plot (a). The empirical densities are depicted in the same figure, plot (b).

The new generalized process 5.41 has another crucial advantage over the Vasicek process when its long-run mean is not zero. In this case, the process can generate skewness which is another well-known property of financial processes. To show this interesting property and compare with the Vasicek process, we simulate $10,000$ paths for both processes for five years and plot empirical densities shown in figure 5.8. Figure 5.8 clearly depicts this important property of this new mean-reverting process. When μ is negative, the process is skewed to the left and vice versa. Since the reverting force asymptotically becomes constant and no longer increases when the spread deviate enough far from the long-run mean. Therefore, the long-run mean partially plays as a role of drift similar to Brownian motion with drift. Although we derived the stationary solution of this new process in section 5.5.2 when $\mu = 0$, we rederive the stationary distribution due to fundamental changes in the process as follows:

$$P_{st}(x) = \begin{cases} c^{-1} \exp\left(2\delta\left[2\ln(2+x) + (\mu-1)x\right]\right), & \text{if } x \geq 0, \\ c^{-1} \exp\left(2\delta\left[2\ln(2-x) + (\mu+1)x\right]\right), & \text{otherwise,} \end{cases} \tag{5.42}$$

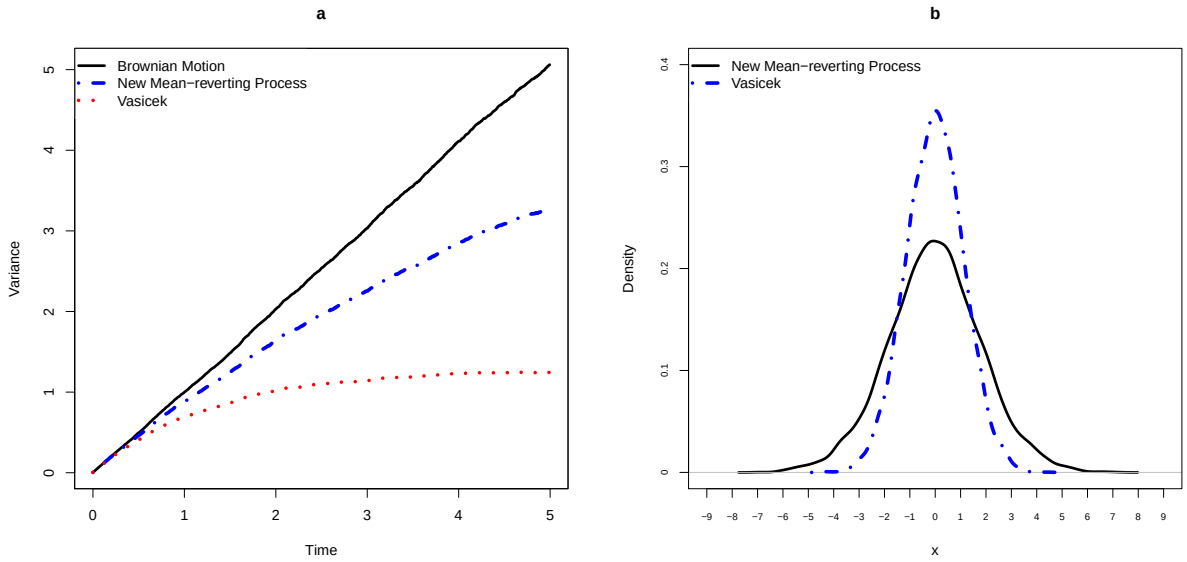

Figure 5.7: Time plots of variance and empirical densities for 10,000 simulated paths: Plot (a) shows how empirical variances evolve through the time in these three models. Plot (b) shows comparison between empirical densities in these two models after five years. We assumed $\kappa = 0.4$, $\mu = 0$, $\sigma = 1$, $\Delta t = \frac{1}{252}$, and $T = 5$ for both models by using identical random generated numbers.

where c is the normalization constant and equals to:

$$c = \int_{-\infty}^{0} \exp\left(2\delta\left[2\ln(2-x) + (\mu+1)x\right]\right) dx + \int_{0}^{\infty} \exp\left(2\delta\left[2\ln(2+x) + (\mu-1)x\right]\right) dx,$$

Using *Mathematica*, we derive the c as follows:

$$c = 2^{1+4\delta}\, e^{-4\delta(-1+\mu)} \left\{E_1 + e^{8\delta\mu} E_2\right\}, \tag{5.43}$$

We also derive the mean (long-run mean), the variance and the skewness of the stationary

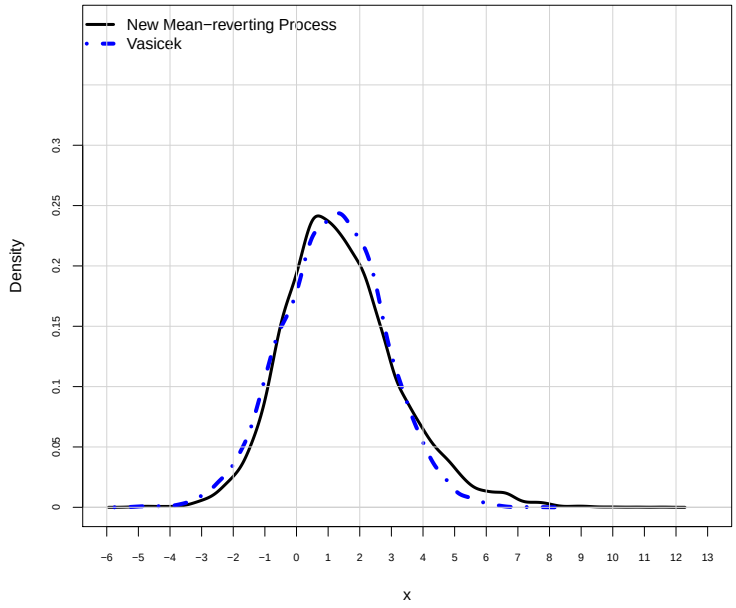

Figure 5.8: The plot depicts comparison between empirical densities in these two models after five years using 10000 simulated paths. We assumed $\kappa = 4$, $\mu = 0.3$, $\sigma = 2.1$ for this new process 5.41 and we also assumed $\kappa = 0.86$, $\mu = 1.2$, $\sigma = 2.1$ for the Vasicek process 5.1. for both models $\Delta t = \frac{1}{252}$, and $T = 5$ are considered by using identical random generated numbers.

distribution as follows:

$$E[X] = \frac{-2\mu\sigma^2 e^{4\delta(\mu-1)} - (\mu+1)(4\kappa\mu+\sigma^2)E_1 + (\mu-1)(4\kappa\mu-\sigma^2)e^{8\delta\mu}E_2}{2\kappa(\mu^2-1)\left\{E_1 + e^{8\delta\mu}E_2\right\}}, \quad (5.44)$$

$$\begin{aligned} E[X^2] = {} & e^{8\delta\mu}\frac{2^{-1-8\delta}}{(\mu+1)^3}(\delta(\mu+1))^{-4\delta}(8\kappa^2\mu^2 + 2\kappa(1-2\mu)\sigma^2 + \sigma^4) \times \\ & \left\{\frac{(32\kappa^3 + 48\kappa^2\sigma^2 + 22\kappa\sigma^4 + 3\sigma^6)\Gamma(4\delta)}{\kappa^2(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1 + e^{8\delta\mu}E_2\right\}}\right\} - \\ & \frac{4\sigma^6 e^{8\delta\mu}\ \mathbf{M}\left[1+4\delta, 4+4\delta, -4\delta(\mu+1)\right]}{(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1 + e^{8\delta\mu}E_2\right\}} - \\ & 2^{-1-8\delta}\frac{1}{(\mu-1)^3}(\delta(1-\mu))^{-4\delta}(8\kappa^2\mu^2 + 2\kappa(1+2\mu)\sigma^2 + \sigma^4) \times \\ & \left\{\frac{(32\kappa^3 + 48\kappa^2\sigma^2 + 22\kappa\sigma^4 + 3\sigma^6)\Gamma(4\delta)}{\kappa^2(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1 + e^{8\delta\mu}E_2\right\}}\right\} - \\ & \frac{4\sigma^6\ \mathbf{M}\left[1+4\delta, 4+4\delta, 4\delta(\mu-1)\right]}{(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1 + e^{8\delta\mu}E_2\right\}}, \end{aligned} \quad (5.45)$$

$$
\begin{aligned}
E[X^3] &= 2^{-2-8\delta}\,(\delta(1-\mu))^{-4\delta}(32\kappa^3\mu^3+24\kappa^2\mu(\mu+1)\sigma^2+2\kappa(6\mu+5)\sigma^4+3\sigma^6)\times \\
&\left\{\frac{(32\kappa^4+80\kappa^3\sigma^2+70\kappa^2\sigma^4+25\kappa\sigma^6+3\sigma^8)\Gamma(4\delta)}{\kappa^3(\mu-1)^4(\kappa+\sigma^2)(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1+e^{8\delta\mu}E_2\right\}}\right\}+ \\
&\frac{6\sigma^8\ \mathbf{M}\left[1+4\delta,5+4\delta,4\delta(\mu-1)\right]}{(\kappa+\sigma^2)(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1+e^{8\delta\mu}E_2\right\}}+ \\
&2^{-2-8\delta}\,e^{8\delta\mu}(\delta(1+\mu))^{-4\delta}(32\kappa^3\mu^3-24\kappa^2\mu(\mu-1)\sigma^2+2\kappa(6\mu-5)\sigma^4-3\sigma^6)\times \\
&\left\{\frac{(32\kappa^4+80\kappa^3\sigma^2+70\kappa^2\sigma^4+25\kappa\sigma^6+3\sigma^8)\Gamma(4\delta)}{\kappa^3(\mu+1)^4(\kappa+\sigma^2)(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1+e^{8\delta\mu}E_2\right\}}\right\}- \\
&\frac{6\times e^{8\delta\mu}\sigma^8\ \mathbf{M}\left[1+4\delta,5+4\delta,-4\delta(\mu+1)\right]}{(\kappa+\sigma^2)(2\kappa+\sigma^2)(4\kappa+\sigma^2)(4\kappa+3\sigma^2)\left\{E_1+e^{8\delta\mu}E_2\right\}}, \qquad (5.46)\\
Var(X) &= E[X^2]-(E[X])^2, \qquad (5.47)\\
\boldsymbol{\gamma_1} &= \frac{E[X^3]-3E[x]Var(X)-(E[X])^3}{(Var(X))^{\frac{3}{2}}}, \qquad (5.48)
\end{aligned}
$$

where
$\delta=\frac{\kappa}{\sigma^2}$, $E_1=E_{\_4\delta}\left(-4\delta(\mu-1)\right)$, $E_2=E_{\_4\delta}\left(4\delta(\mu+1)\right)$,
$\mathbf{M}[a,b,z]$ is the *Kummer confluent hypergeometric* function defined as:

$$\mathbf{M}[a,b,z]=\sum_{n=0}^{\infty}\frac{(a)_n\,2^n}{(b)_n\,n!},$$

$(a)_0=1$, $(a)_n=a(a+1)(a+2)\ldots(a+n-1)$,
$\Gamma(z)$ is the gamma function and defined as $\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}dt$
$\boldsymbol{\gamma_1}$ is the skewness of the X,
$E_n(z)$ is the exponential integral function, which is defined as $E_n(z)=\int_1^\infty\frac{e^{-zt}}{t^n}dt$.

