MEAN REVERSION MODELING WITH APPLICATION IN ENERGY MARKETS

LUO WEI

(B.Eng.(Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING
DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2012

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

LuoWei

____

Luo Wei

1 Aug 2012

Acknowledgement

I am grateful to my supervisor Associate Professor Ng Kien Ming for his invaluable guidance, continuous support and advice during the last two academic years from April 2011 to January 2013. His unconditional commitment and confidence in the success of this Master thesis was an invaluable source of motivation. He guided me to the interesting problem of mean reversion with evidence and modeling. His enthusiasm on the topic and positive outlook has always inspired me.

Besides, I would like to thank

My beloved parents and grandparents, for giving me strength and courage throughout my studies and have supported my study overseas since December 2005. Without them I will not be able to complete this thesis in time.

Professor Sun Jie for wonderful teaching in graduate optimization course IE6001 in which I sharpened my knowledge of Convex Optimization, and contributed to my preparation of this thesis.

Dr. Kim Sijun for rigorous Mathematics proofs in graduate Stochatics Process course IE6004, her way of teaching Probabilities has inspired me on solving problems in a more pragmatic manner.

My classmates at ISE-Computing Lab for their friendship, help and support from August 2011 to August 2012. They have made my research experience a pleasant and enriching journey.

Summary

A phenomenon observed in energy prices is that they tend to exhibit mean-reversion behavior. This thesis proposes two new models on mean-reversion patterns of energy assets: Time-invariant Wavelet-Schwartz Model and Time-Varying State Space Model. The first model is capable of describing stationary time series with fixed degree of mean-reversion by incorporating wavelet-decomposition techniques into the one-factor Schwartz model. As a de-noising method, the wavelet filter is a useful tool to track the cycles of the price movements which can be modeled by mean-reversion. The second model can be used to describe mean reverting processes with constantly changing parameters by adopting a Bayesian estimation approach. The prediction step uses the Kalman Filter, while the Bayesian approach with variance gamma assumption is applied on the calibration of the time-varying mean reversion model. The proposed two models are applied to historical energy price data to test their performance in trading activity. The simulation results generated by the two models are then compared and discussed. This application shows that when different measures are taken, similar sensitivity appears by fixing a relationship between symmetric parameters.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Problem Description

The term mean reversion was first introduced by Fama and French [16] on American stock markets. They found a negative serial correlation in terms of market returns over horizons of three to five years and estimated that more than 40 % of the variation in asset returns were predictable, which were mainly attributed to a mean reverting stationary component in asset prices. Historically, Fama and French [16] and Poterba and Summers [43] are the pioneers who provided direct empirical evidence for mean reversion phenomenon. However, Lo and MacKinlay [33] found evidence against mean reversion in U.S. stock prices using weekly data; Kim, Nelson, and Startz [29] argued that the mean reversion results are only detectable in pre-war data; and Richardson and Stock [44] and Richardson [45] reported that correcting for small-sample bias problems may reverse the Fama and French [16] and Poterba and Summers [43] results. Recently, there were furious discussions about whether there is mean reversion pattern in commodity investment returns; the essence of the Mean reversion concept is the assumption that both a financial instrument's high and low prices are temporary and that asset price will tend to move to the average price over time. From the perspective of an investor, when the current market price is less than the average price, the asset is considered attractive for purchasing, with the expectation that the price will rise; when the current market price is above the average price, the market price is expected to fall. In other words, deviations from the average price are expected to revert to the average, any force that pushes the energy price process back to the mean would imply negative autocorrelation at some time scale and would thus induce the systematic success of trading strategies (as in Figure 1.1). The concept of a process that returns to its mean is too general that many formal definitions can reproduce it with slight differences, either deterministic or stochastic. There is no existing universal measure of mean reversion and the definition that people believe investment professionals often struggle towards (a form involving in fact simultaneously both mean reversion and aversion) is not the same as the standard definition

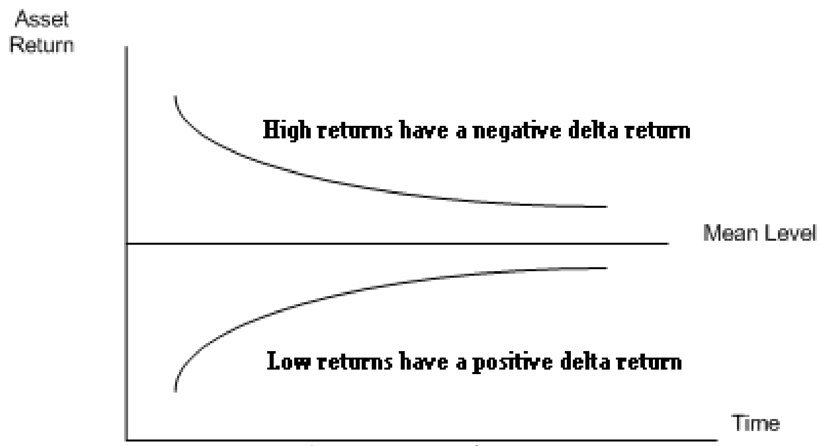

Figure 1.1: The concept of mean reversion

in time series analysis. Due to the multiple perspectives of the concept, there is a lack of precision in what is exactly mean by the term mean reversion. Also in real time, average price in mean-reversion models is time dependent, and there is no existing model accounting for this time-varying property of this important coefficient.

Energy products such as crude oil and Brent oil, are among the most liquid and actively traded commodities on energy capital markets. None of the modern industry could survive without energy products. As the price of a source of energy rises, it is likely to be consumed less and produced more by suppliers. This dynamic creates a downward bias on the prices of products. As the price of a source of energy declines, it is likely to be consumed more, but the production is likely to be less economically viable. This creates upward bias on the price. Mean reversion pattern is pervasive on the energy future contracts on commodity market.

The term structure of mean-reversion deserves being redefined with appropriate time-varying coefficients and a solution needs to be proposed to address this problem. What goes up must come down turned out to be a highly non-trivial fact about capital markets. The problem of this thesis is to provide a concrete solution to the Mean reversion modeling with time-dependent descriptive coefficients. A time-varying Mean reversion model is needed to account for the real time changes on moving mean values.

Given a variety of existing mean reversion models in the literature, there is a lack of model independent from the degree of mean reversion test, neither deterministic nor stochastic. On the other hand, all the existing models assume time invariant mean, speed and variance of mean reversion. The major problem of this thesis is therefore to address the time-varying property of mean-reversion models and provide a calibration approach to them. The problem also include making the time-invariant model independent from mean reversion testing. In the meanwhile, it is important to have a comparison of the proposed time-varying model with the proposed time-invariant mean-reversion model.

1.2 Motivation

The classical definition of mean-reversion is linked to autoregressive moving average model with parameter $(1, 1)$, $ARMA(p, q)$[5]. One important assumption is the weak stationarity of the time series. In order to have an accurate measure of the speed and extent of mean reversion, the classical definition here is slightly different from $ARMA(1, 1)$ model in terms of coefficients interpretation.

Consider a simple time-invariant autoregressive process of order one with drift:

$$x_t - x_{t-1} = a - bx_{t-1} + \varepsilon_t \tag{1.2.1}$$

where ε_t is a zero-mean white noise, i.e. identically independent normal distribution, and $a \in (0, 1)$. Assuming that $0 < b < 2$, otherwise the process is considered non-stationary. The expectation or unconditional mean of the process is

$$\mathbb{E}(x_t) = (1 - b)^{t-1}[\mathbb{E}(x_1) - \frac{a}{b}] + \frac{a}{b} \tag{1.2.2}$$

and the persistent parameter $1 - b$ governs the reversion to the mean

$\frac{a}{b}$. Intuitively, the shock ε_{t-1} enters x_t with weight $1-b$, it enters x_{t+1} with weight $(1-b)^2$ and so forth. That is, the fraction $1-b$ of the shock is carried forward per unit of time and hence the fraction b is washed out per unit of time. The inverse, $\frac{1}{b}$ is the average time for a shock to be washed out. It is the mean reversion time. This argument will be extended to time-varying version in modeling part. The variance of the mean reverting process is

$$\mathbb{V}\mathrm{ar}(x_t) = (1-b)^{2(t-1)}[\mathbb{V}\mathrm{ar}(x_1) - \frac{1}{1-(1-b)^2}] + \frac{\mathbb{V}\mathrm{ar}(\varepsilon_t)}{1-(1-b)^2} \quad (1.2.3)$$

When $\mid 1-b \mid < 1$, in the long run, the expectation converges to

$$\mathbb{E}(x_\infty) = \frac{a}{b}$$

and the variance converges to

$$\mathbb{V}\mathrm{ar}(x_\infty) = \frac{\mathbb{V}\mathrm{ar}(\varepsilon_t)}{(1-(1-b)^2)}$$

Besides, if the process is stationary (i.e. $b \in (0,2)$), there are

$$\mathbb{E}(x_t) = a + (1-b)\mathbb{E}(x_{t-1}) \quad (1.2.4)$$

$$\mathbb{E}(x_t - x_{t-1} \mid x_{t-1}) = a - b\mathbb{E}(x_{t-1}) \quad (1.2.5)$$

Intuitively, if $\mathbb{E}(x_{t-1})$ is below the long-term mean $\frac{a}{b}$, $a - b\mathbb{E}(x_{t-1})$ is positive. Hence $\mathbb{E}(x_t \mid x_{t-1})$ is expected to be higher than $\mathbb{E}(x_{t-1})$ because $\mathbb{E}(x_t - x_{t-1} \mid x_{t-1}) \geq 0$. On the other hand, if $\mathbb{E}(x_{t-1})$ is above the long-term mean, $\mathbb{E}(x_s \mid x_t), \forall s \geq t$ is decreasing towards the mean. As we could see in this model, one important assumption is the time-invariant property of mean, speed and variance. In practice, it makes the calibration process as trial and error based on various choices of the calibration window length. To better avoid this dependency, a time-varying model is necessary. On the other hand, all the existing model is itself a testing on mean reversion; together with the existing mean reversion tests, like unit root tests and Hurst exponent test, the mean reversion model has to rely on the degree of mean reversion given by various tests. This makes the model less convincing. A new model should be applied universally on extracting cycles which could be the mean reversion essence of a time series.

The main motivation of using wavelet decomposition technique is to extract the mean-reverting details hidden in a price time-series, which could be applied by one-factor Schwartz mean-reversion model more appropriately due to the nature of a cycle. On the other hand, all the existing models are cumbersome in calibration. One important factor is calibration has dependency on the length of historical information. A real time on-line estimation is needed for the simplicity of mean-reverting calibration. Therefore a second approach of using Bayesian modeling is worth being studied. In addition, to address the concern of using Bayesian approach in estimation, a comparison of time-varying model to time-invariant model needs to be re-examined to see the improvement of introducing time-varying coefficients.

1.3 Contributions

This thesis has provided a novel approach on mean-reversion modeling with wavelet decomposition techniques. A review of three main forms of mean reversion is firstly done and a formal mathematical definition of what most investment practitioners seem to mean by mean reversion, based on the correlation of returns between disjoint intervals, is proposed in Chapter 3. The main contributions of this thesis include an improvement of classical time-invariant mean reversion model based on wavelet decomposition techniques. The mean-reversion nature of asset price becomes divisible and is conquered on different small time series cycles, called details, after the wavelet-decomposition. A corresponding calibration methodology, using Maximum likelihood estimation and indirect inference is also proposed in the time-invariant mean-reversion model. Besides, another main contribution is the development of time-varying model based on linear state space analysis. A calibration methodology based on Kalman filter and Bayesian probability is also constructed in the time-varying model. Lastly, an application of both two mean reversion models is conducted on the energy future contracts.

1.4 Subsequent Chapters

Chapter 2 summarizes the definitions of mean reversion and the existing methodologies on detecting mean reversion. In addition, there is a concise introduction of wavelet decomposition method. Then in the third chapter, there are two models being analyzed, one is time-invariant stochastic model with wavelet transformation, the other model is time-varying state-space model. Subsequently, the chapter four discusses the application of the two models on energy future contracts on capital market. Lastly, chapter five gives the conclusion of the thesis and some further works of this research topic.

CHAPTER 2

Literature Review

2.1 Overview

In this chapter, a literature review is firstly done on energy derivative models with mean-reverting jumps and stochastic volatility. Apart from the models proposed and analyzed by scholars, two important definitions have been emphasized, namely the famous Granger and Joyeux model and Orhenstein-Uhlenbeck model. Then a collection of mean reversion testing methodologies is categorized with different problem-solving criteria. Certain procedures are provided in the most widely-used methods. Subsequently the chapter discusses the introduction of wavelet decomposition techniques, which is going to be applied into pre-modeling part of mean-reversion model. The current research gap is presented at the end of this chapter.