5.7 Parameter Estimation:

As discussed in section 5.5.3, this mean-reverting process is a nonlinear SDE with a unique but unknown transition density. As a results, we must apply a numerical or approximation method. Here, we apply the new local linearization method (NLLM), explained in detail in chapter 2, to estimate the model parameters.

Since the first and second derivatives of the drift function $h(x, \boldsymbol{\theta}) = \kappa\left(\mu - \frac{x}{2+|x|}\right)$ for this mean-reverting diffusion 5.41 do not exist at zero, as they violate conditions of equation 2.49, this algorithm cannot be applied without some treatments. Fortunately, the discontinuity of the first derivative at zero can be removed. So the treated first derivative is $h'(x, \boldsymbol{\theta}) = \frac{-2\kappa}{(2+|x|)^2}$. The second derivative is $h''(x, \boldsymbol{\theta}) = \frac{4\kappa x}{|x|(2+|x|)^3}$ which clearly shows that second derivative has a vertical asymptote at zero. To resolve this issue, we take a viscosity-like approach and redefine the second derivative as follows:

$$h''(x, \boldsymbol{\theta}) = \begin{cases} 0 & \text{if } -\varepsilon < x < \varepsilon, \\ \frac{4\kappa x}{|x|(2+|x|)^3} & \text{otherwise,} \end{cases} \tag{5.49}$$

where ε is a small positive value.
One can see that in this mean-reverting process, since the first derivative is always positive, when time approaches infinity, the variance approaches to $\lim_{t\to\infty} Var[X_t \,|\, X_s] = \frac{-\sigma^2}{2L_s}$.
There is also another important fact that we want to express here. Any function that has similar behavior and looks (see figure 5.9) like the drift term of this mean-reverting process can generate identical results and identical properties. Here, we list a few alternative processes, having drifts which are twice continuously differentiable functions with respect to X_t:

$$dX_t = \kappa\left(\mu - \frac{X_t}{\sqrt{2+X_t^2}}\right) dt + \sigma\, dW_t, \tag{5.50}$$

$$dX_t = \kappa\left(\mu - \frac{e^{X_t}-1}{e^{X_t}+1}\right) dt + \sigma\, dW_t, \tag{5.51}$$

$$dX_t = \kappa\left(\mu - \arctan(X_t)\right) dt + \sigma\, dW_t. \tag{5.52}$$

One should note that since the Vasicek process is a linear process, NLLM method will calculate the analytical transition density (normal distribution) and the pseudo-maximum likelihood will turn into exact maximum likelihood method to estimate parameters of the model.

5.8 Empirical Results

The spot spread between WTI and WTS crude oils is considered from January 2000 to January 2013 in this study, as empirically analyzed in section 5.2. Here we intend to compare these two models, the Vasicek and the new mean-revering processes and show which one can better explain the reality of the spread processes. We apply the new local

estimation method (2.4.1) to estimate the parameters of the Vasicek and the new mean-reverting models for the daily spread between WTI and WTS crude oils from January 2000 to January 2013. We will also consider time transformation (time scale) to the estimated parameters. In other words, if we consider $s = f(t) = vt$, the resulting process will be identical to the original process. For example, the time transformed process for our mean-reverting process 5.41 will be as follows:

$$dX_s = \frac{\kappa}{v}\left(\mu - \frac{X_s}{2+\mid X_s \mid}\right) dt + \frac{\sigma}{\sqrt{v}}\, dW_s. \tag{5.53}$$

One can see that the time transformation is nothing but scaling κ and σ in a way that guaranteed to generate identical process. We apply the new local linearization method to estimate parameters and summarize the estimation results of the data fitted both to the Vasicek and the new mean-reverting dynamics and their time scaled diffusions in table 5.3. Figure 5.9 depicts the fitted drift functions in both models: the new mean-reverting process and the Vasicek process using the time scaled estimated parameters for κ, and μ that are summarized in table 5.3. The plot shows that these drift functions impose force to revert back to long-run mean quite differently in these two models.

Table 5.3: Estimated parameters for two processes: new mean-reverting and Vasicek one-factor models using daily spread between WTI and WTS oils from January 2000 to January 2013 and comparison between them. We also apply estimation method by scaling the spread empirical data in two different ways namely scale a day as a day and scale a day as a month.

Model	Scale	κ	μ	σ	$\frac{\sigma^2}{2\kappa}$	LRM^1	-2 $logl^2$	AIC^3	BIC^4
New MR	No Scale	117.23	0.564	7.58	0.245	3.147	3961.5	3967.5	3977.5
Process	$s = \frac{250}{12}t$	5.6273	0.564	1.662	0.245	3.147	3961.5	3967.5	3977.5
Vasicek	No Scale	9.092	3.143	7.548	3.133	3.143	3968.5	3974.5	3984.5
Process	$s = \frac{250}{12}t$	0.4364	3.143	1.654	3.133	3.143	3968.5	3974.5	3984.5

Note: 1-Long-run Mean, 2-Log-likelihood, 3-Akaike information criterion, 4- Bayesian information criterion.

Using the fitted parameters (summarized in table 5.3) for both Vasicek and the new mean-reverting processes with their time scaled dynamics, we simulated 10,000 paths for four fitted processes with $\Delta t = \frac{1}{252}$, and $T = 15$ are considered using identical random generated numbers. The simulations results after 15 years, and empirical observations are summarized in table 5.4. Figure 5.10 depicts the empirical distributions using the fitted parameters (summarized in table 5.3) for both the Vasicek and the new mean-reverting processes with their scaled diffusions. The left plot compares the results for the fitted with its time scaled for the new mean-reverting process and right plot compares for the

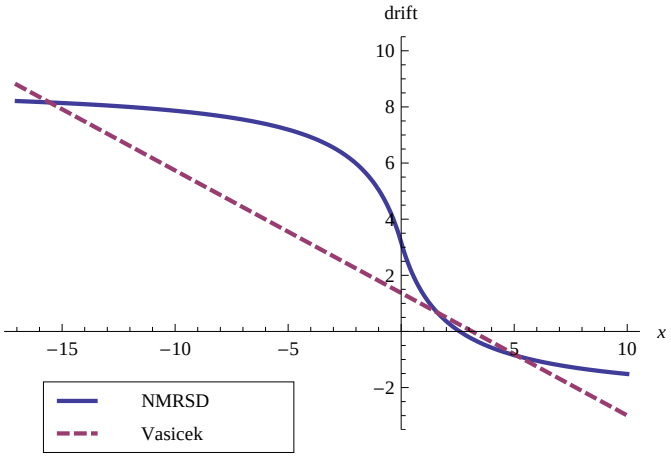

Figure 5.9: The plot depicts the fitted drift functions for the new mean-reverting process and the Vasicek process using the time scaled estimated parameters for κ, and μ that are summarized in table 5.3.

results for the fitted with its time scaled variant for the Vasicek process. Observation results for both processes confirm that the fitted and its scaled process for both models are identical and we can choose upon our preference and logic. We can also observe that the stationary distributions for the fitted and its time scaled diffusions in both models are exactly the same.

Table 5.4: The simulations results for the fitted parameters (summarized in table 5.3) for both the Vasicek and the new mean-reverting processes with time scaled processes, we simulated $10,000$ paths for four fitted processes where the $\Delta t = \frac{1}{252}$, and $T = 15$ are considered using identical random generated numbers with 21745 as a seed. The last row is the summarized information of the observed dataset (the daily spread between WTI and WTS from January 2000 to January 2013).

Simulated by	T. Scale	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skew.
New MR	No Scale	-2.192	1.910	2.953	3.145	4.169	12.420	0.70
Process	$s = \frac{250}{12}t$	-1.517	1.908	2.932	3.124	4.173	13.930	0.66
Vasicek	No Scale	-4.268	1.906	3.115	3.122	4.333	9.340	0.03
Process	$s = \frac{250}{12}t$	-3.920	1.938	3.101	3.113	4.329	9.276	0.00
Observed D.	Daily	-0.540	1.840	2.700	3.146	4.300	10.750	0.87

We conclude that estimation results comes from the original and the time scaled

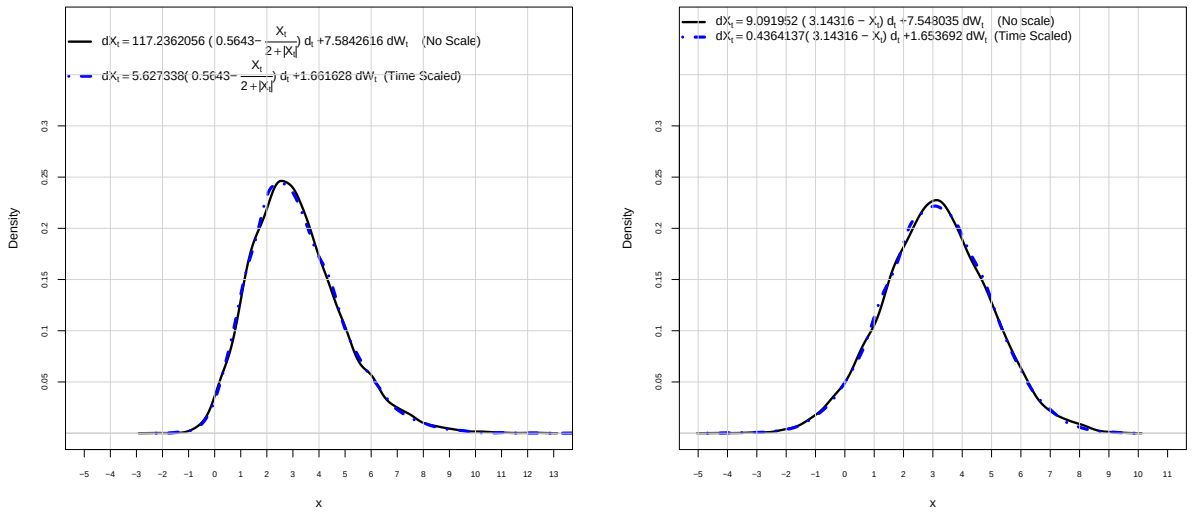

Figure 5.10: The empirical distributions using the fitted parameters (summarized in table 5.3) for both the Vasicek and the new mean-reverting processes with their time scaled dynamics. The left plot compares the results for the fitted with its time scaled for the new mean-reverting process and right plot compares for the results for the fitted with its time scaled for Vasicek process. We simulated $10,000$ paths for four fitted processes with $\Delta t = \frac{1}{252}$, and $T = 15$ are considered using identical random generated numbers with 21745 as a seed.

results are identical to the fitted process and can be deployed interchangeably. However, we prefer the time scaled estimated parameters because the fitted mean-reversion speeds in both processes give misleading impression that mean-reverting feature of the observed spread is quite strong. Figure 5.1 does not illustrate a high mean-reversion property. Here high mean-reversion speed is merely to offset the their higher volatilities. As a result, the fitted processes with time scaled make more sense.