2.2 Mean reversion research methodologies

In recent years, the notion of mean reversion has attracted a considerable amount of attention in the Financial Economics. The term structure of futures prices is tested over the period January 1982 to December 1991, for which mean reversion is found in eleven different capital markets examined , and it is also concluded that the magnitude of mean reversion is large for crude oil and substantially less for precious metals [3]. One important reason is the proliferation of financial instruments linked to the price of financial asset on capital markets. Modeling the derivative price as mean-reverting stochastic process has provided a new systematic approach to the valuation of contingent claims and made the fair pricing issue easier. It is important that the models capture the empirical properties of asset price processes. Secondly, the extent to which financial assets exhibit mean-reverting behavior is crucial in building long-short trading strategies. Many financial institutions including hedge funds and proprietary trading firms are allowed to short-sell derivatives on financial markets, which urge them to seek for market neutral strategies, for instance mean reverting strategies. Balvers and Wu [1] found that strategies based on mean reversion typically yield excess returns of around 1.1-1.7 % per month which in turn outperform a random-walk based strategy. Thirdly, if the asset prices are predictable to some degree with mean-reverting modeling (normally there are measure on the speed of mean reversion and predicted mean level), the asset allocation problem can be considerably more interesting because the optimal investment strategy based on mean reverting model is path dependent [7]. This reduces the burden of investment manager to find out a closed-form allocation path. Most importantly, mean reversion has the appearance of a more scientific method of choosing asset buying and selling points than charting or traditional technique analysis. In typical technique analysis, the standard deviation of the most recent values (e.g., the last 20) is often used as a buy or sell indicator. In charting analysis, most asset reporting services offer moving averages for different periods such as 50 and 100 days. While reporting services provide the averages, identifying the high and low prices for the study period is still inaccurate, no extract numerical values for buying and selling points are derived in technique analysis. However, precise numerical values can be derived in mean reversion modeling from historical data to identify the buy/sell values.

Particularly, some asset classes, such as energy commodities are observed to be mean reverting. The energy commodity derivative market has strongly increased in recent years, both in trading volumes and the variety of offered products. The price of crude oil topped at around \$150 in July 2008 and dropped below \$40 by December 2008. The high and time-varying volatility of natural gas has reached 50% to 100%; similarly electricity has soared 100% to 500% in 2008. These huge price jumps and spikes with steep volatility smiles in future options have challenged the typical trend followers in the commodity market. Unlike financial assets, supply and demand for commodities are to a large extent influenced by production costs and consumer behavior. When prices are high, consumption will decrease and low-cost producers will enter the market. This leads to a decrease in prices. When prices are relatively low, consumers and producers will react vice versa, putting a upward pressure on prices. Additionally, the level of inventories plays an important role in determing the value for storable goods [28] [51]. The physical commodity owner decides whether to consume it immediately or store it for future disposal. Hence, the price of the commodity is the maximum of its current consumption and asset values [47]. The fluctuation of energy prices has stimulated renewed Mean Reverting modeling and application. As shown in Figure 2.1, the closed price of crude oil futures

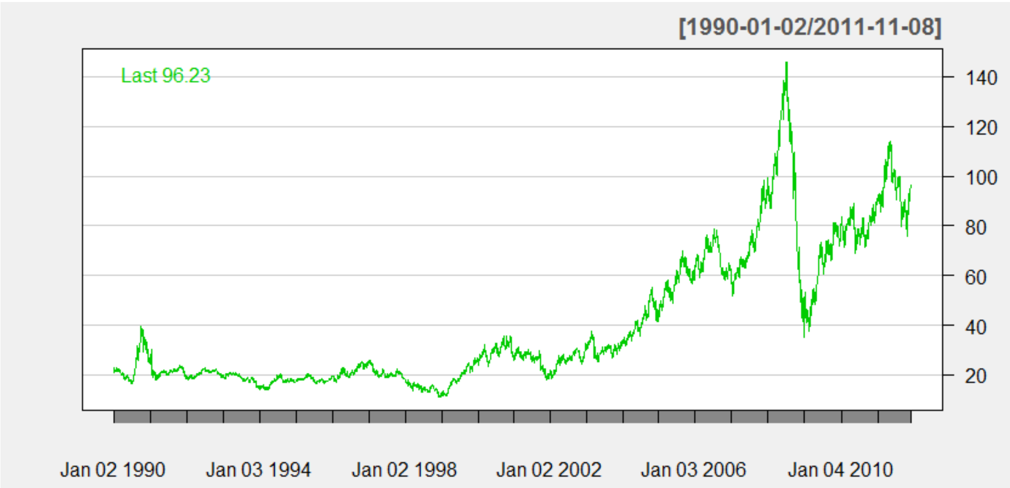

Figure 2.1: Crude Oil Futures Close Price on NYMEX with Front Month as maturity ranging from 1990 to 2010
source: Bloomberg Data retreived from IFS Commodity Derivatives FREE 15-Minute Delayed Pricing Service

experienced two main regime shift in the last 20 years. In the period before 2002, the price clearly shows a mean-reverting pattern. In 2002, the first regime-switching happened. From 2002 to 2008, the futures price increased sharply from 20 $ per barrel to 140 $ per barrel. Then due to the financail crisis, the price dropped to 40 $ per barrel in 2009, where I considered it as second regime-switching point. From 2009 onwards, the crude oil futures is experiencing another rising regime period. Figure 2.2 displays the oil price trend for a one-year period. Comparing Figure 2.2 and Figure 2.1, the models with mean-reverting oil price process could be appropriate for short-term oil futures, while in long-term models additional risk of regime switching should be included into analysis. The main reason for short-term price peaks and regime-switching

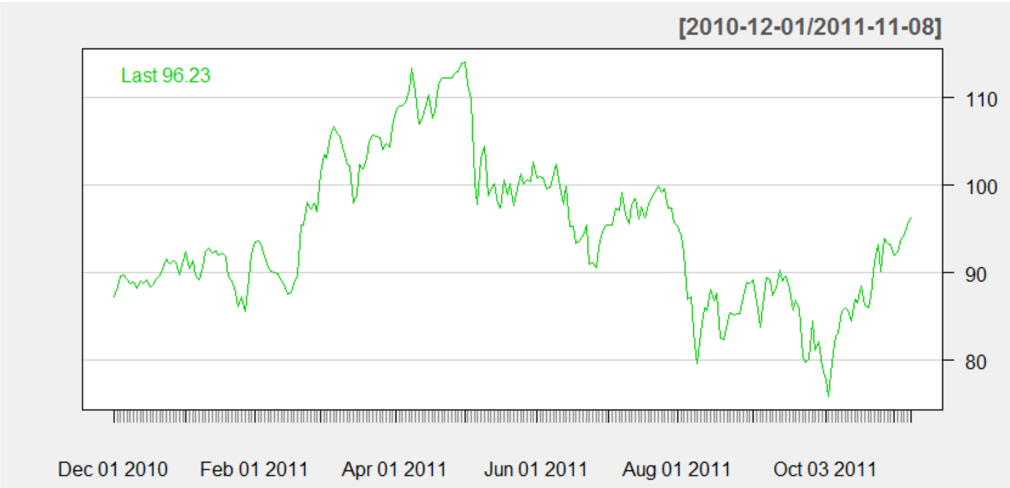

Figure 2.2: Crude Oil Futures Close Price on NYMEX with Front Month as maturity for one year
source: Bloomberg Data retreived from IFS Commodity Derivatives FREE 15-Minute Delayed Pricing Service

in recent years is supply and demand. On one hand, the current limit of the oil production capacity is fairly reached and there exists uncertainty about the remaining global oil resources. For instance, recently in the Southern China Sea there was furious discussion between China, Vietnam and Philippines on the sovereignty of potential petroleum under sea. On the other hand, considering oil demand, particularly the oil demand of China increased tremendously in last ten years. All of this has stimulated renewed modeling and application of mean reversion.

Historically, the majority of work on mean reversion modeling of energy future prices has been focused on the stochastic process used for the spot price and other key variables, such as interest rates and the convenience yield. Mean reversion is classically modeled by Ornstein-Uhlenbeck process. In the spirit of the Black-Scholes-Merton [4] formula, Schwartz [48] has proposed three model settings for the spot price process of commodities. In all three types, either directly in a price process of Ornstein-Uhlenbeck type as in his model 1 or indirectly through a subordinated convenience yield process as in models 2 and 3. Model 3 incorporates also stochastic interest rates. While model 2 and 3 are based upon standard arbitrage theory, model 1 is similar to Ross [46] in which the logarithm of the spot price of the commodity is assumed to follow a mean-reverting process. Besides, the Kalman filter methodology is applied to estimate the parameters in his model. Litzenberger and Rabinowitz [32] introduced a mean-reverting drift in the stochastic differential equation driving oil price dynamics. Later, Eydeland and German [15] introduced stochastic volatility into an Ornstein-Uhlenbeck process, namely the log of price follows an Ornstein-Uhlenbeck process with the square of volatility following the CIR process [10]. Early literature on jump modeling added state-independent compound Poisson jumps to Ornstein-Uhlenbeck process. Hilliard and Reis [25], Deng [13] have utilized jump diffusions respectively in their mean-reverting models specifically. More recently, German and Roncoroni [19] introduced a model with the jump direction dependent on the state, but the jump size is still state-independent.

In addition, seasonality is introduced in mean-revering modelling by including deterministic and periodic functions of time in model specification (time inhomogeneity). Moreover, some scholars have modeled futures curve directly, like Cortazar and Schwartz [9], Clelow and Strickland [8]. Mean reversion modeling on Jump-diffusion with CPP jumps was studied by Hilliard and Reis [25] and Crosby [11].

However, these approaches have three fundamental disadvantages. Firstly the key state variables such as the convenience yield is unobservable. Modeling the unidentifiable factors can only make the calibration step more complicated and it is not helpful to be implemented by practitioners. Secondly the future price curve is an endogenous function of the model parameters. Therefore it will not be necessarily consistent with the market observable future prices. Thirdly, the stochastic treatment of future prices is only applicable to the portfolio exhibiting stationarity.

In most of the time, the price time series of the two Financial instruments may not be stationary, but their price difference, the spreads exhibit stationarity if a common stochastic trend indeed exists between the two assets. The state-space modeling of mean-reverting spreads is analyzed and its parameter estimation is done under a hierarchical Bayesian framework using Markov chain Monte Carlo (MCMC) methods. Among the state-space modeling, linear dynamic systems are useful in financial application. Since the publication of the seminal work of Harrison and Stevens [22], the state space model have become an important time series analysis tool from Bayesian viewpoint, it can be represented as a system of equations specifying how observations of a process are stochastically dependent on the current process state and can be represented by how the process parameters evolve in time.

Pair trading was first appeared in 1987. Since that pairs trading has increased in popularity and has become a potential candidate to deal with mean reverting property of financial instrument.

2.3 Mean reversion models and testings

2.3.1 Overview

Several mean reversion detecting methodologies have evolved over the last twenty years. Pindyck [42] analyzed 127 years of data on crude oil. Using a unit root test, he showed that prices mean revert to stochastically fluctuating trend lines that represent long-run total marginal costs but are themselves unobservable. In section 2.3.4 and section 2.3.5, there is a detailed description of Unit root test. Section 2.3.5 also introduced Hurst Exponent test. All of them are used to justify to what extent of mean-reverting a time series is. Pindyck also found that during the time period of analysis, the random walk distribution for log-prices is a worse approximation for oil. Frankel and Rose [17] first applied the unit root test on the exchange market for the detection of mean reversion. Poterba and Summers [43] analyzed a modified variance ratio test based on the one proposed by Lo and MacKinlay [33]. Recently, due to the popularity of fraction in Finance, there are new detecting tools using fractional applicatoin - hurst exponent. Besides, backwardation is also an implication of mean reversion and it can be used as a predictor for mean-reverting spot prices [18].

2.3.2 Definition of Granger and Joyeux

Granger and Joyeux [21] suggested an important class of mean reversion models and with that they started the literature on long memory time series.

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_t \tag{2.3.1}$$

where ε_t is a zero-mean variate (different from classical definition these noises might be dependent) and $\alpha \in (0, 1)$.

The solution conditional on the state variable x_0 is given by

$$x_t = \frac{\alpha_0(1 - \alpha_1^t)}{1 - \alpha_1} + \alpha_1^t x_0 + \sum_{i=0}^{t-1} \alpha_1^i \varepsilon_{t-i} \tag{2.3.2}$$

and for large t and stationary autoregressive model with parameter one, $AR(1)$ where $0 < \alpha_1 < 1$ we have

$$x_t \simeq \mathbb{E}(x) + \sum_{i=0}^{t-1} \alpha_1^i \varepsilon_{t-i} \tag{2.3.3}$$

From the above expression the autocorrelation coefficients $\rho(k) = \alpha_1^k$ can be read directly. In general, it declines geometrically for stationary $ARMA(p, q)$ models [6].

Suppose x_t is an integrated $AR(1)$ model of order $d \in \mathbb{N}$, that is, the d^{th} difference series

$$y_t = \bigtriangledown^d x_t = (1 - L)^d x_t \tag{2.3.4}$$

where L is the lag operator, is $AR(1)$ without drift:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \tag{2.3.5}$$

Granger and Joyeux generalized the available theory for $d \in \mathbb{N}$ to $d \in (-0.5, 0.5)$, that is, to allow for fractional d. The shift operation is defined by the infinite binomial expansion

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots \tag{2.3.6}$$

The autocorrelation can be shown to have the order

$$\rho(k) \sim Ck^{2d-1} \text{when } k \to \infty \tag{2.3.7}$$

where $C > 0$. The decay is thus slower than geometrical series.

In addition, the process x_t reverts to its mean zero. A non-zero mean μ can be introduced to generalize the model.

$$y_t = (1-L)^d (x_t - \mu) \tag{2.3.8}$$

The case of $d \in (0.5, 1)$ is treated by differencing $x_t - \mu$ once:

$$y_t = (1-L)^{d-1}(1-L)(x_t - \mu) \tag{2.3.9}$$

so that for $y_t^{'} = (1-L)(x_t - \mu)$ the fractional parameter $d^{'} = d - 1 \in (-0.5, 0)$ and the theory applies.

2.3.3 Definition of Orhenstein-Uhlenbeck process

Mean reversion as a concept opposite to momentum should demonstrate a form of symmetry with respect to time since such a group of financial assets may be above its historical average approximately as often as below, which was predicted by the efficient-market hypothesis. Besides, a rigorous mathematical definition of mean-reverting stochastic process should demonstrate the unobserved phenomenon that an energy commodity future may hit zero and stay there forever. Recognizing a financial asset is overpriced is a rare and unconsciously difficult statement in continuous terms. It requires a definition on the fair price of asset in risk-neutral measure. The future movement of a mean-reverting time series can potentially be forecasted using mean-reverting models based on historical data. For the framework intended to demonstrate a tendency to remain near, or tend to return over time to a long-run average value, stochastic process is better than the first two models. For random walk, any shock is permanent and there is no tendency for price level to return to a constant mean over certain time. The changes of variance in asset price do not grow linearly with time as they would be if it was a random walk. Opposed to random walks (with drift), the Ornstein-Uhlenbeck Mean Reverting process does not exhibit an explosive behavior, but rather tends to fluctuate around the long-term mean level. Mathematically, the definition of an Orhenstein-Uhlenbeck process [38] adapted to mean reversion is the solution S_t of the following stochastic differential equation:

$$\mathrm{d}S_t = \lambda(\mu - S_t)\mathrm{d}t + \sigma \mathrm{d}B_t \tag{2.3.10}$$

where B_t is a standard Brownian motion on a risk-neutral probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and it is controlled by three parameters $\lambda, \mu, \sigma \in \mathbb{R}$. According to German [20], given a Markov diffusion process (X_t), the process exhibits mean reversion if and only if it admits a finite invariant measure. The Ornstein-Uhlenbeck process does admit a finite invariant measure, and this probability measure is Gaussian. Compared to the other definitions, the parameters in Ornstein-Uhlenbeck process has a more straight-forward understanding, λ is a measure of the speed for mean reversion, μ is considered as the fixed mean, and σ is the volatility.

2.3.4 Augmented Dickey-Fuller Test

Consider a simple general autoregressive process with p lags, i.e. $AR(p)$

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

Set $\beta_j = \sum_{i=j}^{p} \phi_i, \forall j \in 1, \cdots, p,$

$$y_t = \mu + \beta_1 y_{t-1} - \sum_{j=1}^{p-1} \beta_{j+1} \Delta y_{t-j} + \varepsilon_t$$

where $\Delta y_t = y_t - y_{t-1}$. In terms of lag operator $\phi(B)$, the $AR(p)$ process is written as $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$. Then it is straightforward to notice that $\phi(1) = 0$ if and only if $\beta_1 = 1$.

Therefore, a unit-root test can be formulated as testing the null hypothesis $H_0 : \beta_1 = 1$. The augmented Dickey-Fuller (ADF) statistics is $\frac{(\hat{\beta}_1 - 1)}{\hat{\sigma}(\hat{\beta}_1)}$, in the following texts, ADF is defined as augmented Dickey-

Fuller statistics.

2.3.5 Phillips-Perron Test

Phillips and Perron [40] have relaxed the identically independent distribution assumption on the noises ϵ_t, then a modified test statistics called Phillips-Perron statistics, or PP statistics was introducted to unit-test,

$$\mathrm{PP} = \{\mathrm{ADF}\sqrt{\hat{r}_0 - n(\hat{\lambda}^2 - \hat{r}_0)\sigma\hat{\beta}_1/2s}/\hat{\lambda}\}$$

where $\hat{r}_j = \frac{\sum_{t=j+1}^{n} \hat{\epsilon}_t \hat{\epsilon_{t-j}}}{n}$, $\lambda = \hat{r}_0 + 2\sum_{j=1}^{q}[1 - \frac{j}{q+1}]\hat{r}_j$, and s^2 is the OLS estimate of $\mathbb{V}\mathrm{ar}\{\epsilon_t\}$. Phillips and Perron have shown that PP has the same limiting distribution as ADF under Null assumption $H_0 : \beta_1 = 1$.

Davidson and MacKinnon [12] report that the Phillips-Perron test performs worse in finite samples than the augmented Dickey-Fuller test.

2.3.6 Hurst exponent Test

In the original definition given by Mandelbrot and Van Ness [35], a unique Hurst exponent value characterizes both a fractional Brownian Motion and its increments. The Hurst Exponent occurs in several areas of applied mathematics, including fractals and chaos theory, long memory processes and spectral analysis. Hurst Exponent estimation has been applied in areas ranging from biophysics to computer networking. The Hurst Exponent is directly related to the fractal dimension of a process, which gives a measure of the roughness of the process. There are many ways to estimate the Hurst Exponent, and the most common way of estimation is to run a linear regression.

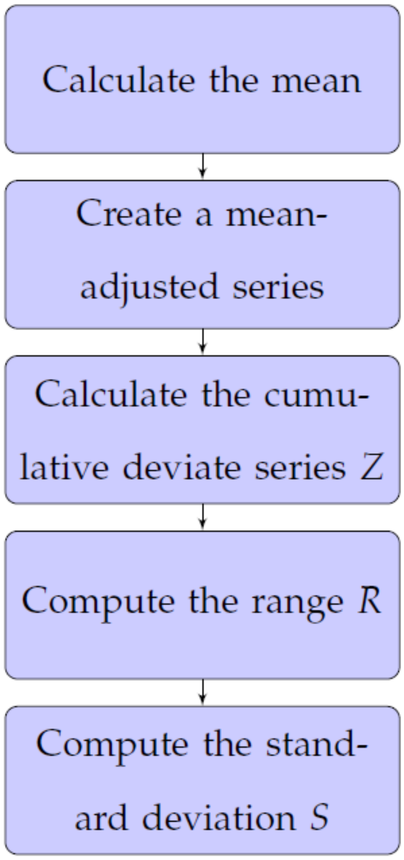

Figure 2.3: Procedure for Hurst Exponent Test

For a time series of length N, we can divide it into A subgroups of length n, where $n << N$. For each subgroup, we will perform the five steps below. Without loss of generality, considerring the subgroup with index one.

1. Calculate the mean

$m = \frac{1}{n} \sum_{t=1}^{n} X_t$

2. Create a mean-adjusted series

$Y_t = X_t - m \quad \text{for } t = 1, 2, \ldots, n$

3. Calculate the cumulative deviate series Z

$Z_t = \sum_{i=1}^{t} Y_i \quad \text{for } t = 1, 2, \ldots, n$

4. Compute the range R

$R(n) = \max\left(Z_1, Z_2, \ldots, Z_n\right) - \min\left(Z_1, Z_2, \ldots, Z_n\right),$

5. Compute the standard deviation S

$S(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2}$

Calculate the rescaled range $\frac{R(n)}{S(n)}$ and average over all the partial time series(subgroups) of length n. We should be able to have one expression for a fixed n, $(\frac{R}{S})_n := \frac{1}{A}\sum_{a=1}^{A}\left(\frac{R}{S}\right)_{an}$. Moreover, it should be able to run a linear regression $log(\frac{R(n)}{S(n)})$ over $log(n)$ to estimate Hurst Exponent given that $(\frac{R}{S})_n = Cn^H$, and to choose k different values of n in order to generate enough data.

2.4 Wavelet Transformation

The term wavelets itself was coined in the geophysics literature by Morlet et al [36]. In many situations, wavelets often offer a kind of insurance: they will sometimes work better than certain competitors on some classes of problems, but typically work as well as other methods for the rest of the problem categories. For example, one-dimensional nonparametric regression has mathematical results of this type [37]. On time-series decomposition techniques, wavelet has structure extraction, localization, efficiency and Sparsity advantage compared to other techniques.

2.4.1 Discrete Wavelet Transform

The Discrete Wavelet Transform is developed from Continuous Wavelet Transform with discrete inputs [2] [34], but the mathematical derivation is simpler. However, there is no simple formula for the relationship between input and output, but people can use the hierarchical structure to describe the concept.

2.4.2 1-D DWT

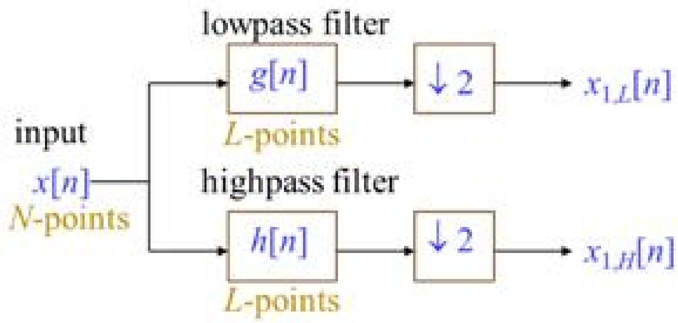

Figure 2.4: Block Diagram of Filter Analysis

The relation equations are represented as follows:

$$x_{1,L} = \sum_{k=0}^{K-1} x[2n-k]g[k]$$

$$x_{1,H} = \sum_{k=0}^{K-1} x[2n-k]h[k]$$

where $g[n]$ is a low-pass filter like scaling function, and $h[n]$ is a high-pass filter like mother wavelet function, in a two stage structure it looks like this:

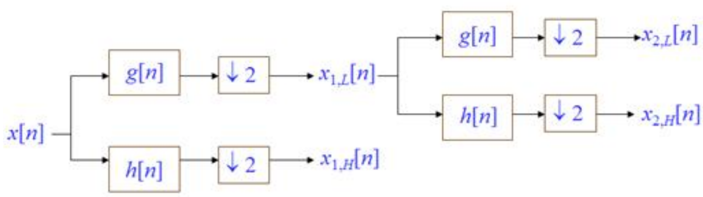

Figure 2.5: A Two Stage Structure

The relation equations can be modified as follows:

$$x_{\alpha,L} = \sum_{k=0}^{K-1} x_{\alpha-1,L}[2n-k]g[k]$$

$$x_{\alpha,H} = \sum_{k=0}^{K-1} x_{\alpha-1,H}[2n-k]h[k]$$

Hence, we can decompose the high-pass component to gain more details, and the number of output bits will still be close to the number of input bits if the length of input is much larger than the length of the filter. The decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input, so the frequency resolution has been doubled.

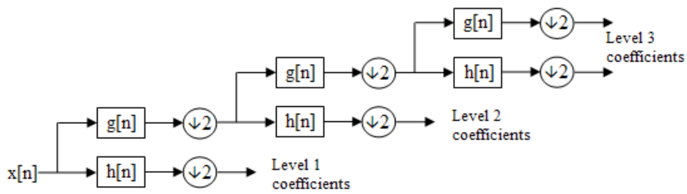

Figure 2.6: Three-stage 1-D DWT

2.5 Research Gap

The Granger-Joyeux and the Orhenstein-Uhlenbeck models have been used extensively by researchers and industry practitioners for the purpose of understanding mean reversion [6]. However, they assume time-independent degree of mean-reversion. As a result, users of the models have to find out the degree of mean-reversion of a price-time series through certain testing procedures at different points in time. But the

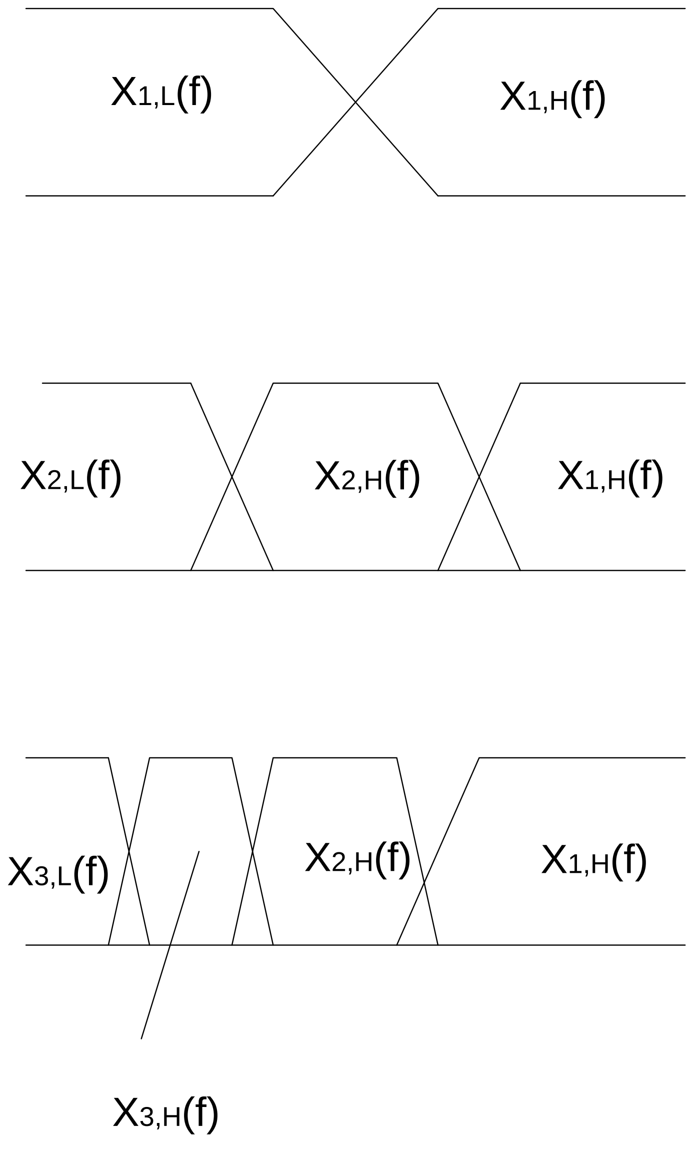

Figure 2.7: Three-stage 1-D DWT in frequency domain

problem with the testing procedures is that the different procedures generally produce different values for the degree of mean-reverting for the same price-time series. There are two possible ways of solving the problem: either proposing a time-varying model which will allow us to totally abandon the dependence on the testing procedures; or continuing using the time-independent models but with a technique on extracting the correct length of historical rolling window other than the testing procedures to determine the degree of mean-reversion – the wavelet decomposition technique will just be a feasible way.