5.8.1 Goodness of Fit

We checked whether this novel stochastic process can explain the reality of the spread process better than the Vasicek process or not. We summarized the estimation results with the tests results for goodness of fit in table 5.3. As we can see both AIC and BIC values suggest superiority of our mean-reverting model over the Vasicek model. Figure 5.11 depicts the comparisons results for the fitted models for both models (estimated parameters are summarized in table 5.3): The left plot depicts comparison between simulated empirical densities in the Vasicek and the new mean-reverting after fifteen

years using 10,000 simulated paths. The right plot shows the empirical density for the observed daily spread between WTI and WTS crude oil data, the fitted stationary distributions for these models. Both left; right subplots of figure 5.11 clearly confirm our claims that this new diffusion can capture the most important characteristics of the observed data including skewness, kurtosis, and heavy tails, which are crucial attributes for an appropriate model.

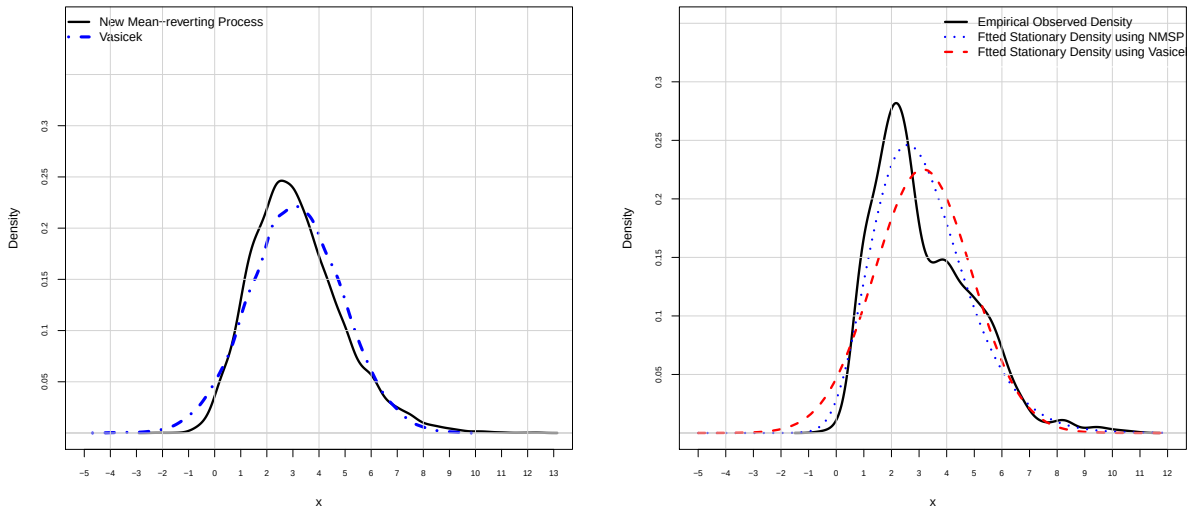

Figure 5.11: The left plot depicts comparison between simulated empirical densities in these two models: Vasicek and the new mean-reverting after fifteen years using 10,000 simulated paths using the fitted parameters (summarized in table 5.3). The right plot shows the comparison between the empirical density for the observed daily spread between WTI and WTS crude oil data with the fitted stationary distributions of these two models. We assumed that the $\Delta t = \frac{1}{252}$, and $T = 15$ using identical random generated numbers with 21745 as a seed for the simulation.

5.9 Conclusion:

In this chapter, we introduced a new mean-reverting random walk, and derive its continuous time limit. Its stationary distribution was derived, but all attempts to solve this nonlinear dynamics analytically for transition density unfortunately failed. We also generalized this new mean-reverting process to equip the diffusion to capture possible skewness of the spread process. We compare this one-factor model to Elliott *et al.* (2005)'s one-factor model and using simulation results, we showed that this new mean-reverting

one-factor model has capability to capture the potential heavy tail or tails, kurtosis, and skewness of the spread processes. To estimate parameters of this nonlinear dynamics, the new local linearization method was deployed. We applied this new SDE and Vasicek processes to the daily spread between WTI and WTS crude oils from January 2000 to January 2013. The maximum log-likelihood value, AIC, BIC all endorsed that this new SDE outperformed compare to Vasicek process. In chapter 6, we will introduce optimal trading strategies for a spread process in which we assume the spread process follows models discussed in chapter 4 and this chapter.

Chapter 6

TRADING STRATEGIES FOR THE SPREAD PROCESSES

6.1 Introduction

The main concept behind any pairs trading strategy relies on fluctuations about the equilibrium level (long-run mean) of the spread process. In other words, the key property of spread process is mean-reversion. We attempt to take advantage of mean-reversion property of the spread process while we consider realistic features of the tradable spread process. Each of the methods to model spread processes explained earlier in this thesis can be applicable for some particular cases, and the model is chosen based on the behavior of the underlying assets or commodities in the spread portfolio. Therefore, a spread trading strategy should build upon the method that we choose to price the spread process. As we explained in chapter 4, we briefly review three essential steps involved in any pairs trading strategies as follows:

1. Identifying assets (stocks) or commodities pairs that could potentially share a common non-stationary trend (stochastic trend). This process can be proceeded using the economical and fundamental information (sharing similar risk factors). In other words, the potential pairs are chosen when they are showing certain characteristics of comovement, arising from common fundamental drivers. Vidyamurthy (2004) lays out a method to choose potential pairs namely the *score/distance measure* in the stock markets.

2. Applying statistical methods on historical data, we should verify existence of cointegration or long-term relationship (mean-reversion) in the potential pairs. We should also estimate the cointegration coefficient β to build the spread portfolio. One should notice that in most cases, we confront pairs that are not precisely cointegrated and differ from ideal cointegration conditions even though the pair is still

tradable. Therefore, we should select an appropriate method to model the spread process by investigating empirical evidence.

3. Once we construct the spread portfolio (the linear combination of the underlying assets or commodities) and choose a proper method to model the spread process, we need to specify when we should consider placing a trade. When the spread deviates sufficient enough from the equilibrium level, we enter a trade in the direction hoping that the temporary deviation form the long-run mean will revert back and we can unwind the trade. It is important to identify what are the specifications of substantial deviation from the long-run mean to decide when to enter into a trade. We also need to specify upper and lower optimal threshold or thresholds (in case of multiple thresholds) in the intended spread dynamics. We attempt to define set of simple trading rules known as *trading signals* in proper fashion based on *maximization of profits*. Finally we need to identify when we should unwind the trade position and realize profit.

In this chapter, we attempt to specify some trading strategies for different classes of the spread dynamics and answer the questions in step 3. This chapter is organized as follows: In section 6.2, we empirically illustrate an existing trading strategy for the cointegration approach. We explain how two different trading strategies can be implemented for Elliott *et al.* (2005)'s one-factor model in section 6.3. In section 6.5, we first review the optimal trading strategies for the cointegration approach. Second we investigate the optimal trading strategies introduced by Bertram (2010) in which he exploits *first passage time results* for the Ornstein-Uhlenbeck (OU) process to find optimal upper and lower boundaries. We propose to deploy the stationary distribution to obtain optimal barriers in the Vasicek dynamics for which is appropriate for the long-term investment strategy. Finally, we introduce two new trading strategies for the spread process of WTI crude oil and Brent oil by considering the empirical facts and behavioral process change.

6.2 Trading Strategy for Cointegration Approach

The fundamental notion behind this strategy is the mean reversion property of the spread series. Here we assume that first the tradable pair is chosen. Second the cointegration tests and economical and fundamental drivers reviewed in chapter 4 confirm the existence of cointegration in the specified pair. Finally we also identified the long-run mean, μ_r, the cointegration coefficient, β and constructed the spread series. The residual series, r_t ($r_t = p_t^A - \beta p_t^B - \mu_r$, where p_t^A and p_t^B are the underlying assets or commodities prices or log prices depending on the model) must be a stationary $ARMA$ series model. We fit $ARMA$ to the sample residual series using the Kalman filter, estimate the model parameters including the volatility σ_r. Consider η be the trading costs of executing a pairs

trade which are bid-ask spreads of two underlying assets or commodities components of the transaction costs are known as *slippage*, marginal interest rates, and commissions. Let also Δ be the optimal deviation from the equilibrium level, μ_r when we decide to open a new position or unwind the existing position. One obvious condition for executing a trade is $2\,\Delta > \eta$. We attempt to determine a value for Δ so as to maximize the profits. The profitability of the trades can be measured by mean-reverting rate, which here is related to the *zero-crossing rate*. When the zero-crossing rate is high, the process is quick to revert back to its equilibrium level. This implies that the holding time of a paired position is significantly short and the frequency of trades is high. Note that we can deploy Rice's formula to estimate the zero-crossing rate for r_t after estimating the stationary $ARMA$ parameters. We can establish the optimal upper and lower barriers (in this case, since r_t is a symmetric process, the upper and lower barriers are $\mu_r + \Delta$ and $\mu_r - \Delta$ respectively) by maximizing profit (expected return) while minimizing the risk strategies in terms of volatility. Assume the optimal Δ is estimated. Then a simple trading strategy can be as follows:

- When the spread process has diverged Δ from the long-run mean, μ_r, we place an appropriate position in the two underlying assets or commodities based on the cointegration coefficient β (at time t, when $p_t^A - \beta p_t^B = \mu_r \pm \Delta$,we sell (buy) one unit of A and simultaneously buy (sell) β units of B),
- Wait until the mispricing has been corrected. Depending on the spread series, we can either unwind the position at time T $(T > t)$, when $p_t^A - \beta p_t^B = \mu_r$ with profit $\Delta - \eta$ or unwind the position at time T $(T > t)$, when $p_t^A - \beta p_t^B = \mu_r \mp \Delta$ with profit of $2\Delta - \eta$ and enter into a new position in the opposite direction.

Notice that we assume that any given time we hold at most one paired trade and we do not open a new position unless we close the previous trade.

6.2.1 Empirical Demonstration

To illustrate the cointegration strategy, we consider the spread process between WTI crude oil, P_t^C and Brent oil, P_t^B for daily price of front contracts from April 1994 to January 2005. In chapter 4, we showed that WTI crude oil and Brent oil are cointegrated in our sample data. We again apply Engle-Granger two step test. In the two step test, first we apply simple linear regression: $P_t^B = \alpha + \beta P_t^C + r_t$, where r_t is the residual series. The fitted model is:

$$P_t^B = 0.932\, P_t^C + \hat{r}_t, \text{ where } \sigma_r = 0.816.$$

Figure 6.1 depicts the residual time series, $\hat{r}_t$. The plot illustrates that the residual series is indeed a stationary process and shows long-run relationship between WTI and Brent oils series. We also apply ADF to test for the unit root in the residual series.