CHAPTER 3

Proposed Models

3.1 Overview

In the mean-reversion modeling section, there are two mean-reversion models being proposed, analyzed and compared. The research gap exists in the area of modeling which is independent from mean-reversion testing and the area of time-dependent coefficients, i.e. time-varying mean, speed, and variance of mean reversion model. The first model is a time-invariant stochastic framework, based on the famous one factor stochastic Schwartz modeling but applied on extracted small cycles after wavelet decomposition. Wavelet decomposition is a methodology of decomposing time series into cycles and trends. As a de-noising method, wavelet filter was an useful tool to track the cycles of the price movements which can be modeled by mean-reversion. This technique enables independent usage of mean-reversion modeling from mean-reversion testings. The other model is to address the lack of time dependency in mean reversoin modeling, i.e. time-varying state space dynamic linear model. Bayesian approach with variance gamma assumption was applied on the on-line calibration of time-varying mean reversion model. The advantage of state space model is the on-line estimation of time-varying mean reversion parameters.

3.2 Time Invariant Model

3.2.1 Wavelet-Schwartz model

Assume the risky future contracts follow a mean-reverting pattern as shown previously (Equation 2.3.10) in definition section with time-invariant speed of mean reverting, long term mean and variance. Certain constraints and modification should be imposed. The first constraint of a price process is the non negativity. If S_t is the price of a risky asset at time t, the one factor Schwartz model [48] is written as

$$\mathrm{d}S_t = \lambda(\mu - \ln S_t)S_t\mathrm{d}t + \sigma S_t\mathrm{d}B_t \tag{3.2.1}$$

where B_t is the standard Brownian motion starting from 0, λ is the speed of mean-reversion, μ is the equilibrium level, and σ is the measure of process volatility[1]. Applying Itô's formula, it is found that $X_t = \ln S_t$ is an Ornstein-Uhlenbeck process with modified coefficients. The main reason to treat the drift part as a difference between long term mean and the log of asset prices is to prevent stochastic process from going negative which is undesirable.

[1]see Appendix for the derivation of the parameters' meanings

The stochastic differential equation for the stochastic process X_t is therefore written as

$$\mathrm{d}X_t = \lambda(\mu - \frac{\sigma^2}{2\lambda} - X_t)dt + \sigma \mathrm{d}B_t \tag{3.2.2}$$

The general solution[2] to the Equation 3.2.2 of a log Ornstein-Uhlenbeck process X_t is

$$X_t = e^{-\lambda t}X_0 + (\mu - \frac{\sigma^2}{2\lambda})(1 - e^{-\lambda t}) + \int_0^t \sigma e^{\lambda(s-t)} dB_s \tag{3.2.3}$$

The variable X_t is Normally distributed with two moments

$$\mathbb{E}[X_t] = e^{-\lambda t}X_0 + (\mu - \frac{\sigma^2}{2\lambda})(1 - e^{-\lambda t}) \tag{3.2.4}$$

and

$$\mathbb{V}\mathrm{ar}[X_t] = \frac{\sigma^2}{2\lambda}(1 - e^{-2\lambda t}) \tag{3.2.5}$$

The half-life $H = \frac{\ln 2}{\lambda}$ is defined as the time for the expected value of X_t to reach the intermediate (middle) price between the current value X_0 and the equilibrium level $\mu - \frac{\sigma^2}{2\lambda}$, to which the mean reverting process converges in the long run. It is consistent to the meaning of λ that higher of the speed will lead to a smaller value of half-life.

In discrete settings, it is common to treat mean reverting stochastic process as an autoregressive and moving average model with parameter (1,1), which is both applicable to the calibration and simulation. Moreover, it is shown in Chapter 1 (Equation 1.2.1) that $ARMA(1,1)$ coefficients' definitions have to be adjusted and adapted to the system so that the parameter in time-invariant model makes sense in the environment of

[2]see Appendix for the derivation of the explicit solution

mean reversion. It is noticed that in the setting of the three dimensional parameter $\theta = (\lambda, \mu, \sigma)$, it is treated as time-invariant, i.e. it depends on the scope of data history and the rolling window length.

3.2.2 Simulation

In order to calibrate the time-invariant coefficients describing the mean-reverting property, it is important to be able to simulate the mean reverting process in discrete terms. It is necessary to treat the risky asset in a discrete time frame, with Δt (the difference between time t_{k+1} and t_k) sufficiently small. In a close-to-close envrionment, Δt is one-day. In high frequency trading, Δt could be as small as 0.01 seconds.

$$X_{t_{k+1}} = e^{-\lambda\Delta t} X_{t_k} + (1 - e^{-\lambda\Delta t})(\mu - \frac{\sigma^2}{2\lambda}) + \sigma\sqrt{\frac{1 - e^{-2\lambda\Delta t}}{2\lambda}}\epsilon_t \qquad (3.2.6)$$

where $\epsilon_t \sim \mathcal{N}(0,1)$, $\Delta t = t_{k+1} - t_k$. Based on this sample path simulation equation, it is easy to generate a series of mean reverting asset prices with different mean-reverting degrees through a linear combination of random walk and Ornstein-Uhlenbeck process. Figure 3.1 is a pure mean reversion simulation with initial value $S_0 = 3$, $\mu = 1$, $\lambda = 3$ and $\sigma = 0.5$. The two out-most lines are the 2 times standard error, and the middle line is the long time mean. While in Figure 3.2 is the 0 % mean reversion simulation with the same initial value as above. The graphical representation shows that a random walk does not show too much mean reverting pattern. There are trends in random walk which can be extracted with various detrending methods such as wavelet decomposition methods. While in Figure 3.3 is a 50% mean reversion sim-

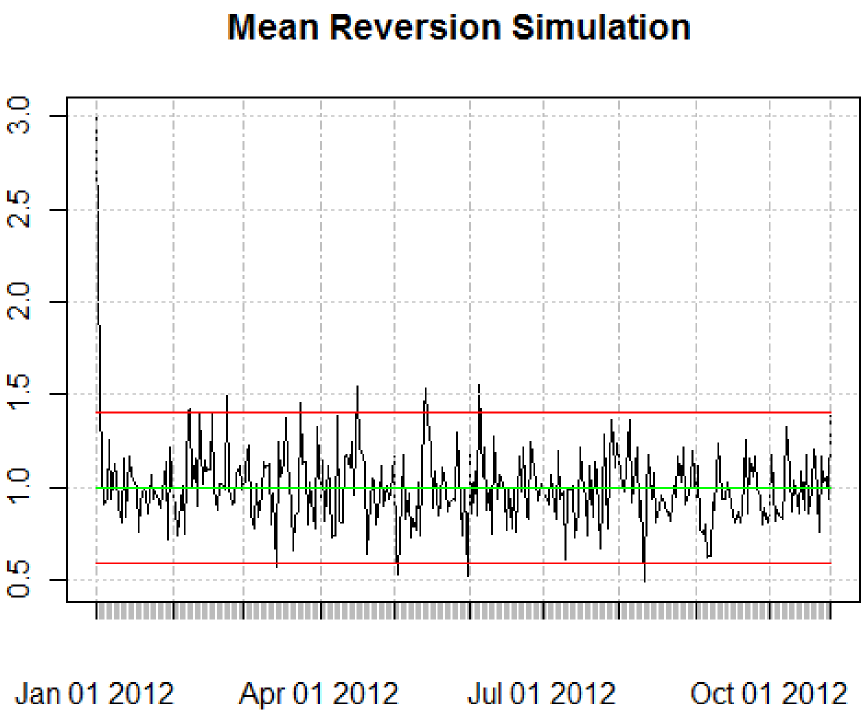

Figure 3.1: 100 percent mean reversion
R simulation

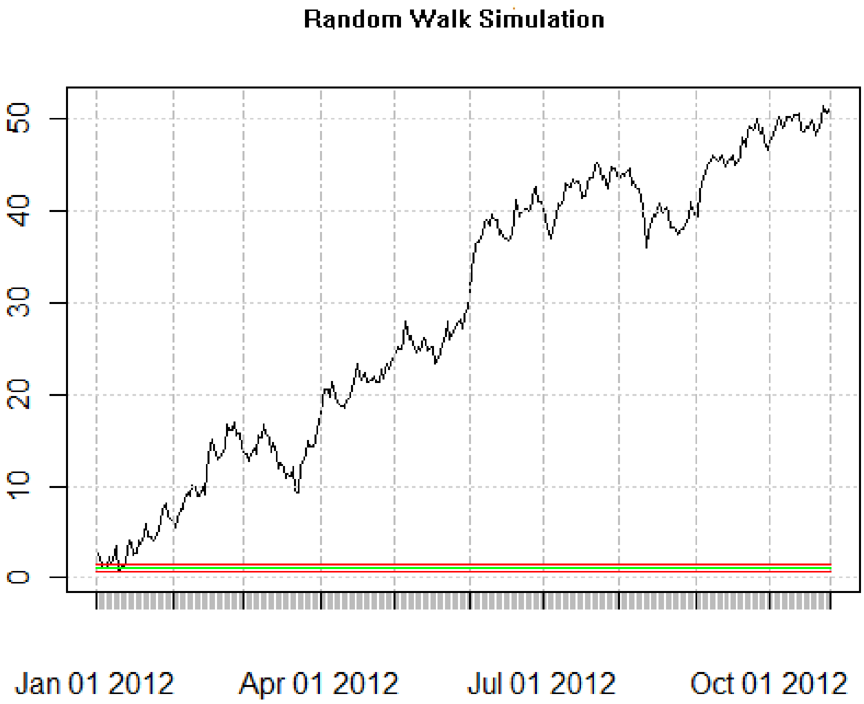

Figure 3.2: 0 percent mean reversion
R simulation

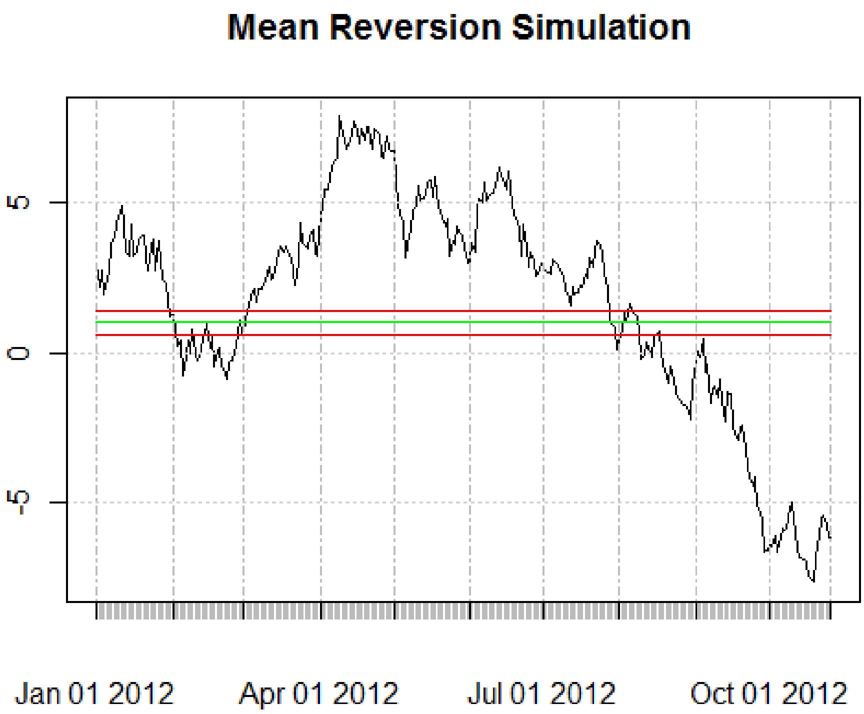

Figure 3.3: 50 percent random walk
R simulation

ulation. It has the same parameter setting as the Figure 3.1. The graph looks more mean reverting than a random walk but less mean reverting than Figure 3.1 as the data go beyond the boundaries. Therefore, to accurately capture mean-reverting phenomenon, filtering technique is critical as it is undesirable to see the noise affects the calibration result of mean-reverting models.

Therefore, in reality a more realistic model is applied to pure mean-reverting time series.

$$y_t = d * S_t + (1-d) * \epsilon_t \tag{3.2.7}$$

y_t is the price value observed, as the sum of a pure mean reverting process and random noise with weightage coefficients d. Besides, it is straightforward that $(X_{t+1} - X_t) \mid X_t$ has mean $(e^{-\lambda \Delta t} - 1)(X_t - \mu)$ and variance $\sigma^2(\frac{1-e^{-2\lambda\Delta t}}{2\lambda})$. If $\lambda \Delta t$ is small, there is $\mathbb{E}[(X_{t+1} - X_t) \mid X_t] = \lambda \Delta t(\mu - X_t)$ and $\mathbb{V}\mathrm{ar}[(X_{t+1} - X_t) \mid X_t] = \sigma^2(\Delta t)$.

3.2.3 Wavelet Decomposition

Based on the Discrete Wavelet Transformation techniques introduced in Percival and Walden [39], the relationship between the smoothies and the details after wavelet decomposition is as follow:

$$\begin{aligned} y &= d_1 + s_1 \\ &= d_1 + d_2 + s_2 \\ &= d_1 + d_2 + d_3 + s_3 \\ &= d_1 + ... + d_i + s_i \end{aligned} \tag{3.2.8}$$

Here y is the original time series. In the literature, the choice of wavelet filters are abundant, for instance Daubechies, Least Asymetric, Best Localized and Coiflet [39]. In this modelling part, the most famous Haar wavelet (it is also called Daubechies filter with coefficient 4) is chosen with $h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}}$, $h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}}$,$h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}$ and $h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}$.

With a pre-determined decomposing degree i, which is defined in Section 2.4 of Chapter 2, the following equation links the price time series y to the filter.

$$y = A^T W = \sum_{k=1}^{i} d_k + s_i \tag{3.2.9}$$

The $N/2$ wavelet coefficients for unit scale are defined as

$$A_{1,t} = \sqrt{2} B_{1,2t+1}, t = 0, \cdots, \frac{N}{2} - 1$$

with

$$\sqrt{2} B_{1,t} = \sum_{l=0}^{3} y_{t-l|N}, t = 0, \cdots, N-1$$

where $y_{t-l|N}$ is defined as the $(t-l)$ mod N th element in the time series y.

Hence,

$$W_{1,t} = \sum_{l=0}^{3} h_l y_{(2t+1-l)|N} = \sum_{l=0}^{N-1} h_l^o y_{(2t+1-l)|N}, t = 0, \cdots, \frac{N}{2} - 1 \quad (3.2.10)$$

where $\{h_l^o\}$ is periodized version of $\{h_l\}$ with length N.

Subsequently, the remaining $\frac{N}{2} - 1$ rows of matrix A can be circularly shifted by $A_{t,k}^T = [U^{2t} A_{0,k}]^T$ where U is $N \times N$ circular shift matrix or unit delay operator with the property $U^{-1} = U^T$.

Wavelet filters are a special type of linear filters that have been developed in engineering to account for non-stationary data and fat tails and potential long-range dependencies. Applying the wavelet transformation as introduced above, the details of the time series are extracted if the order i is given. In the Chapter 4, there is a detailed analysis on different choice of i and the cumulative sum of them which were fitted with mean reversion model.

3.2.4 Calibration

However, fitting or calibration of such models is not easy to come by, if practitioners successfully detect the mean-reverting time series, one way to estimate the parameters is to use Maximum Likelihood Estimations. Alternative methods include least squares regression of discrete autoregressive versions of the Ornstein-Uhlenbeck Meam Reverting model, methods of moments, and simulation based indirect inference methods. The conditional probability density function f_k of X_{t_k} conditional on

$X_{t_{k-1}}$is given by

$$f_k(x_{t_k};\bar{\mu},\lambda,\sigma) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2\lambda}[1-e^{-2\lambda(t_k-t_{k-1})}]}}$$

$$\exp\{-\frac{[x_{t_k}-\bar{\mu}-(x_{t_{k-1}}-\bar{\mu})e^{-\lambda(t_k-t_{k-1})}]^2}{2\frac{\sigma^2}{2\lambda}[1-e^{-2\lambda(t_k-t_{k-1})}]}\} \tag{3.2.11}$$

where, $\bar{\mu} = \mu - \frac{\sigma^2}{2\lambda}$. Therefore the log likelihood function $\mathcal{L} = \sum_{k=1}^{N}\log(f_k)$ is written as

$$\mathcal{L} = -\frac{N}{2}\log[\frac{\sigma^2}{2\lambda}] - \frac{1}{2}\sum_{k=1}^{N}\log[1-e^{-2\lambda(t_k-t_{k-1})}]-$$

$$\frac{\lambda}{\sigma^2}\sum_{k=1}^{N}\frac{[x_{t_k}-\bar{\mu}-(x_{t_{k-1}}-\bar{\mu})e^{-\lambda(t_k-t_{k-1})}]^2}{1-e^{-2\lambda(t_k-t_{k-1})}} \tag{3.2.12}$$

The first order conditions for maximum likelihood estimation require the gradient of the log likelihood to be equal to zero. In other words, the maximum likelihood estimators $\widehat{\bar{\mu}}, \widehat{\lambda}$ and $\widehat{\sigma}$ must satisfy the following conditions:

$$\frac{\partial\mathcal{L}}{\partial\bar{\mu}}\mid_{\widehat{\bar{\mu}}}= 0$$

$$\frac{\partial\mathcal{L}}{\partial\sigma}\mid_{\widehat{\sigma}}= 0$$

$$\frac{\partial\mathcal{L}}{\partial\lambda}\mid_{\widehat{\lambda}}= 0$$

The solution to this non-linear system of equations can be approximated through some simple analytical manipulations of the first order conditions.

$$\frac{\partial\mathcal{L}}{\partial\bar{\mu}} = \frac{2\lambda}{\sigma^2}\sum_{k=1}^{N}\frac{x_{t_k}-\bar{\mu}-(x_{t_{k-1}}-\bar{\mu})e^{-\lambda(t_k-t_{k-1})}}{1+e^{-\lambda(t_k-t_{k-1})}} \tag{3.2.13}$$

Therefore,

$$\widehat{\bar{\mu}} = \frac{\sum_{k=1}^{N} \frac{x_{t_k} - x_{t_{k-1}} e^{-\widehat{\lambda}(t_k - t_{k-1})}}{1+e^{-\widehat{\lambda}(t_k - t_{k-1})}}}{\sum_{k=1}^{N} \frac{1-e^{-\widehat{\lambda}(t_k - t_{k-1})}}{1+e^{-\widehat{\lambda}(t_k - t_{k-1})}}} \tag{3.2.14}$$

$$\widehat{\mu} = \frac{\sum_{k=1}^{N} \frac{x_{t_k} - x_{t_{k-1}} e^{-\widehat{\lambda}(t_k - t_{k-1})}}{1+e^{-\widehat{\lambda}(t_k - t_{k-1})}}}{\sum_{k=1}^{N} \frac{1-e^{-\widehat{\lambda}(t_k - t_{k-1})}}{1+e^{-\widehat{\lambda}(t_k - t_{k-1})}}} + \frac{\widehat{\sigma}^2}{2\widehat{\lambda}} \tag{3.2.15}$$

and

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2\lambda}{\sigma^3} \sum_{k=1}^{N} \frac{[x_{t_k} - \bar{\mu} - (x_{t_{k-1}} - \bar{\mu}) e^{-\lambda(t_k - t_{k-1})}]^2}{1 - e^{-2\lambda(t_k - t_{k-1})}} \tag{3.2.16}$$

$$\widehat{\sigma} = \sqrt{\frac{2\widehat{\lambda}}{N} \sum_{k=1}^{N} \frac{[x_{t_k} - \widehat{\bar{\mu}} - (x_{t_{k-1}} - \widehat{\bar{\mu}}) e^{-\widehat{\lambda}(t_k - t_{k-1})}]^2}{1 - e^{-2\widehat{\lambda}(t_k - t_{k-1})}}} \tag{3.2.17}$$

For estimator of λ, $\widehat{\lambda}$, simply minimize the following expression

$$-\frac{n}{2}\log[\frac{\widehat{\sigma}(\lambda)^2}{2\lambda}] - \frac{1}{2}\sum_{k=1}^{N} \log[1 - e^{-2\lambda(t_k - t_{k-1})}]$$

$$-\frac{\lambda}{(\widehat{\sigma}(\lambda))^2} \sum_{k=1}^{N} \frac{[x_{t_k} - \widehat{\mu}(\lambda) - (x_{t_{k-1}} - \widehat{\mu}(\lambda)) e^{-\lambda(t_k - t_{k-1})}]^2}{1 - e^{-2\lambda(t_k - t_{k-1})}} \tag{3.2.18}$$

To simplify the notation, let y denote the pricing sequence we observed from S_2^i to S_N^i, and x that we observed from S_1^i until S_{N-1}^i, then we have the following results.

$$\widehat{\bar{\mu}} = \frac{(\sum y)\cdot(\sum x^2) - (\sum x)\cdot(\sum xy)}{[N\cdot(\sum x^2 - \sum xy) - (\sum x)^2 + (\sum x)\cdot(\sum y)]} \tag{3.2.19}$$

$$\widehat{s^2} = \frac{\sum y^2 - 2e^{-\widehat{\lambda}\Delta t}\sum xy + e^{-2\widehat{\lambda}\Delta t}\sum x^2 - 2\widehat{\bar{\mu}}(1 - e^{-\widehat{\lambda}\Delta t})(\sum y - e^{-\widehat{\lambda}\Delta t}\sum x)}{N} \tag{3.2.20}$$

$$+\widehat{\mu}^2(1-e^{-\widehat{\lambda}\Delta t})^2$$

$$\widehat{\sigma} = \sqrt{2\widehat{s}^2 \frac{\widehat{\lambda}}{1-e^{-2\widehat{\lambda^2}}}} \tag{3.2.21}$$

$$\widehat{\lambda} = -\frac{\ln(\frac{\sum xy - \widehat{\mu}\sum x - \widehat{\mu}\sum y + N\widehat{\mu}^2}{\sum x^2 - 2\widehat{\mu}\sum x + N\widehat{\mu}^2})}{\Delta t} \tag{3.2.22}$$

To have a better esimation of λ, Phillips and Yu [41] proposed a modified method called jackknife technique, whereby λ is estimated over the whole sample, and denoted by λ_T, as well as m equal partitions of the data, dividing the time period in parts, call the $\lambda_1, \cdots, \lambda_m$. The bias in the estimate is greatly reduced if the modified λ is introduced.

$$\lambda_{jack} = \frac{m}{m-1}\lambda_T - \frac{\sum_{i=1}^{m}\lambda_i}{m^2-m}$$

Once the maximum likelihood estimator θ^{mle} is obtained, the next step is to find the minimizer of the following expression

$$||\frac{1}{N}\sum_{i=1}^{N}\theta_i^{mle} - \theta^{mle}||$$

This minimizer is called the indirect inference estimator, where θ_i^{mle} is the maximum likelihood estimator of i^{th} simulation based on θ^{mle}. Compared to maximum likelihood estimator, indirect inference estimator has an improved finite-sample performance [52].

3.2.5 Summary and limitation in time-invariant model

The time-invariant model has utilized the Wavelet-decomposition techniques, which transformed the stationary price time series into different degrees of cycles and trends. Applying the the one-factor Schwartz model on different cycles or cumulative cycles will address the problem of mean-reversion modeling to the original time series in different depth. One big limitation in the model is the assumption of time-invariant mean, speed and variance. This could make the calibration of coefficients sensitive to the choice of historic window length, but it could be more widely used as wavelet-decomposition is applicable to any stationary time series with the length equals to the power of 2. To make the model more adaptive to the price evolution, a time-varying state space model will relax this assumption.

3.3 Time Varying State Space Model

Time-varying models have a more realistic application since it is an adaptive process. But it is much more difficult to tackle with than the time-invariant model introduced in previous section; we have shown that stochastic mean reversion model keeps the coefficients time-invariant such that the calibration of the coefficient is easier than time varying models. Mathematically, assuming y_t is the price process, then $y_t = A + By_{t-1} + \varepsilon_t$. It is quite uncommon in reality as economic growth and important political issues can both trigger heavily regime-switch on the coefficient A and B. It is reasonable to treat both two coefficients time dependent.

In the case of time-varying speed, mean and volatility, state-space model is a more relevant and useful tool to mean-reversion analysis. In recent years there has been an increasing interest in the application of state space models in time series analysis. State space models consider the pricing time series as the output of a dynamic system perturbed by random distributions. They allow a natural interpretation of a time series as the combination of several components, such as trend, seasonal and regressive components. At the same time, they have an elegant and powerful probabilistic structure, offering a flexible framework for financial applications. Computations can be implemented by recursive algorithms. The problems of estimation and forecasting are solved by recursively computing the conditional distribution of the quantities of interest, given the available information. In this sense, they are quite naturally treated within a Bayesian framework.

3.3.1 Model Identification

Before moving on to the time varying case, consider a stochastic process x_t which follows

$$x_t = A + Bx_{t-1} + \epsilon_t$$

and

$$y_t = x_t + \omega_t$$

where $\epsilon_t \sim N(0, C^2)$, $\omega_t \sim N(0, D^2)$ are independent white noises, $\omega_t \perp x_t \forall t$ and $-1 < B < 1$, $A, B, C, D \in \mathbb{R}$

By combining the above two equations, the observation equation can be also written as

$$y_t = A + By_{t-1} + \psi_t$$

where $\psi_t = \omega_t - B\omega_{t-1} + \epsilon_t$, the noise ψ_t is normally distributed $N(0, \sigma^2)$

with fixed variance σ^2, and $\sigma^2 = D^2 + B^2D^2 + C^2$

It is an $AR(1)$ model with fixed parameters A,B and σ^2

In reality A and B can be both time-dependent in very long time scale, and σ^2 is uncertain conditionally as well. Knowing more information, i.e. more observations on historical price movement y_t, leads to updated estimation of the uncertain volatility σ^2. It is obvious that this approach requires knowledge of Bayesian estimation and forecasting. Therefore, a more dynamic but complicated model is required to describe the moving mean reverting property of financial assets.