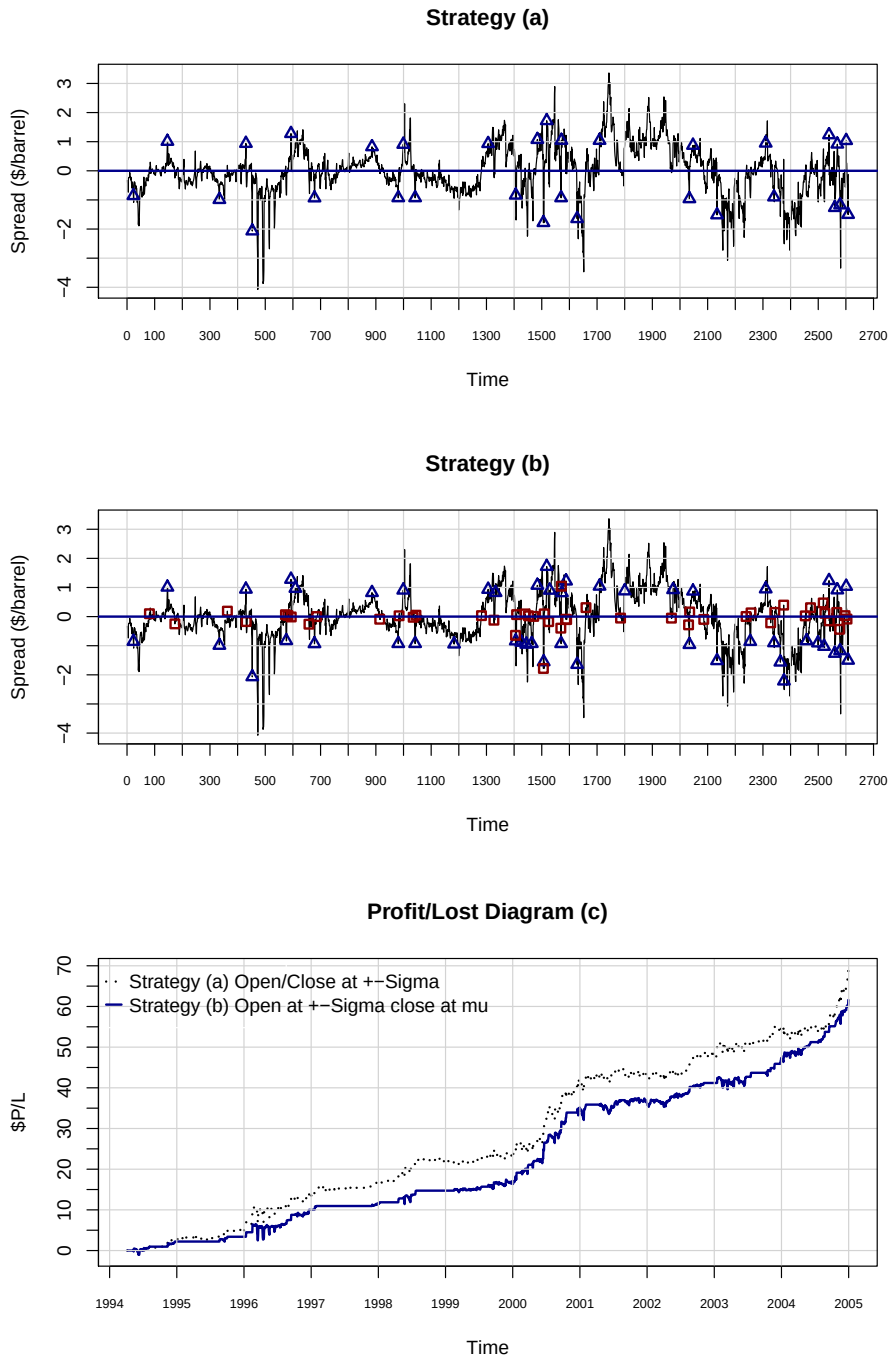

Figure 6.1: Time plot of the estimated spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in cointegration approach: Plot (a) shows the series of trades in strategy (a) in which whenever we hit the lower barrier, $\ell_1 = \mu - \Delta = 0 - 0.816$ or upper barrier, $\ell_2 = \mu + \Delta = 0 + 0.816$, we close existing paired trade and open new position in opposite direction. Plot (b) shows the series of trades in strategy (b) in which whenever we hit the lower barrier, $\ell_1 = \mu - \Delta = 0 - 0.816$ or upper barrier, $\ell_2 = \mu + \Delta = 0 + 0.816$, we only open a new position in proper direction and we wait until to revert to long-run mean, 0 when we unwind existing paired trade. Plot (c) depict how profit/loss growth is both strategies.

The test indicates that the null "unit root" hypothesis for the residual series should be rejected in favor of the alternative hypothesis at one percent level of significance. Moreover, we fit ARMA(1,1) to $\hat{r}_t$ using the *FitARMA* package in the R programming language written by McLeod and we derive:

$$\hat{r}_t = 0.9353\,\hat{r}_{t-1} + \varepsilon_t - 0.1103\,\varepsilon_{t-1}.$$

6.2.1.1 Illustration of Simple Trading Strategy:

Although the optimal Δ can be derived using *the expected first passage time* for the estimated $ARMA(p,q)$ model for the spread series, to illustrate how the algorithm works, we choose Δ to be the volatility of the spread process, $\hat{r}_t$. Therefore, $\Delta = \sigma_r = 0.816$. Figure 6.1 demonstrates the time plot of the spread series $\hat{r}_t$. In figure 6.1, strategy (a), each triangle represents the time when the existing paired trade is closed with profit of at least $2\,\Delta = \$1.632$ (assuming the transition cost is zero) and enter into a new paired trade in the opposite direction of closed position. In figure 6.1, strategy (b), each triangle also represents the time that we decide to open a new paired trade in appropriate direction and square represents the time that we decide to close the only existing position with a profit of at least $\Delta = \$0.816$ (assuming the transition cost is zero). In figure 6.1, profit/loss (c), we also depict how the profit or loss in both strategies, (a) and (b) evolves through the specified time frame. We summarize the trades with respect to two different strategies (a) and (b) in our sample data in table 6.1. As we can see, for this data set, strategy (a) has better performance than strategy (b) in two aspects: first it has higher profit and second it has lower number of trades meaning that it has lower transaction costs. We would like to come up with a rigorous method to select optimal Δ and optimal closing time for the trades.

Table 6.1: Summary of trades in two different strategies (a) & (b) shown in figure 6.1 in the spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in cointegration approach.

Strategy (a)				Strategy (b)			
Number of trades	Profit per trade	Expected total profit	Total profit	Number of trades	Profit per trade	Expected total profit	Total profit
31	\$1.632	\$48.96	\$67.60	48	\$0.816	\$38.35	\$60.22

6.3 Trading Strategy for One-factor Spread Process

This trading strategy assumes that the spread process follows the Vasicek process, proposed by Elliott *et al.* (2005), and explained in detail in section 4.3 of this thesis. Notice

that first, we assume that the linear combination of underlying assets or commodities to form the spread process is already estimated and known. We also estimated the parameters of the model using the sample data (the algorithm was explained in chapter 4). As we explained before, the spot spread process, X_t is assumed to follow the Vasicek stochastic process as follows:

$$dX_t = a(\, b - X_t \,)\, dt + \sigma dW_t \tag{6.1}$$

where the model assumes that the linear combination X_t evolves as an Ornstein-Uhlenbeck process with constant coefficients, $a > 0$ is mean-reversion rate, b is the long-run mean of the spread, σ is the volatility of the process, and dW_t is the increment of a standard Brownian motion.

Let $-\Delta$ and Δ ($\Delta > 0$) be the lower and upper deviations from the equilibrium level b where we decide to open a new position, unwind the existing position, or both. Therefore, the lower barrier is $\ell_1 = b - \Delta$ and the upper barrier is $\ell_1 = b + \Delta$. We attempt to determine Δ so as to maximize the profits of the trades. The profitability of the trades can be measured by the mean-reversion rate a which here is related to the *zero-crossing rate* and the location of Δ. When the zero-crossing rate is high, the process quickly reverts back to the equilibrium level b meaning the holding time of a paired position is significantly short and frequency of trades is high. Now, we need to define notations that are required here. For the stochastic process X_t, *the first passage time* τ_x is defined as $\tau_x = \inf\{t \geq 0 \,:\, X_t = x \mid X_0 = x_0\}$ where if X_t never hits x, τ_x is assigned a value of infinity ($\tau_x = \infty$). We also define the first exit time from interval $[\ell_1\,,\, \ell_2]$ as follows:

$$\tau = \inf\{t \geq 0 \,:\, X_t = \ell_1 \,\text{or}\, \ell_2 \mid X_0 = \ell_0 \text{ and } \ell_0 \epsilon [\ell_1\,,\, \ell_2]\,\}$$

Let τ_Δ be the expected first passage time for X_t starting at ℓ_1 and ending at ℓ_2. Since X_t is symmetric, the expected passage time for starting at ℓ_2 and ending at ℓ_1 will be the same as τ_Δ. The rate R at which the trades deliver is $R = \frac{2\Delta}{\tau_\Delta}$. We select Δ to maximize this rate. Assume the optimal Δ is estimated. Then a simple trading strategy can be as follows:

- When the spread process has diverged Δ from the long-run mean, b, we place an appropriate position in the two underlying assets or commodities based on the linear combination coefficient, β,
- Wait for the mispricing to be corrected. Depending on the spread series, we can either unwind the position at time T ($T > t$), when the spread process reverts back to the long-run mean b with profit $\Delta - \eta$, η is the transaction costs of carrying out of the paired trade, or unwind the position at time T ($T > t$), when the spread process hit the barrier in the opposite direction of previous hit with profit of $2\Delta - \eta$ and enter into a new position in the opposite direction.

Note that we assume that we only hold at most one paired trade at any given time in both strategies, and we do not open new position unless we close the previous trade.

6.3.1 Empirical Demonstration

To illustrate one-factor strategy, we consider the location spread process between WTI crude oil, P_t^C and Brent oil, P_t^B for daily price of five contracts from April 1994 to January 2005 where the involved future contracts are the 1,3,6,9 and 12 months to maturities. In chapter 4, we explained the model parameters's estimation methodology. We fit the model for location spreads between WTI and Brent oils. Figure 6.2 depicts the daily front contract future prices of location spread process between WTI and Brent oils. Table 6.2 lists the estimation results for the one factor model using our sample data.

Table 6.2: Estimated parameters for one-factor model using future contracts of 1,3,6,9 and 12 months to maturities (daily location spread between WTI and Brent oils from April 1994 to January 2005).

	a	σ	b	λ	ζ
value	1.158	0.843	1.649	0.467	0.281
Sd.Err	0.0174	0.0198	0.0498	0.0510	0.0142

6.3.1.1 Illustration of Simple Trading Strategy:

As we mentioned prior in this section, $-\Delta$ and Δ ($\Delta > 0$) is the lower and upper deviations from the equilibrium level b where we decide to open a new position, unwind the existing position or both. Although the optimal Δ can be derived using *the expected first passage time* for the estimated one-factor model for the location spread process (Vasicek), to illustrate how the algorithm works, we consider Δ to be the estimated volatility of the location spread process, $\hat{\sigma}$. Therefore, $\Delta = \hat{\sigma} = 0.843$. Figure 6.2 shows the time plot of the location spread process for the front contract. In figure 6.2, strategy (a), each triangle represents the time when the existing paired trade is closed with profit of at least $2\,\Delta = \$1.686$ (assuming the transition cost is zero) and enter into a new paired trade in the opposite direction of closed position. In figure 6.2, which depicts strategy (b), each triangle also represents the time that we decide to open a new paired trade in appropriate direction and square represents the time that we decide to close the only existing position with profit of at least $\Delta = \$0.843$ (assuming the transition cost is zero). In figure 6.2, profit/loss (c), we also demonstrate how the profit or loss in both strategies, (a) and (b) evolves through the specified time frame.

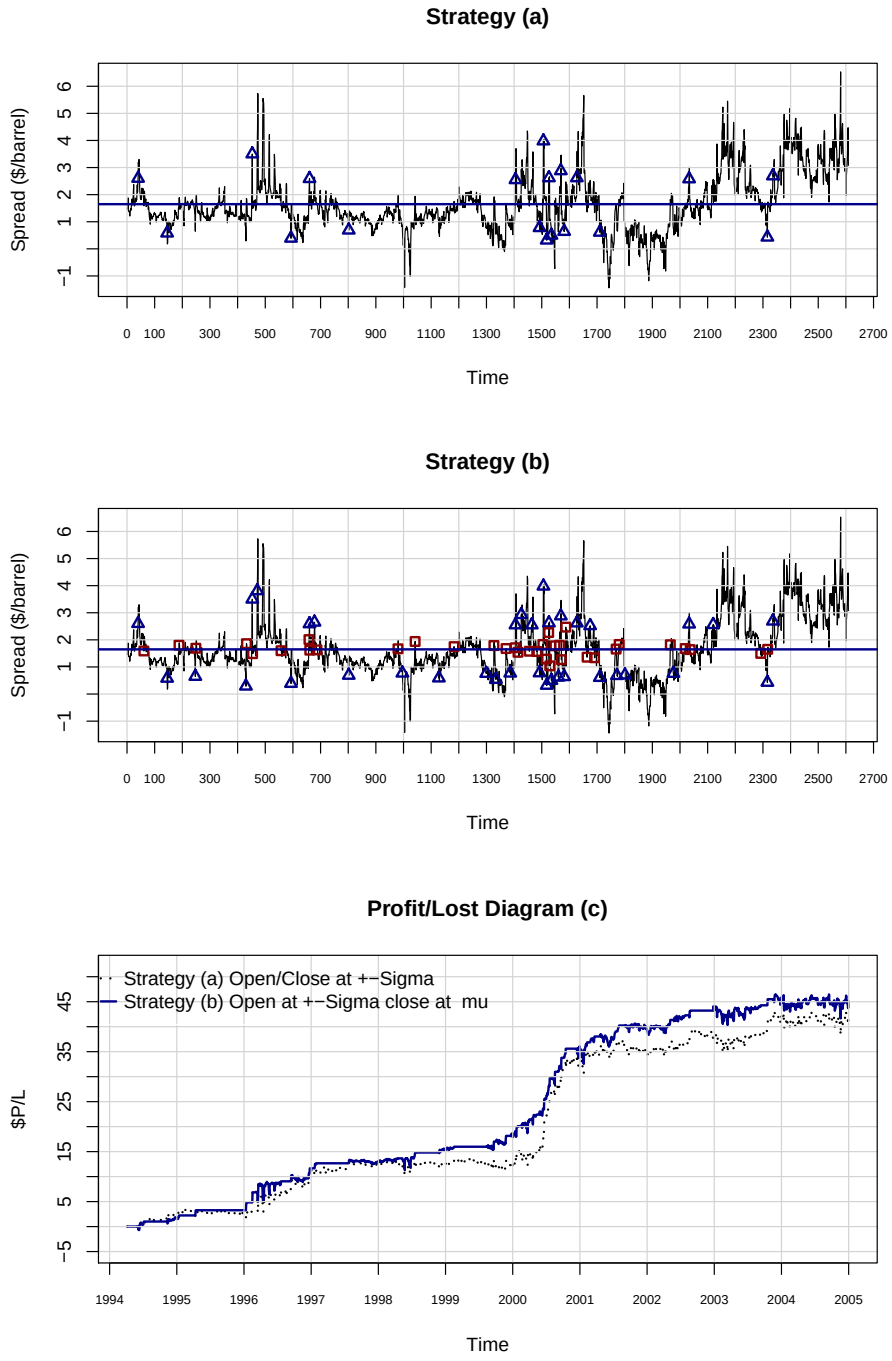

Figure 6.2: Time plot of the location spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in one-factor method: Plot (a) shows the series of trades in strategy (a) in which whenever we hit the lower barrier, $\ell_1 = b - \Delta = 1.649 - 0.843$ or upper barrier, $\ell_2 = b + \Delta = 1.649 + 0.843$, we close existing paired trade and open new position in opposite direction. Plot (b) shows the series of trades in strategy (b) in which whenever we hit the lower barrier, $\ell_1 = b - \Delta = 1.649 - 0.843$ or upper barrier, $\ell_2 = b + \Delta = 1.649 + 0.843$, we only open a new position in proper direction and we wait until to revert to long-run mean, $b = 1.649$ when we unwind existing paired trade. Plot (c) depict how profit/loss growth is both strategies.

We summarize the trades with respect to two different strategies (a) and (b) in our sample data in table 6.3. As we can see, for this data set, unlike the cointegration approach, strategy (b) performs slightly better than strategy (a) because it has higher total profit. However, if we include the transaction costs of trades, we can see these two strategies have more less equal performance.

Table 6.3: Summary of trades in two different strategies (a) & (b) shown in figure 6.2 in the location spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in one-factor method.

Strategy (a)				Strategy (b)			
Number of trades	Profit per trade	Expected total profit	Total profit	Number of trades	Profit per trade	Expected total profit	Total profit
19	\$1.686	\$30.35	\$42.05	36	\$0.843	\$29.51	\$45.5

6.4 Cointegration approach and One-factor Method results Comparison:

Once we compare the outcomes of trades of these two approaches in tables 6.1 and 6.3, we immediately find out that the cointegration approach has a dramatically higher total profit and higher number of trades in both strategies (a) and (b). This comparisons approach might be irrelevant because the spread process in cointegration approach, $r_t = p_t^A - \beta p_t^B$ ($\beta = 0.93$) and location spread process in one-factor method, $X_t = p_t^A - p_t^B$ are fundamentally different. By looking at figures 6.2 and 6.1, the location spread process has significant behavior change in last five years and unlike the spread process, we rarely trade in location spread in last five years. It worth mentioning that if we fit the spread process, $r_t = p_t^A - \beta p_t^B$ sample data to a one-factor model, we may have equivalent results in both approaches.

6.5 Analysis of Optimal Trading Strategies

Here, we attempt to derive the optimal strategies for both models namely the cointegration and Elliott *et al.* (2005)'s one-factor process. In other words, we will derived optimal Δ for both strategies, (a) and (b) for both models. As we explained, Δ is considered to be the deviation from the equilibrium level μ_r. Strategy (a) is defined as follows: Once the lower barrier, $\ell_1 = \mu_r - \Delta$ or upper barrier, $\ell_2 = \mu_r + \Delta$ is crossed, the existing paired trade will be closed and a new position in the opposite direction will be opened. Strategy

(b) is defined the same as strategy (a) for opening a new position, but liquidating the opened position will be at equilibrium level, μ_r. Notice that in the one-factor model μ_r is replaced by b (long-run mean). Vidyamurthy (2004) shows that if the spread process assumes to be a Gaussian white noise series ($\mathcal{N}(0, \sigma^2)$), a profitability measure function for trading strategy (a) in the time interval $(0, T)$ is $pmf(\Delta) = 2T\Delta\left(1 - \Phi(\frac{\Delta}{\sigma})\right)$ where $\Phi(x)$ is the standard normal cumulative distribution. Therefore, the optimal deviation Δ^* can be derived by maximizing of $pmf(\Delta)$ and the optimal deviation Δ^* occurs at 0.75σ. It worth mentioning that the profitability measure function remains identical for the strategy (b), which means $\Delta^* = 0.75\sigma$. Figure 6.3 clearly illustrates the optimal amount of sigma away from the long-run mean in the given profitability measure function for a Gaussian white noise. Since in the cointegration approach, simple linear regression

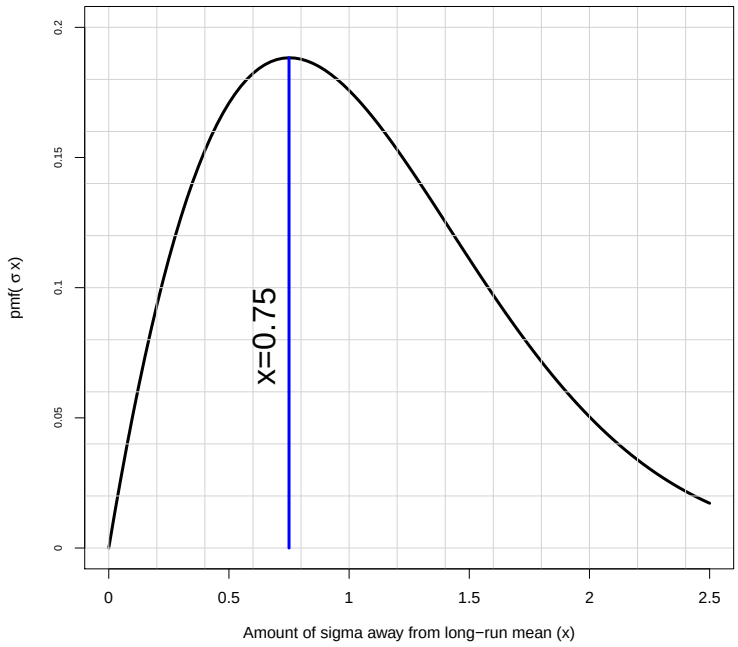

Figure 6.3: The plot depicts the optimal amount of sigma away from long-run mean in the given profitability measure function for a Gaussian white noise

is deployed to estimate the model parameters such as the hedging ratio, the spread or residual series r_t can be considered to be a Gaussian white noise. Therefore, the optimal deviation Δ^* can be considered for this approach. It worth mentioning that in section 6.2.1, using Engle-Granger two step method, we showed that the residual series r_t for our dataset follows the stationary ARMA(1,1), so we might do better with the optimal devia-

tion Δ^* derived using ARMA(1,1)'s optimal hitting time; however, the results are already impressive based on this simplifying assumption. In section 6.2.1, the cointegration approach is deployed to the spread process between WTI crude oil and Brent oil for daily price of front contracts from April 1994 to January 2005 and the estimated parameters are given. The optimal deviation Δ^* for this dataset is $\Delta^* = 0.75\hat{\sigma} = 0.75 * 0.816 = 0.612$. The trades with respect to two different strategies (a) and (b) using the optimal deviation Δ^* in our sample data are summarized in table 6.4. As we can see, likewise, the analysis of trades in the cointegration approach using arbitrary $\Delta = 0.816$, summarized in table 6.1, strategy (a) has better performance than strategy (b) using optimal Δ^* in two aspects: first it has higher profit and second it has lower number of trades meaning that it has lower transaction costs. As it was expected, the profitability of trades in both strategies (a) and (b) using optimal $\Delta^* = 0.612$ are quite higher than that of arbitrary $\Delta = 0.816$. In general, investigating the first passage time (define as a random variable)

Table 6.4: Summary of trades, implemented for the observed dataset, in two different strategies (a) & (b) using the optimal deviation $\Delta^* = 0.612$ in the spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in cointegration approach.