3.3.2 Time-varying Formulation

This formulation is mainly based on the dynamic linear model from West and Harrison [50]. Specifically, in the language of mean reversion, for each time t, the general univariate dynamic linear model is defined by three equations: price observation, state evolution and initial information.

⋆ Price Observation Equation:

$$y_t = F_t^{'}\theta_t + \psi_t \tag{3.3.1}$$

where $\psi_t \sim \mathcal{N}(0, \sigma^2)$, and

$$F_t = \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix}$$

$$\theta_t = \begin{pmatrix} A_t \\ B_t \end{pmatrix}$$

In scalar form, it is written as

$$y_t = A_t + B_t y_{t-1} + \psi_t$$

As we can see clearly in the above equation, y_t represents the price of mean-reverting asset. Both A and B are time variant with index t, it is called the state variable in our model. ψ_t is a white noise with random variance. In the subsection Bayesian Framework, there will be an explanation on the choice of probability distribution for σ^2 which facilitates the forecasting distribution.

⋆ Systems equation(State Evolution equation):

$$\theta_t = \Phi\theta_{t-1} + v_t \tag{3.3.2}$$

Systems equation defines how the hidden state variable θ_t evolves over time:

$$\Phi = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix}$$

$\Phi \in \mathbb{R}^{2\times 2}$ is a parametric matrix with fixed coefficients ϕ_1 and ϕ_2

$$v_t = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

$v_t \sim N_2(0, \sigma^2 V_t)$ is the noise vector in state evolution equation.

In the section of induction, it will be shown that V_t has the following specific diagonal form

$$V_t = \begin{pmatrix} \frac{1-\delta_1}{\delta_1}\phi_1^2 P_{11,t} & 0 \\ 0 & \frac{1-\delta_2}{\delta_2}\phi_2^2 P_{22,t} \end{pmatrix}$$

In scalar form, it is written as

$$A_t = \phi_1 A_{t-1} + v_{1t}$$

$$B_t = \phi_2 B_{t-1} + v_{2t}$$

where $v_{1t} \sim \mathcal{N}(0, \sigma^2 \frac{1-\delta_1}{\delta_1}\phi_1^2 P_{11,t})$ $v_{2t} \sim \mathcal{N}(0, \sigma^2 \frac{1-\delta_2}{\delta_2}\phi_2^2 P_{22,t})$

ϕ_1 and ϕ_2 are the $AR(1)$ coefficients of A_t and B_t respectively, usually being assumed to be between $(-1, 1)$ so that A_t and B_t can be weakly stationary processes.

⋆ Initial information:

$$(\theta_1 | \sigma^2) \sim N_2(m_1, \sigma^2 P_1) \tag{3.3.3}$$

where

$$m_1 = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$$

$$P_1 = \begin{pmatrix} p_{11,t=1} & 0 \\ 0 & p_{22,t=1} \end{pmatrix}$$

and $p_{11,t}, p_{22,t}$ are the elements of matrix P_t, both of them are real numbers.

In order to make the process adaptive to a Bayesian approach, an assumption was made on σ^2, it is Inverse Gamma distributed, i.e. $\sigma^2 \sim \mathcal{IG}(n_1/2, d_1/2)$. In the language of joint distribution, (θ_1, σ^{-2}) follows Normal Gaussian distribution[3] such that $(\theta_1|\sigma^2) \sim \mathcal{N}(m_1, \sigma^2 P_1)$. Basically, the unobservable hidden Markov Chain, state process θ_t, has marginal student t-distribution. y_t is an imprecise measurement of θ_t, i.e. price observation is a linear transformation of state variable, in other words the observed process is a noisy realization of the hidden state process.

This state space modeling is a flexible modeling structure that directly embodies temporal movement in data defining qualities. It will be shown in next section, it defines simple rules for the management of uncertainties, based on the laws of probability. The analysis with many different but interacting sources of uncertainty, become problems of mathematical manipulation and so are, in principle, well-defined. Two key assumptions exist in this model:

Assumption 1: The state variable θ_t is a Markov chain.

Assumption 2: Conditionally on θ_t, y_t are independent and depend on θ_t only, i.e. $y_t|\theta_t \perp y_s|\theta_s, \forall(s,t) \in \mathbb{R}^2$

[3]See appendix on Normal Gaussian distribution for derivation

3.3.3 Bayesian Framework

Following the general Bayesian framework in West and Harrison [50], it is straightforward to deal with linear terms in time-varying mean reversion terms.

Simply speaking, the key concept in Bayesian method is to treat the parameter θ_t itself as a random variable.

Suppose $\theta_1|\sigma^2 \sim \mathcal{N}_2(m_1, \sigma^2 P_1)$ and $\sigma^2 \sim \mathcal{IG}(\frac{n_1}{2}, \frac{d_1}{2})$, the posterior distribution of $\theta_t|\sigma^2, \mathcal{F}_t$ (filtering) and the predictive distribution $y_{t+1}|\sigma^2, \mathcal{F}_t$ (prediction) are obtained through modified Kalman Filter. Here $\mathcal{F}_t$ is the filteration in the probability space $(\Omega, \mathbb{P}, \mathcal{F})$. In some of the notations, $\mathcal{F}_t$ is exchangeable with $y^t := (y_1, y_2, \cdots, y_t)$. The estimation procedure can be obtained through the following induction of Bayesian approach.

Induction

Assume that at time $t-1, \forall t \geq 2$, for some given coefficients

$$(m_{t-1}, P_{t-1}, n_{t-1}, d_{t-1})$$

the posteriors are given by

$$(\theta_{t-1}, \sigma^{-2})|y^{t-1} \sim \mathcal{NG}(m_{t-1}, P_{t-1}, \frac{n_{t-1}}{2}, \frac{d_{t-1}}{2}) \tag{3.3.4}$$

By definition of Normal Gamma distribution,

$$\theta_{t-1}|\sigma^2, y^{t-1} \sim \mathcal{N}_2(m_{t-1}, \sigma^2 P_{t-1})$$

and

$$\sigma^2|y^{t-1} \sim \mathcal{IG}(\frac{n_{t-1}}{2}, \frac{d_{t-1}}{2})$$

because

$$\theta_t|\theta_{t-1}, \sigma^2 \sim \mathcal{N}_2(\Phi\theta_{t-1}, \sigma^2 V_t)$$

and plus the system equation, it is easy to verify that

$$\theta_t|\sigma^2, y^{t-1} \sim \mathcal{N}_2(\Phi m_{t-1}, \sigma^2 V_t + \sigma^2 \Phi P_{t-1} \Phi^{'})$$

therefore, together with distribution of $\sigma^2|y^{t-1}$,

$$(\theta_t, \sigma^{-2})|y^{t-1} \sim \mathcal{NG}(\Phi m_{t-1}, \Phi P_t \Phi^{'} + V_t, \frac{n_{t-1}}{2}, \frac{d_{t-1}}{2})$$

again according to the property of Joint Normal Gamma Distribution[4], it is straightforward to notice that

$$\theta_t|y^{t-1} \sim t(\Phi m_{t-1}, (\Phi P_t \Phi^{'} + V_t)\frac{d_{t-1}}{n_{t-1}}, n_{t-1}) \quad (3.3.5)$$

then by

$$y_t|\theta_t, \sigma^{-2} \sim \mathcal{N}(F_t^{'}\theta_t, \sigma^2)$$

there is the prediction distribution

$$y_t|y^{t-1} \sim t(F_t \Phi m_{t-1}, F_t(\Phi P_t \Phi^{'} + V_t)F_t^{'}\frac{d_{t-1}}{n_{t-1}}, n_{t-1}) \quad (3.3.6)$$

[4]See Appendix of Joint Normal Gamma Distribution

where

$$\mathbb{E}[y_t|y^{t-1}] = F_t^{'}\Phi m_{t-1}$$

$$\mathbb{V}\mathrm{ar}[y_t|y^{t-1}] = F_t(\Phi P_t \Phi^{'} + V_t)F_t^{'}\frac{d_{t-1}}{n_{t-1}}\frac{n_{t-1}}{n_{t-1}-2}$$

Accordingly,

$$(\theta_t, \sigma^{-2})|y^t \sim \mathcal{NG}(m_t, P_t, \frac{n_t}{2}, \frac{d_t}{2}) \tag{3.3.7}$$

where

$$m_t = \Phi_t m_{t-1} + (\Phi P_t \Phi^{'} + V_t)F_t(F_t(\Phi P_t \Phi^{'} + V_t)F_t^{'} + 1)^{-1}(y_t - F_t\Phi m_{t-1})$$

$$P_t = \Phi P_t \Phi^{'} + V_t - (\Phi P_t \Phi^{'} + V_t)F_t(F_t(\Phi P_t \Phi^{'} + V_t)F_t^{'} + 1)^{-1}(\Phi P_t \Phi^{'} + V_t)^{'}F_t^{'}$$

$$n_t = n_{t-1} + 1$$

$$d_t = d_{t-1} + (y_t - F_t\Phi m_{t-1})^{'}(F_t(\Phi P_t \Phi^{'} + V_t)F_t^{'} + 1)^{-1}(y_t - F_t\Phi m_{t-1})$$

In other words,

$$\sigma^{-2}|y^t \sim \mathcal{G}(\frac{n_t}{2}, \frac{d_t}{2})$$

and

$$\theta_t|y^t, \sigma^2 \sim \mathcal{N}(m_t, \sigma^2 P_t)$$

3.3.4 Estimation

The estimation procedure is a systematic way of doing Bayesian forecasting. The following on-line estimation procedure is to make the conditional investigation of pricing time-series y_t easier by defining a few more variables.

Simplified on-line algorithm

The estimation algorithm is an adaptive and recursive algorithm using $AR(1)$ coefficients (ϕ_1, ϕ_2) and discount factors (δ_1, δ_2). The whole process can be simplified into following steps by denoting $f_t = F_t^{'}\Phi m_t$, $R_t = \Phi P_t \Phi^{'} + V_t$ and $Q_t = F_t^{'} R_t F_t + 1$,

$\forall t = 2, \cdots, N$

$$e_t = y_t - f_{t-1}$$

$$m_t = \Phi m_{t-1} + \frac{R_{t-1} F_{t-1} e_t}{Q_{t-1}}$$

$$P_t = R_{t-1} - K_t K_t^{'} Q_{t-1}$$

$$n_t = n_{t-1} + 1$$

$$d_t = d_{t-1} + e_t^{'} Q_{t-1}^{-1} e_t$$

$$V_t = diag(\frac{1-\delta_1}{\delta_1}\phi_1^2 p_{11,t}, \frac{1-\delta_2}{\delta_2}\phi_2^2 p_{22,t})$$

$$R_t = \Phi P_t \Phi' + V_t$$

$$Q_t = F_t^{'} R_t F_t + 1$$

$$f_t = F_t^{'} \Phi m_t$$

After introducing the simplified notation, it is easy to verify that $\theta_t | y^t \sim t_2(n_t; m_t, \frac{P_t d_t}{n_t})$, i.e. the confidence interval of Type I error α is

$$\theta_{it} | y^t \in [m_{it} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{p_{ii,t} d_t}{n_t}}], \forall i = 1, 2 \tag{3.3.8}$$

where $\theta_{1t} = A_t$ and $\theta_{2t} = B_t$.

Different from the time-invariant stochastic modeling, time-varying mean is time-dependent and defined as

$$Z_t = \frac{A_t}{1 - B_t}$$

because $y_t - Z_t = B_t(y_{t-1} - Z_t) + \psi_t$.

The corresponding forecasting distribution is

$$y_t | y^{t-1} \sim t(f_{t-1}, \frac{Q_{t-1} d_{t-1}}{n_{t-1}}, n_{t-1}) \tag{3.3.9}$$

the confidence interval of Type I error α is

$$y_t|y^{t-1} \in [f_{t-1} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{Q_{t-1}d_{t-1}}{n_{t-1}}}] \tag{3.3.10}$$

3.3.5 Calibration

The calibration of the four parameters $(\phi_1, \phi_2, \delta_1, \delta_2)$ of time-varying mean-reverting model involves non-linear optimization with no constraints. The objective functions could be a measure of goodness of fit or loglikelihood function and its derivatives.

Firstly, Goodness of fit can be evaluated through Mean Squared Standardized Errors (MSSE) As it is shown previously in Equation 3.3.9

$$y_t|y^{t-1} \sim t(f_{t-1}, \frac{Q_{t-1}d_{t-1}}{n_{t-1}}, n_{t-1})$$

$$\frac{(y_t - f_{t-1})\sqrt{n_{t-1}}}{\sqrt{Q_{t-1}d_{t-1}}} \sim t(0, 1, n_{t-1})$$

$$u_t = \frac{(y_t - f_{t-1})\sqrt{n_{t-1}}}{\sqrt{Q_{t-1}d_{t-1}}} = \frac{e_t\sqrt{n_{t-1}}}{\sqrt{Q_{t-1}d_{t-1}}} \tag{3.3.11}$$

From the property of a student t-distribution we know

$$\mathbb{E}(u_t) = 0$$

and

$$\mathbb{V}ar(u_t) = \frac{n_{t-1}}{n_{t-1} - 2}$$

Let

$$w_t = \sqrt{1 - 2n_{t-1}^{-1}} u_t$$

be the mean squared standardized errors, then

$$\mathbb{E}(w_t^2) = 1$$

By Law of large numbers, when the price history is very long, it is easy to have the following observation

$$\frac{\sum_{t=2}^{T} w_t^2}{T-1} \to 1$$

The criteria of deciding if the choice of $(\phi_1, \phi_2, \delta_1, \delta_2)$ is fit is to check the mean squared standardized errors of the price series. If it is closer to 1, better is the choice of the parameters for the time-varying mean reverting model.