Strategy (a)				Strategy (b)			
Number of trades	Profit per trade	Expected total profit	Total profit	Number of trades	Profit per trade	Expected total profit	Total profit
41	\$1.225	\$50.21	\$70.73	63	\$0.612	\$ 38.57	\$65.61

involves calculating the probability density function of first passage time for any given stochastic process. This derivation, in most cases is complicated and there is no explicit solution for the first passage time density. The moments of the first passage time have been studied for several special cases. Linetsky (2004) and Wang and Yin (2008) are some of these studies. In particular, the first passage time for Ornstein-Uhlenbeck (OU) process has been investigated in several studies. For instance, for the standard case of OU where $a = 1$, $b = 0$, and $\sigma = \sqrt{2}$ in equation 6.1, the density and the moments of the first passage time have been derived in Blake and Lindsey (1973), Thomas (1975), and Ricciardi and Sato (1988). By using $Y_\tau = \frac{\sqrt{2a}}{\sigma}(X_t - b)$ and $\tau = at$, and transforming equation 6.1 into the dimensionless process, Bertram (2010) derives analytical solution for optimal lower and upper thresholds when the spread dynamics is considered to follow an OU process. He also shows that the optimal thresholds are symmetric around the long-run mean b for both his optimal strategy choices namely maximizing *the expected return* and maximizing the *Sharpe ratio*. Since Bertram (2010) does not consider short selling, he merely builds his strategy to enter in a trade at lower optimal threshold ($-\Delta^*$) and exit at upper optimal threshold (Δ^*). However, his optimal thresholds are consistent

with both trading strategies (a) and (b).
Here, we state the final results of Bertram (2010)'s paper for the optimal expected return choice. To do so, we first explain his model's assumptions as follows: Let $X_t = \ln(S_t)$ to be the logarithm of an asset or commodity that is mean-reverting; he assumes X_t to follow an OU process. Without loss of generality, he considers the long-run mean of X_t to be zero ($b = 0$ in equation 6.1). He defines his continuous trading strategy by entering at $x_t = l$, exiting at $X_t = m$ ($l < m$), and waiting until the process to revert back to $x_t = l$, when the trading cycle is complete. He also considers c to be the transition costs of entering and exiting a trade. For profitability of a trade, it is required to have $m > l + c$. Bertram (2010) derives the expected profit per unit time for the maximum expected return strategy as follows:

$$\mu(l, m, c) = \frac{a(m - l - c)}{\pi\left(\text{Erfi}\left(\frac{m\sqrt{a}}{\sigma}\right) - \text{Erfi}\left(\frac{l\sqrt{a}}{\sigma}\right)\right)}, \tag{6.2}$$

where $\text{Erfi}(z)$ is an entire function called the imaginary error function ($\text{Erfi}(z) = i\,\text{Erf}(iz)$ and $\text{Erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-u^2}du$).
By maximizing the expected return function given in equation 6.2 with respect to l and m, Bertram (2010) shows that $l < 0$ and $m = -l$. This finding confirms that optimal lower (entry) and upper (exit) thresholds are symmetric about long-run mean b. Using Taylor series expansion, he approximates the optimal l^* for small values of $\frac{l\sqrt{a}}{\sigma}$ as follows:

$$\begin{aligned}\Delta^* = -l^* \;\; = \;\; & \frac{c}{4} + \frac{c^2 a}{4(c^3a^3 + 24ca^2\sigma^2 - 4\sqrt{3c^4a^5\sigma^2 + 36c^2a^4\sigma^4})^{\frac{1}{3}}} \\ & + \frac{(c^3a^3 + 24ca^2\sigma^2 - 4\sqrt{3c^4a^5\sigma^2 + 36c^2a^4\sigma^4})^{\frac{1}{3}}}{4a}.\end{aligned} \tag{6.3}$$

One can see that when the transaction cost c equals zero, the difference between optimal entry and exit thresholds of a trade approaches zero. This means that at any given time, the trader should progressively balance the positions to fulfil as many trades as possible. In such a case, the frequency of the trades are unprecedentedly high in theory. This observation leads to the fact that the optimal strategy based on this model will be more appropriate for high frequency trading. Sometimes the spread traders may consider to have an optimal trading strategy which is more suitable for long term investment. In this case we propose to derive the optimal thresholds based on stationary distribution, which is normal for Elliott *et al.* (2005)'s one-factor process. Therefore, since the stationary distribution is $\mathcal{N}(0, \frac{\sigma^2}{2a})$, the optimal deviation Δ^* based on the stationary distribution is $\Delta^* = 0.75\frac{\sigma}{\sqrt{2a}}$. Note that the optimal deviation based on the stationary distribution is independent of the transaction cost. The trades with respect to two different strategies (a) and (b) using the optimal deviation Δ^* derived by Bertram (2010)'s method and

stationary distribution for Elliott *et al.* (2005)'s one-factor model in our sample data are summarized in table 6.5. The results show that in both strategies (a) & (b), the total expected profit and the total profit derived by stationary distribution's optimal deviation are higher than that of Bertram (2010)'s method. This observation confirms our claim that the optimal deviation derived by stationary solution is more suitable to long-term investment strategy.

Table 6.5: Summary of trades, implemented for the observed dataset, in two different strategies (a) & (b) using the optimal deviation obtained by Bertram (2010)'s method and stationary distribution in the spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in Elliott *et al.* (2005)'s one-factor model. The estimated parameters are given in table 6.2 and transaction cost c is assumed to be \$0.08.

Method	Δ^*	**Strategy (a)**				**Strategy (b)**			
		# of trades	Profit per trade	Expected total profit	Total profit	# of trades	Profit per trade	Expected total profit	Total profit
Bertram	0.354	40	\$0.628	\$25.11	\$41.89	69	\$0.274	\$ 18.90	\$43.69
St. Dis.	0.415	34	\$0.751	\$25.53	\$43.27	65	\$0.335	\$ 21.80	\$45.56

In chapters 4 and 5, it was discussed that the spread processes are usually highly volatile, exhibit weak stationary, skewness (asymmetric), kurtosis, and heavy tails in their transition densities. The model parameters might also switch from time to time. To consider these undeniable characteristics two new models, namely the generalized one-factor regime switching and the new one-factor mean-reverting process, are introduced. The analysis confirms that the transition densities of the spread processes are mainly non-Gaussian. As a result, the Gaussian assumption of the transition densities are not quite correct. As another consequences is that the optimal thresholds derived based on the Gaussian assumption cannot be the best choices. Here, we attempt to discuss other measures that can be taken to improve the optimal trading strategies. Constructing the optimal thresholds based on a more realistic model for the spread process such as our introduced models can considerably improve the trading strategies. Based on the skewness, kurtosis, and heavy tails appropriate to a given spread process, the optimal thresholds are not necessarily symmetric. The opened position by hitting the optimal upper threshold, and opened position by hitting the optimal lower threshold might be needed to close based on different closing thresholds. To show how these considerations can improve the optimal trading strategies, we illustrate these ideas empirically by implementing in our dataset. The trades based on two altered versions of strategies (a) & (b) are summarized in table 6.6. Strategy (a^*) is the asymmetric version of strategy (a) which means that lower deviation $\Delta_1 = \Delta_4$ and upper deviation $\Delta_2 = \Delta_3$ have

different distances from the long-run mean b. Strategy (b^*) is the asymmetric and modified version of strategy (b) which means the trader take proper positions (enter in a trade) in the lower deviation Δ_1 and upper deviation Δ_2 from the long-run mean b. and closes the existing position at two different deviations Δ_3 and Δ_4 (if the position was executed at Δ_1, it would close at Δ_3 and if the position was executed at Δ_2, it would close at Δ_4). All deviations are from the long-run mean b. Figure 6.4 illustrates how the trades are placed and are unwound as well as their profit and lost diagrams in these two evolved trading strategies (a^*) & (b^*) for the location spread between WTI and Brent oils in our dataset. Strategy (b^*) has exceptionally the highest performance among all four strategies (a), (b), (a^*) & (b^*) in this particular empirical experiment. Comparison between the summary of the trades based on the new asymmetric strategies (a^*) & (b^*) and the summary of the trades based on symmetric strategies (a) & (b) using the optimal deviations, derived by Bertram (2010)'s method and stationary distribution reveal the advantages of these new asymmetric strategies (a^*) & (b^*) over the the symmetric (a) & (b) including higher expected total profit and total profit. These empirical observations suggest to consider more realistic both for modeling the spread process itself and the optimal trading strategies.

Table 6.6: Summary of trades, implemented for the observed dataset, in two different modified strategies (a^*) & (b^*): strategy (a^*) is the asymmetric version of strategy (a) meaning that lower deviation $\Delta_1 = \Delta_4$ and upper deviation $\Delta_2 = \Delta_3$ have different distances from the long-run mean b. strategy (b^*) is the asymmetric and modified version of strategy (b) which means the trader take proper positions (enter in a trade) in the lower deviation Δ_1 and upper deviation Δ_2 from the long-run mean b. and closes the existing position at two different deviations Δ_3 and Δ_4 (if the position was executed at Δ_1, it would close at Δ_3 and if the position was executed at Δ_2, it would close at Δ_4). All deviations are from the long-run mean b. The empirical data is the observations of the spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005. The strategies are implemented based on Elliott *et al.* (2005)'s one-factor model. The estimated parameters are given in table 6.2 and transaction cost c is assumed to be \$0.08.

Strategy	Δ_1	Δ_2	Δ_3	Δ_4	# of trades	Profit per trade	Expected total profit	Total profit
(a^*)	-0.632	1.345	1.345	-0.632	17	\$1.90	\$ 32.32	\$42.96
(b^*)	-0.628	1.346	0.189	0.332	46	\$0.737(0.934)	\$ 38.04	\$60.24