Secondly, Maximum Likelihood Estimator is another choice for calibration. The classical Likelihood function

$$\prod_{t=2}^{T} f(y_t|y^{t-1})$$

is a candidate for objective function. Since the log function is monotonically increasing, therefore the sum (from 2 to T) of log of probability density function of the standard student t distribution u_t is

$$\log(\prod_{t=2}^{T} f(y_t|y^{t-1})) = \sum_{t=2}^{T} \log(\frac{\Gamma(\frac{n_t}{2})}{\sqrt{\pi n_{t-1}}\Gamma(\frac{n_{t-1}}{2})}) - \frac{1}{2}\sum_{t=2}^{T} n_t \log\{1 + \frac{(e_t)^2}{Q_{t-1}d_{t-1}}\}$$

The other candidates for the objective function includes the Derivatives of Maximum Likelihood Estimator. [5]

3.3.6 Mean Reversion Criteria

Similar to the case in time-invariant model, the criteria for mean-reversion is to look at the B_t value. By the same argument in definition, when B_t lies in $(-1, 1)$, due to

$$y_t - Z_t = B_t(y_{t-1} - Z_t) + \psi_t$$

$$\mathbb{E}[y_s|\mathcal{F}_t] = B_t^{s+1-t}(y_{t-1} - Z_t) + Z_t, \forall s \geq t$$

$$\mathbb{E}[y_s|\mathcal{F}_t] \to Z_t$$

when s is very large.

The extension of the time-varying mean reversion model in above section is the one with discounting factors. The more information we have, i.e. y^t evolves, the conditional variance of volatility is getting bigger, which is not so convincing in practice. It means a fixed treatment of the distribution of σ^2 is not convincing in sense of analysis of large amount of data. However discounting on σ^2 will bring two more discounted factors into calibration, which makes the estimation of parameters more complicated since the optimization problem will add two more degrees of unknown.

[5] See Appendix

CHAPTER 4

Application to Energy Market

4.1 Introduction to energy market

Energy market is a supply demand driven market, unlike other derivatives market, the mean-reversion phenomenon is more common to be observed. Among energy products, the crude oil market is the largest commodity market in the world, with global demand amounting to about 80 million barrels daily. Ten-year-fixed-price supply contracts have been common place in the over-the-counter market for many years. There are swaps where oil at a fixed price is exchanged for oil at a floating price [26]. Besides, natural gas market is the second in all the commodities, and the more mean-reverting pair trading has an origin from natural gas.

Forecasting of energy future price is crucial to every participant in financial markets. Prediction of energy future price certainly involves mean reversion modeling. First of all, it is necessary to know what are the future contracts. In Finance, a futures contract is a standardized contract between two parties to exchange a specified asset of standardized quantity and quality for a price agreed today (the futures price or the strike price) with delivery occurring at a specified future date, the delivery date [26]. The contracts are traded on a futures exchange. The party agreeing to buy the underlying asset in the future, the buyer of the contract, is said to be long, and the party agreeing to sell the asset in the future, the seller of the contract, is said to be short. The terminology reflects the expectations of the parties – the buyer hopes or expects that the asset price is going to increase, while the seller hopes or expects that it will decrease. For the contract itself costs nothing to enter.

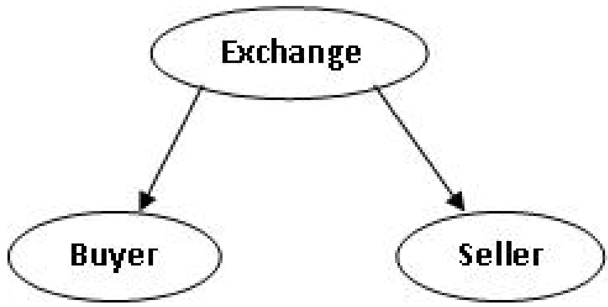

Figure 4.1: Exchange Buyer Seller relationship

In many situations, the underlying asset to a future contract may not be traditional commodities at all – that is, for financial futures the underlying asset or item can be currencies, securities or financial instruments and intangible assets or referenced items such as stock indexes and interest rates. In this application part, the focus mainly lies in energy future contracts.

4.1.1 Data Description

In order to make the month-to-month trading transition smoother, it is common to use price building blocks instead of price generics [1] to monitor the daily price changes. The concept of a building block is approximately similar to that of an Index. For instance, S&P Goldman Sachs Commodity Index is basically a time series of daily returns based on a basket of commodity contracts with pre-defined fixed weights, where all the contracts are tradable in the financial market. With the basic knowledge of how indexes are created and how they are maintained and reported, it is straightforward to apply those concepts to specific case of indexes such as a price building block. A price building block involves at most two contracts instead of a basket. When there is only one contract in the building block, the graph of building block is similar to the generic plot of the corresponding contracts. However, for the two contracts or a spread building block cases, the graph can be very different. In order to understand the case for two contracts, the idea of pair trading is the key. Two contracts combination is called a spread if long one future contract and short another correlated future contract at the same time. The terminology for the two contracts in a spread is leg. The long leg normally correlates with the short leg, which means it can be crude oil contract with gasoline contract, or same commodity underling with different expiry dates.

[1] A generic has no performance change on rolling dates

4.1.2 Terminology

In commodity future market, there are a few naming conventions. To know a contract, it is sufficient to know its underlying commodity name and expiry month and year. The name of the commodities are usually quoted by tickers consist of two characters. The expiry month is quoted according to a one-to-one table. The following two tables list all the necessary information for knowing a commodity future contract. For instance, CLZ1 means Crude Oil contract traded on New York Mercantile Exchange (NYMEX) with expiry month of December, 2011. Similarly, JXX2 means Gasoline contract traded on Tokyo Commodity Exchange (TOCOM) with expiry month of November, 2012.

Table 4.1: Enery products tickers and exchanges

Tickers	Commodities	Exchanges
CL	Crude Oil	NYMEX
CO	Brent Oil	ICE
XB	Gasoline	NYMEX
HO	Heating Oil	NYMEX
NG	Natural Gas	NYMEX
CP	Crude Oil	TOCOM
JV	Gasoline	TOCOM
JX	Kerosene	TOCOM
QS	Gas Oil	ICE
EN	WTI Crude Oil	ICE

4.1.3 Pillars

When the rolling dates of a specific energy contract changes, the corresponding pillar will be different. A pillar could be a combination of commodity name and rolling-dates configuration. For the two front month pillars of a Crude Oil building block, there is a simple convention for

Table 4.2: Energy products months and years

Code	Month
F	January
G	February
H	March
J	April
K	May
M	June
N	July
Q	August
U	September
V	October
X	November
Z	December

its rolling basis. The rolling-dates configurations are normally quoted in such a manner that commodity name followed by its corresponding month-to-expire and year-to-expire ranging from January to December.

Table 4.3: Rolling pillar example for Crude Oil

Rolling Month	H	J	K	M	N	Q	U	V	X	Z	F	G
Rolling Year	0	0	0	0	0	0	0	0	0	0	1	1

Similarly, for three front month Brent Oil, there is

Table 4.4: Rolling pillar example for Brent Oil

Rolling Month	J	K	M	N	Q	U	V	X	Z	F	G	H
Rolling Year	0	0	0	0	0	0	0	0	0	1	1	1

The 0, 1 indicator means the expiry year of the contract held can be either this year or next year. For instance, now we are in year 2012, therefore the corresponding Crude Oil contract for this pillar is 2012 for first 10 months and 2013 for last two months (illustrated on the left column). If the current year is 1990, the case is similar and the corresponding contracts vector refers to the right column.

Table 4.5: Rolling pillar example

CL1S 2012	CO1S 1990
CLH12	COJ90
CLJ12	COK90
CLK12	COM90
CLM12	CON90
CLN12	COQ90
CLQ12	COU90
CLU12	COV90
CLV12	COX90
CLX12	COZ90
CLZ12	COF91
CLF13	COG91
CLG13	COH91

4.2 Numerical Example

After a brief introduction to the basic concepts and terminology in energy future analysis, the next step is to compare the performance of the two models developed in Chapter 3 using historical data.

4.2.1 Wavelet-Schwartz Model

In this section, the calibration and predicting results of wavelet-Schwartz Model and the sensitivity analysis on the simple sign prediction strategy are presented. Considering a basket of Energy Future Contracts, namely CO1S, CO2S, CL1S, CL2S, XB1S, XB2S, following graphs show the price dynamics of their corresponding building blocks. First assuming the dynamics of the six risky asset are all pure mean-reverting, i.e. the degree d is set to be 1, then their prices follow the presumed Wavelet-Schwartz model. The numerical result when the three parameters are calibrated for log price by using the indirect inference method is as below,

Figure 4.2: Brent Oil with maturity on November 2011

Figure 4.3: Brent Oil with maturity on December 2011

Figure 4.4: Crude Oil with maturity on November 2011

Figure 4.5: Crude Oil with maturity on December 2011

Figure 4.6: Gasoline with maturity on November 2011

Figure 4.7: Gasoline with maturity on December 2011

Table 4.6: Time-invariant Model Results I

Estimator	COX1	COZ1	CLX1	CLZ1
λ	0.003462456	0.003473071	0.005260256	0.003904866
$\bar{\mu}$	4.571446	4.571596	4.476605	4.440749
σ	0.01700883	0.01691612	0.01780006	0.01652342

Table 4.7: Time-invariant Model Results II

Estimator	XBX1	XBZ1
λ	0.007285	0.007366415
$\bar{\mu}$	5.538515	5.537321
σ	0.01622564	0.01603127

Based on these time-invariant coefficients λ, μ, σ, a simple trading strategy could be the signal (relative value) based on the one-step ahead prediction from this time-invariant mean reversion model, that is to say,

$$\frac{\mathbb{E}[\hat{y}_{t+1}] - y_t}{C \times \sigma_{\hat{y}_{t+1}}}$$

where $\hat{y}_{t+1}$ is the forecast value for the next trading session $t+1$, C is the z-score time-independent and varying from 1 to 3, and $\sigma_{y_{t+1}}$ is the volatility of price building block. However, to avoid lookahead bias, a rolling window of different length can be applied on the original time series. In following sensitivity analysis, the length of rolling basis is fixed to be one year period. The wavelet-Schwartz time invariant mean reversion model has not achieved outstanding performance due to the degree of mean reversion may not be perfect and the assumption on time independence is too strict to be realistic. On the other hand, to certain extend, the depth of mean-reversion is the key of concern, log Orhstein-Uhlenbeck model can be applied on the decomposed pricing time series. By denoting s for smoothies, d for details after wavelet decomposition, applying log price mean reversion model on different details, the trading strategy was simply

$$\frac{\mathbb{E}[\hat{d}_{i,t+1}] - d_{i,t}}{C * \sigma_{d_{i,t+1}}}$$

with a fixed choice $C = 1.5$, where $d_{i,t}$ are the various cycles extracted from decomposition.

On the two-most traded energy futures, Crude Oil and Natural Gas, there is following sensitivity analysis based on different choice of details, rolling window length and the C values. The purpose of the first two sensitivity analysis is to test to what degree the Sharpe ratio varies with respect to different choice of cycles, namely the details. Even through different degrees of extraction of the original price time series generate various cycles, the peak of the Sharpe ratios is around a common choice of z-score 1.5. As it can be seen in Figure 4.8 and Figure 4.9, based on different details of the wavelet decomposition, both commodities show the similar evolution of performance with respect to different choice of C. However the Sharpe ratio based on all different details are barely sufficient in time-invariant model, on average it is 0.3 - 0.4, which means the link between the purely mean reverting details in not strong enough to have good prediction of the price. However, if a change of trading underlying to spread with the model applied on cumulative details, the Sharpe ratio of CL2S-CO1S could go as high as 1.5. This might bring more profits when co-integration techniques are used. The Sharpe ratio is a standard profit and loss measure defined as $\frac{\mathbb{E}(return)}{\sigma_{return}}$. Unlike Figure 4.9, Figure 4.10 and 4.11 shows that for separate details Cumulative strategy Sharpe ratio is generally lower than that of non-cumulative. This could be partially explained by the interdependency among the decomposed details. However, if we look at the spread performance in Figure 4.13 and Figure 4.12, the performance in terms of Sharpe ratio are much better than the Single asset based ones. This is in line with our observation that a spread between two non-perfect correlated assets is more mean-reverting than the single assets themselves, which fits more properly into the model assumed in the strategy.