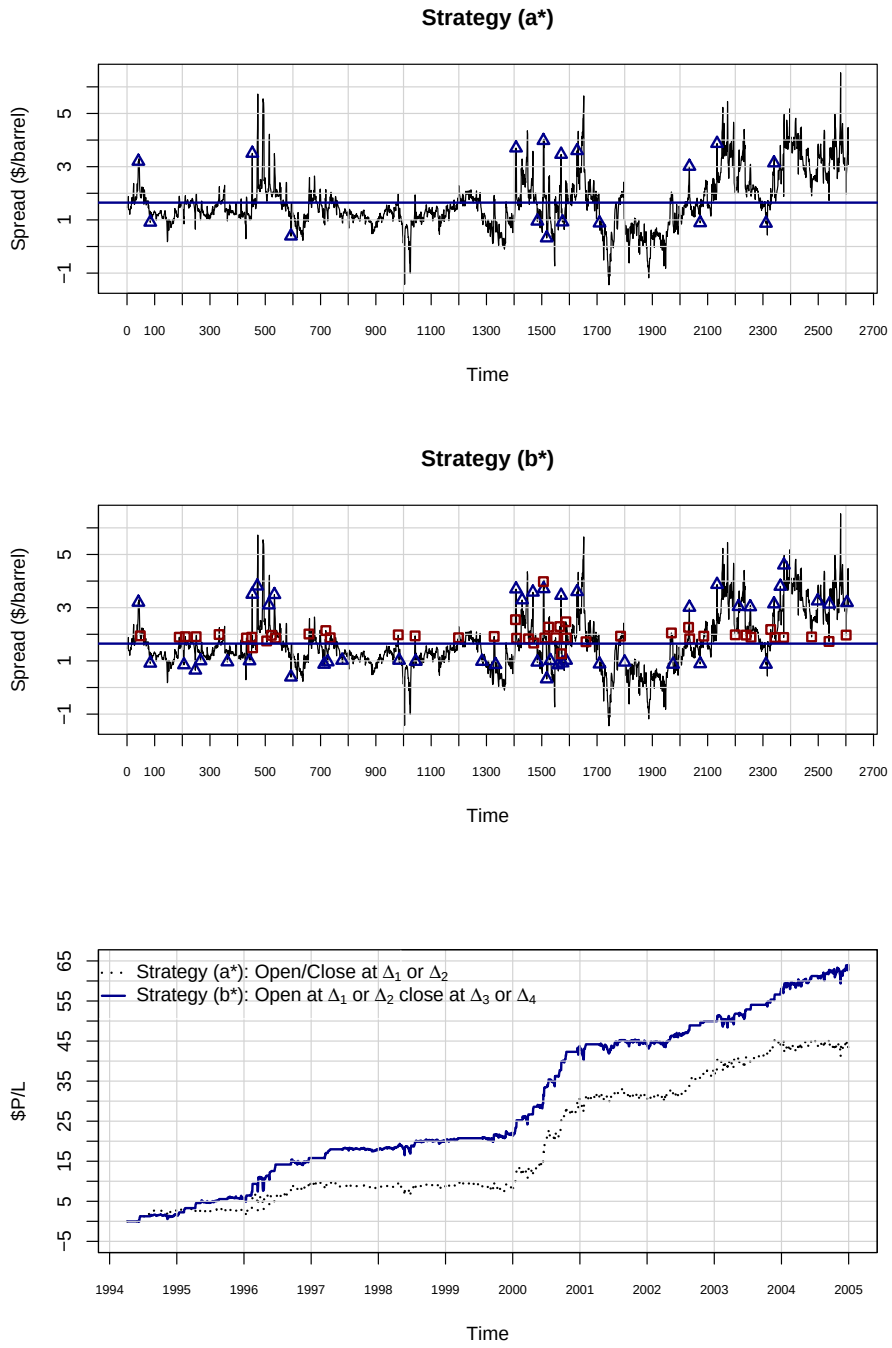

Figure 6.4: Time plot of the location spread process between WTI crude oil and Brent oil daily front contracts prices from April 1994 to January 2005 in one-factor method: Plot (a) shows the series of trades in revised strategy (a^*) in which whenever we hit the lower barrier, $\ell_1 = b + \Delta_1 = 1.649 - 0.632$ or upper barrier, $\ell_2 = b + \Delta_2 = 1.649 + 1.345$, we close existing paired trade and open new position in opposite direction. Plot (b) shows the series of trades in revised strategy (b^*) in which whenever we hit the lower barrier, $\ell_1 = b + \Delta_1 = 1.649 - 0.628$ or upper barrier, $\ell_2 = b + \Delta_2 = 1.649 + 1.346$, we only open a new position in proper direction and we wait until to revert to the proper closing barriers: $\ell_3 = b + \Delta_3 = 1.649 + 0.189$ (when existing trade was opened at ℓ_1) or $\ell_4 = b + \Delta_3 = 1.649 + 0.332$ (when existing trade was opened at ℓ_2) when we unwind existing paired trade. Plot (c) depict how profit/loss growth is both strategies.

6.6 Conclusion:

In this chapter, we focused on building the optimal trading strategies for the spread processes. The study is inspired by the main property of a spread process: mean-reversion. Two trading strategies (a) & (b) were explained in detail and empirically shown how they work in two cases: when cointegration approach is considered for modeling the spread and when the spread dynamics is assumed to follow the Elliott *et al.* (2005)'s one-factor process. The optimal trading strategies based on Vidyamurthy (2004)'s algorithm for Gaussian white noise, applicable for cointegration approach, were explained and applied for our observed dataset, the spread between WTI and Brent crude oils. Bertram (2010)'s optimal trading strategies, constructed based on maximization of the expected return, were reviewed when the spread process is assumed to follow the Ornstein-Uhlenbeck (OU) process. For the long-term investment strategies, we proposed to deploy the stationary distribution to obtain the optimal thresholds when the spread processes are considered to follow Vasicek dynamics. Two new trading strategies (a^*) & (b^*), modified asymmetric versions of (a) & (b), were developed. To demonstrate how these new modified strategies can enhance the optimal trading strategies, the strategies were empirically applied to our dataset. The results confirm the superiority of these newly defined strategies. The study of how to obtain the optimal thresholds (Δ_1^* and Δ_2^*) for strategy (a^*), and how to derive the optimal entering thresholds (Δ_1^* and Δ_2^*) and exiting thresholds (Δ_3^* and Δ_4^*) for strategy (b^*) were left for future research.

Chapter 7

SUMMARY AND FUTURE RESEARCH DIRECTIONS

7.1 Conclusions and Contributions

This thesis is mainly concentrated on the commodity pricing models, the commodity spread pricing models, and the optimal trading strategies for the spread processes. Extensive numerical works, stochastic models and mathematical formulations, performed in each model, are the major features of this thesis. First, in Chapter 1, we reviewed some of crucial notions, concepts, and difficulties that we confront when we study the commodity markets and derivatives on the commodity. The Kalman filter algorithm and the new local linearization method were investigated and reviewed in Chapter 2. Based on characteristics of the model SDE, one of these two powerful algorithms was exploited to estimate parameters of the commodity pricing models and the spread pricing models.

In Chapter 3, we described three commodity pricing models namely preliminary, Schwartz (1997)' one-factor and Gibson and Schwartz (1990)' two-factor models in detail. In these models, we also derived futures and forward prices. This two-factor Gibson & Schwartz model is one of the most commonly applied for pricing by market participants. We developed a new generalized one-factor mean-reverting dynamics to model a commodity spot price process. Since the new process is a nonlinear dynamics with unknown transition density, the new local linearization algorithm was applied to estimate the model parameters. We argued that the new generalized SDE posses some key characteristics to capture the most essential properties including skewness, heavy tails and excess kurtosis, observed in a commodity spot-price evolution. The new one-factor and Schwartz one-factor models were compared by fitting both models to daily front futures prices of WTI crude oil from October 1, 2004 to December 31, 2014. The calibration

results confirmed the superiority of the new SDE.

In Chapter 4, we also investigated three main approaches to price a spread process namely cointegration, one-factor and two-factor models. We apply these three models to our real empirical sample data and compare the results. Later, we analysis the recent behavioral change in the location spread between WTI crude oil and Brent oil. Since these models are not flexible enough to capture all behavior changes, we extended the one-factor and two-factor spread models by adding a compound Poisson process where jump sizes follow double exponential distribution. Later, the generalized one-factor mean-reverting dynamics is introduced to capture the major properties of spread processes, compared to Vasicek process and showed the new model's absolute superiority by fitting to the spread between WTI and Brent oils. Since the spread between WTI and Brent in observed data is experienced fundamental change in early 2011, the generalized model and Vasicek process are deployed in regime switching framework by considering two regimes. The estimation results fully justified applying the regime switching algorithm. The AIC and BIC criteria in both regimes chose this new generalized model.

In Chapter 5, we introduced a new mean-reverting random walk, and derive its continuous stochastic process. Its stationary distribution was derived, but all attempts to solve this nonlinear dynamics analytically for transition density unfortunately failed. We also generalized this new mean-reverting process to equip the diffusion to capture possible skewness of the spread process. We compare this one-factor model to Elliott *et al.* (2005)'s one-factor model and using simulation results, we showed that this new mean-reverting one-factor model has capability to capture the potential heavy tail or tails, kurtosis, and skewness of the spread processes. To estimate parameters of this nonlinear dynamics, the new local linearization method was deployed. We applied this new SDE and Vasicek processes to the daily spread between WTI and WTS crude oils from January 2000 to January 2013. The maximum log-likelihood value, AIC, BIC all endorsed that this new SDE outperformed compare to Vasicek process.

Finally, in Chapter 6, we mainly focused on building the optimal trading strategies for the spread processes. The study is inspired by the main property of a spread process: mean-reversion. Two trading strategies (a) & (b) were explained in details and empirically shown how they work in two cases: when cointegration approach is considered for modeling the spread and when the spread dynamics is assumed to follow the Elliott *et al.* (2005)'s one-factor process. The optimal trading strategies based on Vidyamurthy (2004)'s algorithm for Gaussian white noise, applicable for cointegration approach, were explained and applied for our observed dataset, the spread between WTI and Brent crude oils. Bertram (2010)'s optimal trading strategies, constructed based on maximization of the expected return, were reviewed when the spread process is assumed to follow the

Ornstein-Uhlenbeck (OU) process. For the long-term investment strategies, we proposed to deploy the stationary distribution to obtain the optimal thresholds when the spread processes are considered to follow the Vasicek dynamics. Two new trading strategies (a^*) & (b^*), modified and asymmetric versions of (a) & (b), were developed. To demonstrate how these new modified strategies can enhance the optimal trading strategies, the strategies were empirically applied to our dataset. The results confirm the superiority of these newly defined strategies.

7.2 Principal Contributions

The four main contributions of this thesis can be summarized as follows: (i) introducing of a new generalized one-factor mean-reverting dynamics to model a commodity spot price process in Chapter 3; (ii) generalizing one-factor mean-reverting dynamics to model the spread processes and implementing regime switching framework with two regimes in the new model, which were investigated in Chapter 4; (iii) presenting a novel mean-reverting random walk and deriving its continuous stochastic process which is a great choice to model spread pricing models as presented in Chapter 5; (iv) revising two existing spread trading strategies: (a) & (b) by considering asymmetric thresholds as described in Chapter 6.

7.3 Future Research Directions

There are several possible extensions to improve the models studied in this thesis that can be investigated further in future. Here, we mention some of these extensions in detail as follows:

- The state variables of the dynamics, studied in this research, are unobserved. The state space form is a framework for unobservable variables. Since the state-space forms that can be generated to the newly developed models in this thesis are non-linear, we exploited the new local linearization method by using the front futures prices as proxy for the spot prices. To deal with this limit, we suggest to deploy *Particle Filters or sequential Monte Carlo methods*, which can be applied in nonlinear state-space form, for future work.

- In Chapter 4, we extended the one-factor and two-factor spread models by adding a compound Poisson process where jump sizes follow double exponential distribution to capture sudden abrupt breaks. However, the implementation of these two processes with jumps have not been carried out and can be considered for future studies.

- In Chapter 5, we attempted to find analytical transition density for the novel mean-reverting spread process; however, we failed to solve the SDE analytically. Perhaps, it is worth investigating further to solve the dynamics analytically.

- In Chapter 6, the optimal trading strategies have not been discussed for Dempster *et al.* (2008)'s two-factor spot spread process, so future study of the optimal thresholds for this model may lead to high performance spread trading strategies. Performance of the optimal trading strategies heavily rely on the assumption of statics spread model parameters, so considering regime-switching framework for the optimal strategies can end up increasing the profit and the performance. We introduced asymmetric spread trading strategies (a^*) and (b^*); however, the study of how to obtain the optimal thresholds (Δ_1^* and Δ_2^*) for strategy (a^*), and how to derive the optimal entering thresholds (Δ_1^* and Δ_2^*) and exiting thresholds (Δ_3^* and Δ_4^*) for strategy (b^*) can be considered for future work.

.1 Appendix

Commodity: In economics, a commodity is a term that refers to any marketable item produced to fulfill wants or needs. Commodities comprise both goods and services, but the term is more particularly used to refer to goods only

Par grade: Minimum acceptance standard for a commodity is called "a basic grade", "par grade", or "contract grade"

Spot price: Spot price is the price for immediate delivery of a particular commodity

Front month contract: Front month contract used in futures trading to refer to the contract month with an expiration date closest to the current date, which is often in the same month. In other words, this would be the shortest duration contract that could be purchased in the futures market

Convenience yield: The convenience yield is defined as the overall benefits that holder of commodity receive minus the costs especially the cost of storage with exception of the cost of financing

Mean reversion: Mean reversion" is a theory proposing that most of economic markets fluctuates around or move back towards a "mean" or "average" (equilibrium level)

Seasonality: During different periods of time in year, the supply and demand have dramatic changes in particular commodities in which drive seasonality in commodity markets

Unobservable variables: In statistics, latent variables, hidden variables or unobservable variables, are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed

Bibliography

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, **70**(1), 223–262.

Alexander, C. (1999). Correlation and cointegration in energy markets. *Managing energy price risk*, **2**.

Bessembinder, H., Coughenour, J., Monroe M. and Seguin P. (1995). Mean Reversion in Equilibrium Asset Prices: Evidence from the Futures Term Structure. *Journal of Finance*, **50:1**, 361–374.

Bertram, W. K. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical mechanics and its applications*, **389**(11), 2234–2243.

Bjork, T. (2004). *Arbitrage theory in continuous time*. Oxford university press.

Blake, I. F., and Lindsey, W. C. (1973). Level-crossing problems for random processes. *Information Theory, IEEE Transactions on*, **19**(3), 295–315.

Bloomberg L.P. (2014). Daily WTI's fornt contract prices and daily WTS's spot prices 1/4/00 to 12/31/12. *Retrieved September 23, 2014 from Bloomberg terminal.*

Borenstein, S., and Kellogg, R. (2012). The Incidence of an Oil Glut: Who Benefits from Cheap Crude Oil in the Midwest?. *Working paper*

Cappe, O., Moulines, E., & Ryden, T. (2005). *Inference in hidden Markov models (Vol. 6)*. New York: Springer.

Carmona, R., and Ludkovski, M. (2004). Spot convenience yield models for the energy markets. *Contemporary Mathematics*, **351**, 65–80.

Chatnani, N. N. (2010). *Commodity Markets: operations, instruments, and applications.* Tata McGraw-Hill Education, New Delhi.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1981). The relation between forward prices and futures prices. *Journal of Financial Economics*, **9**(4):321–346.

Craddock, M. J., & Dooley, A. H. (2001). Symmetry group methods for heat kernels. *Journal of Mathematical Physics*, **42**(1), 390–418.

Cvitanic, J., and Zapatero, F. (2004). *Introduction to the economics and mathematics of financial markets*. MIT press.

Dempster, M. A. H., Medova, E., and Tang, K. (2008). Long term spread option valuation and hedging. *Journal of Banking & Finance*, **32**(12):2530–2540.

Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space models. *Biometrika*, **84**, 669–684.

Durbin, J. and Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.

Elliott, R. J., Aggoun, L., & Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control*. Springer-Verlag.

Elliott, R. J., Hoek, J. V. D. and Malcolm, W. P. (2005). Pairs Trading. *Quantitative Finance* , **5**(3):271–276.

Elliott, G. R., Rothenberg T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, **64**: 813–836.

Engle, R. F. and Granger, C. W. J. (1987). Cointegration and error correction representation, estimation and testing. *Econometrica*, **55**:251–276.

Erlwein, C., and Mamon, R. (2009). An online estimation scheme for a HullWhite model with HMM-driven parameters. *Statistical Methods and Applications*, **18**(1), 87–107.

Gallant, A. R., and Tauchen, G. (1996). Which moments to match?. *Econometric Theory*, **12**(04), 657–681.

Geman, H. (2005). *Commodities And Commodity Derivatives: Modelling And Pricing For Agriculturals, Metals And Energy*. John Wiley & Sons, Ltd.

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

Hamilton, J. D. (1994). *Time series analysis (Vol. 2)*. Princeton: Princeton university press.

Hamilton, J. D. (2008). Regime-switching models. *The new palgrave dictionary of economics*, **2**.

Harvey, A. C. (1993). *Time Series Models*. Harvester Wheatsheaf, Hemel Hempstead, UK.

Ingrid, P., Aug 29,(2013). Why is there volatility in the WTI-Brent crude oil spread? *¡http://marketrealist.com/2013/08/whats-causing-volatility-wti-brent-spread/¿*

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, **59**(6):1551–1580.

Kalman, R. E., (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**(1):35–45.

Kalman, R. E. and Bucy, R. S.(1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, **83**(1):95–108.

Kedem, B., and Fokianos, K.(2002). *Regression Models for Time Series Analysis*. Wiley, New York.

Kim, C. J. and Nelson, C. R. (1999). *State Space Models with Regime Switching*. Academic, New York.

Kou, S. (2002). A jump diffusion model for option pricing. *Management Science*, **48**(8):1086–1101.

Latte, T. (2011). Morgan Stanley and the Birth of Statistical Arbitrage. *Online website*.

Linetsky, V. (2004). Computing hitting time densities for CIR and OU diffusions: Applications to mean-reverting models. *Journal of Computational Finance*, **7**, 1–22.

McLeish, D. L. (2005). *Monte Carlo simulation and finance*. John Wiley & Sons.

Merton, R. C., (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*. **3** 125–144.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, **75**(2):335–346.

Phillips, P. C. (1986). Understanding spurious regressions in econometrics. *Journal of econometrics*, **33**(3):311–340.

Pilipovic, D. (2007). *Energy risk: Valuing and managing energy derivatives* . New York: McGraw-Hill.

Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: A local linearization approach. *Statistica Sinica*, **2**, 25–83.

Ricciardi, L. M., and Sato, S. (1988). First passage-time density and moments of the Ornstein-Uhlenbeck process. *Journal of Applied Probability*, 43–57.

Risken, H. (1989). *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Series in Synergetics.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, **52**(3):923–973.

Serletis, A. (2007). *Quantitative and Empirical Analysis of Energy Markets Vol. 1*. World Scientific Publishing Co. Pte.Ltd.

Shoji, L. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, **16**(4):733–752.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, **52**(5), 1973–2002.

Thomas, M. U. (1975). Some mean first-passage time approximations for the Ornstein-Uhlenbeck process. *Journal of Applied Probability*, 600–6004.

Tsay, R. S.(1991). Two canonical forms for vector $ARMA$ processes. *Statistica Sinica*, **1**:247–269.

Tsay, R. S. (2010). *Analysis of financial time series*. Wiley, New York.

United Nations, (2011). Price Formation in Financialized commodity markets: the role of information. *The secretariat of the United Nations Conference on Trade and Development*, New York and Geneva.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, **5**: 177–188.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. John Wiley and Sons, Canada.

Wang, H., and Yin, C. (2008). Moments of the first passage time of one-dimensional diffusion with two-sided barriers. *Statistics & Probability Letters*, **78**(18): 3373–3380.

Zivot, E. and Andrews, K. (1992). Further Evidence On The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis. *Journal of Business and Economic Statistics*, **10**(10): 251–270.

Mir Hashem Moosavi Avonleghi

SUMMARY OF QUALIFICATIONS:

- Ph.D. in Financial Modeling
- M.Sc. in Financial Mathematics
- Well-developed knowledge of quantitative finance, commodity pricing models , time series analysis, trading strategies for energy spread processes, option pricing, and risk management
- Over ten years experience in system analysis, programming and technical support
- Experienced in designing, developing, implementing, analyzing, programming, maintaining, mathematical modeling and modifying systems
- Skilled at organizing complex projects, defining project priorities, and delegating tasks
- Self-starter, goal-oriented strategist with confidence, perseverance to promote success
- Experienced team player, bringing enthusiasm and energy into group efforts
- Over eight years experience in teaching various courses

EXPERIENCE:

TEACHING AND RESEARCH ASSISTANT **2009 to 2015**

University of Western Ontario (UWO), London, Ontario, Canada

- Worked as a teaching assistant for Financial Markets and Investments, Advanced Financial Modeling , Calculus , and Statistics & Probability for Engineers
- **Ph.D. Thesis:** **2010 to 2015**

  Topic: Quantitative Techniques for Pairs Trading in Commodity Markets
  - Reviewed three major approaches to price a spread process : cointegration, one-factor and two-factor models and their extension to include jumps
  - Introducing optimal trading strategies based on the main properties of the spread process mean-reversion and relative pricing in commodity markets
  - Introduced a novel mean-reverting random walk, deriving its continuous stochastic process, and applying it to model a spread process
  - Confronted all models with real location & calendar spread data
- **Master Project:** **2010**

  Topic: Drawbacks of Black-Scholes Model and Applications of the Skewed Generalized Error Distribution (SGED)

INSTRUCTOR /BOARD OF EDUCATION/ HEAD OF SCIENCE DEPARTMENT

Karr Higher Education Institution, Ghazvin, Iran **1999 to 2007**

Achievements:

- Experienced instructor of courses including Operations Research, Calculus, Statistics and Probability for Engineers, Fundamentals of Computer and Programming (C, Turbo Pascal)
- Led various research projects
- Led the science department faculty members from 2002 to 2007

SYSTEM ANALYST / PROGRAMMER **2004 to 2007**

NIOPDC (National Iranian Oil Products Refinery and Distribution) Tehran, Iran

Cost Minimization of Oil Product Transportation Project (2006 to 2007)

- Supervised, consulted and collaborated in Mathematical Modeling, and System Analysis
- Consulted with users to identify operating procedures and clarify program objectives
- Developed mathematical models, computer analysis to optimize processes
- Successfully managed technical meetings for analyzing project

Achievements:

- Decreased monthly oil products distribution costs by 5% (approximately $23,000 CDN/m)

Tank Lorry Measuring System Project (2004 to 2005)

- Designed, developed, programmed and maintained system
- Assisted users and wrote manuals

Achievements:

- Greatly increased efficiency

SYSTEMS ANALYST / MATHEMATICAL MODELER **2004 to 2005**

Consulting Contracts, Tehran, Iran

- Collaborated to develop several decision making models which used data envelopment analysis
- Designed mathematical models and analyzed systems
- Liaised with senior executives and technical personnel
- Researched methods and indices used by similar companies

Achievements:

- Ranked companies based on productivity

MATHEMATICAL MODELER / SYSTEMS ANALYST / PROGRAMMER **1997 to 2004**

NIOPDC, Tehran, Iran

Optimization of liquid gas transportation system project (1997 to 2004)

- Developed program for solving mathematical model of project (Linear program-Big M Method)
- Successfully managed technical meetings for analyzing project
- Automated traditional methods of transportation system analysis
- Maintained the project after finishing the project, in which I restructured and improved developed application (restructured the software's databases) and revised mathematical model of project (by adding new constraints)

Achievements:

- Increased efficiency and subsidiary satisfaction and minimized transportation costs reduction (approximately $12,000 CDN /m)

TECHNICAL SKILLS:

- Well-shaped programming skills on Object-Oriented R, C#/C++, Delphi, Matlab
- Proficient in the use of Microsoft Office (Word, Excel, Project and PowerPoint)
- Well-developed knowledge of SQL Server (2000) to design database, entity-relationship modeling, implementation, query construction and optimization, providing documentation

EDUCATION:

Ph.D. Candidate in Financial Modeling **2015**

University of Western Ontario (UWO), London, Ontario, Canada

Master of Science in Financial Mathematics **2010**

University of Western Ontario (UWO), London, Ontario, Canada

Master of Science in Applied Mathematics (Operations Research) **1999**

Tarbiat Modarres University (T.M.U), Tehran, Iran

Bachelor of Science in Mathematics- Applications in Computer **1996**

Amirkabir University of Technology (Polytechnic), Tehran, Iran