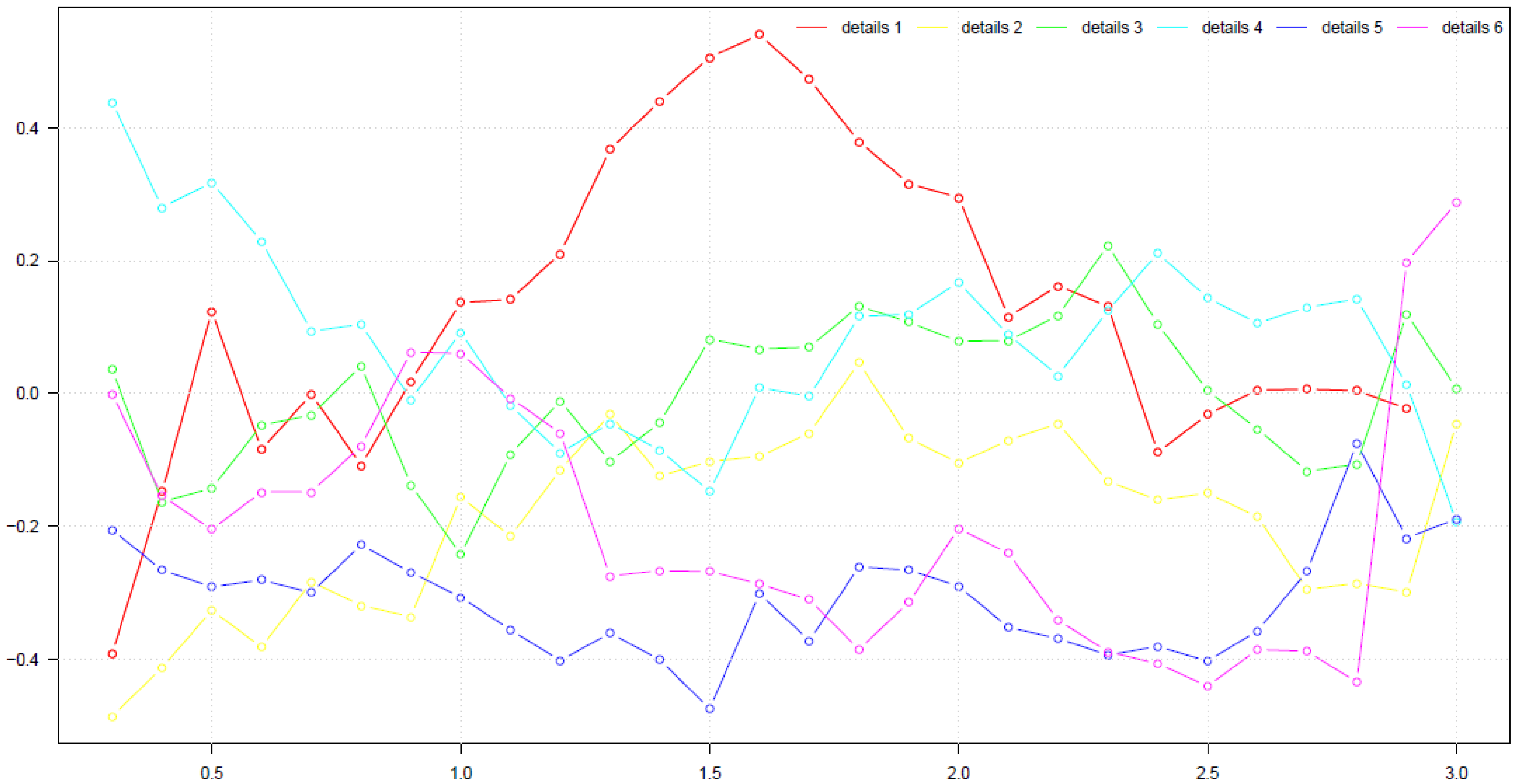

Figure 4.8: Sharpe Ratio with different C values for Crude Oil front month wavelet-schwartz Senstivity Analysis

Figure 4.9: Sharpe Ratio with different C values for Natural Gas front month wavelet-schwartz Senstivity Analysis

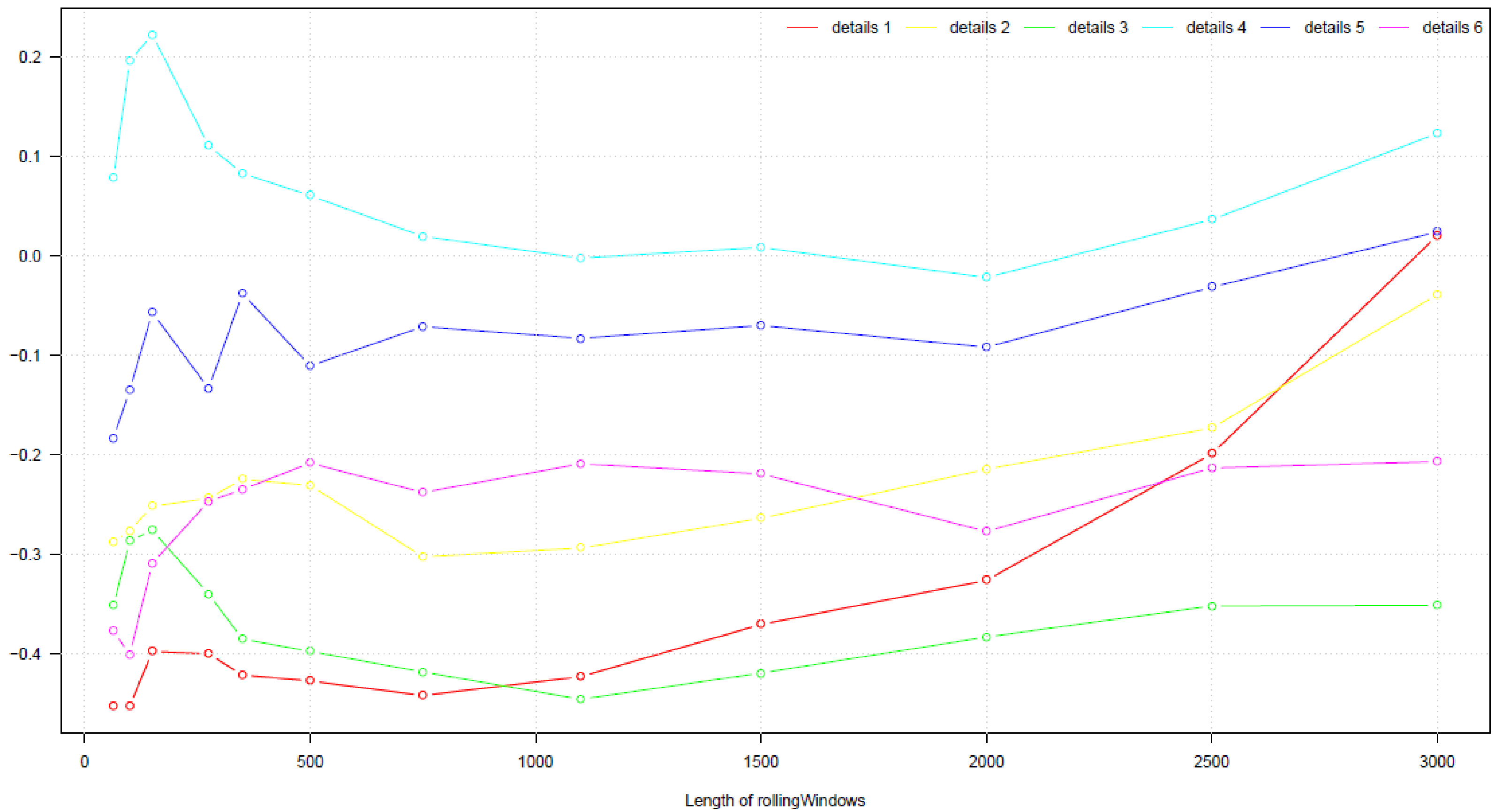

Figure 4.10: Sharpe Ratio with different historical sample length for Crude Oil front month wavelet-schwartz Senstivity Analysis with cumulative details

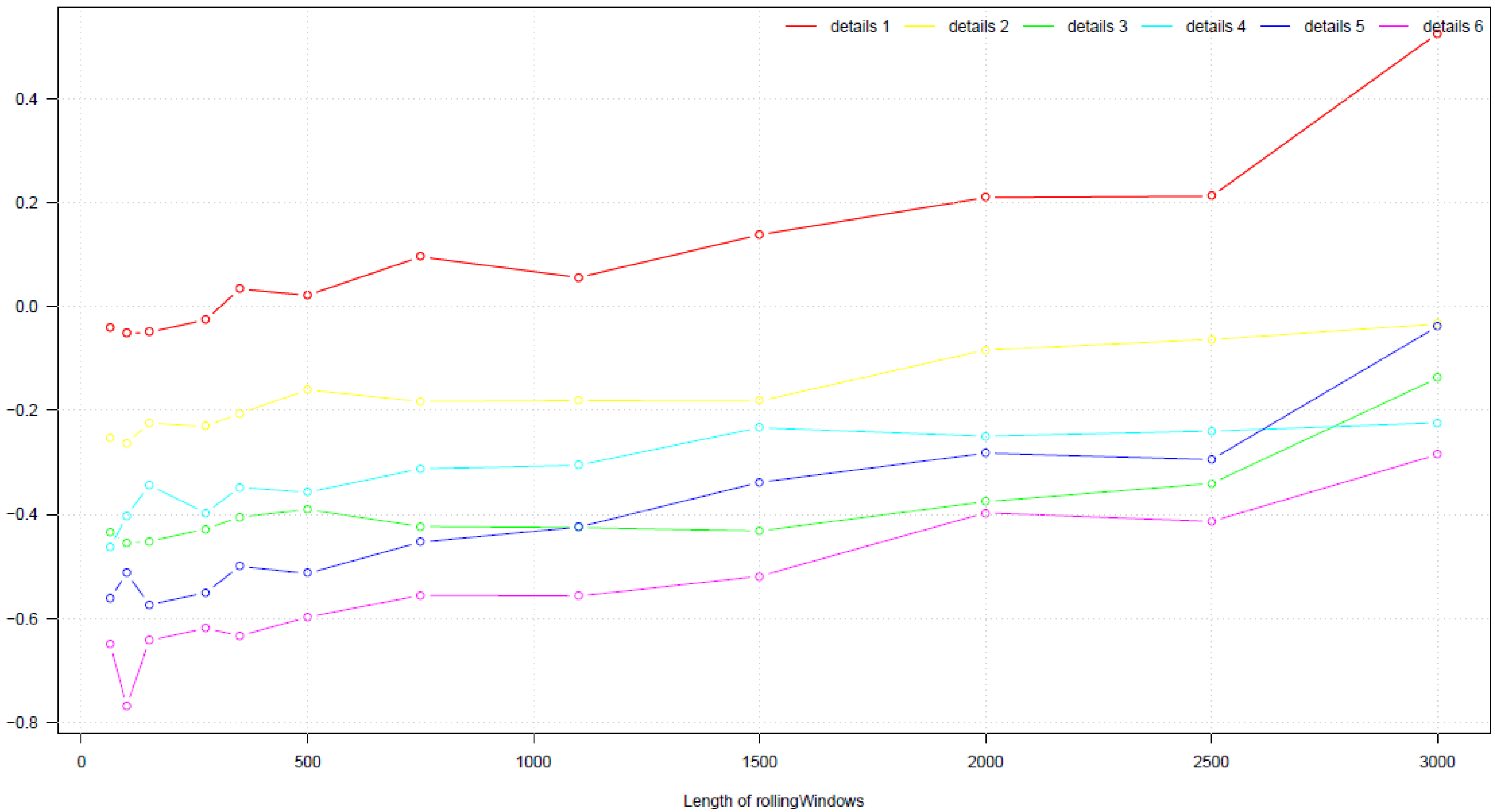

Figure 4.11: Sharpe Ratio with different historical sample length for Natural Gas front month wavelet-schwartz Sensitivity Analysis with cumulative details

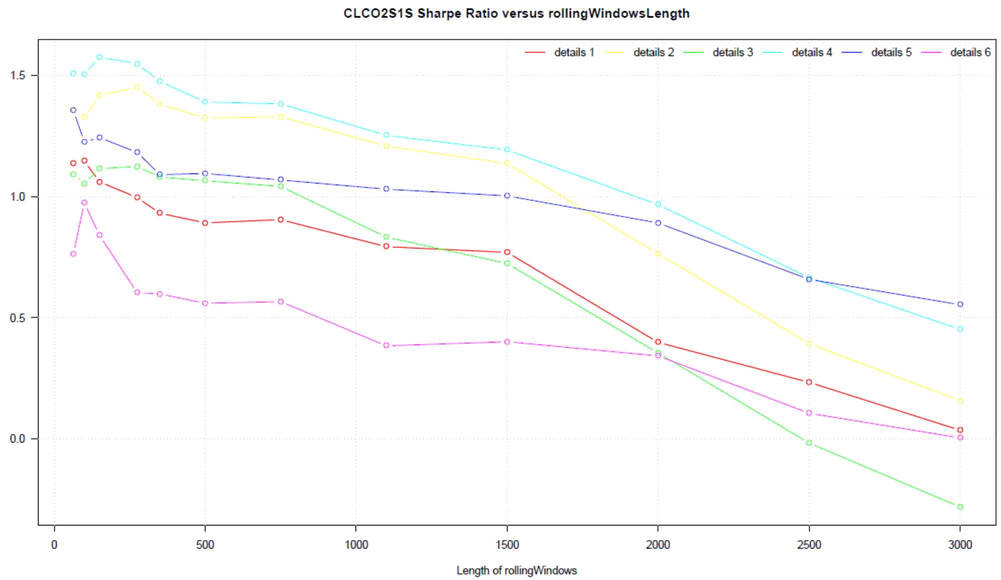

Figure 4.12: Sharpe Ratio with different historical sample length for Crude Oil and Brent Oil spread wavelet-schwartz Senstivity Analysis with cumulative details

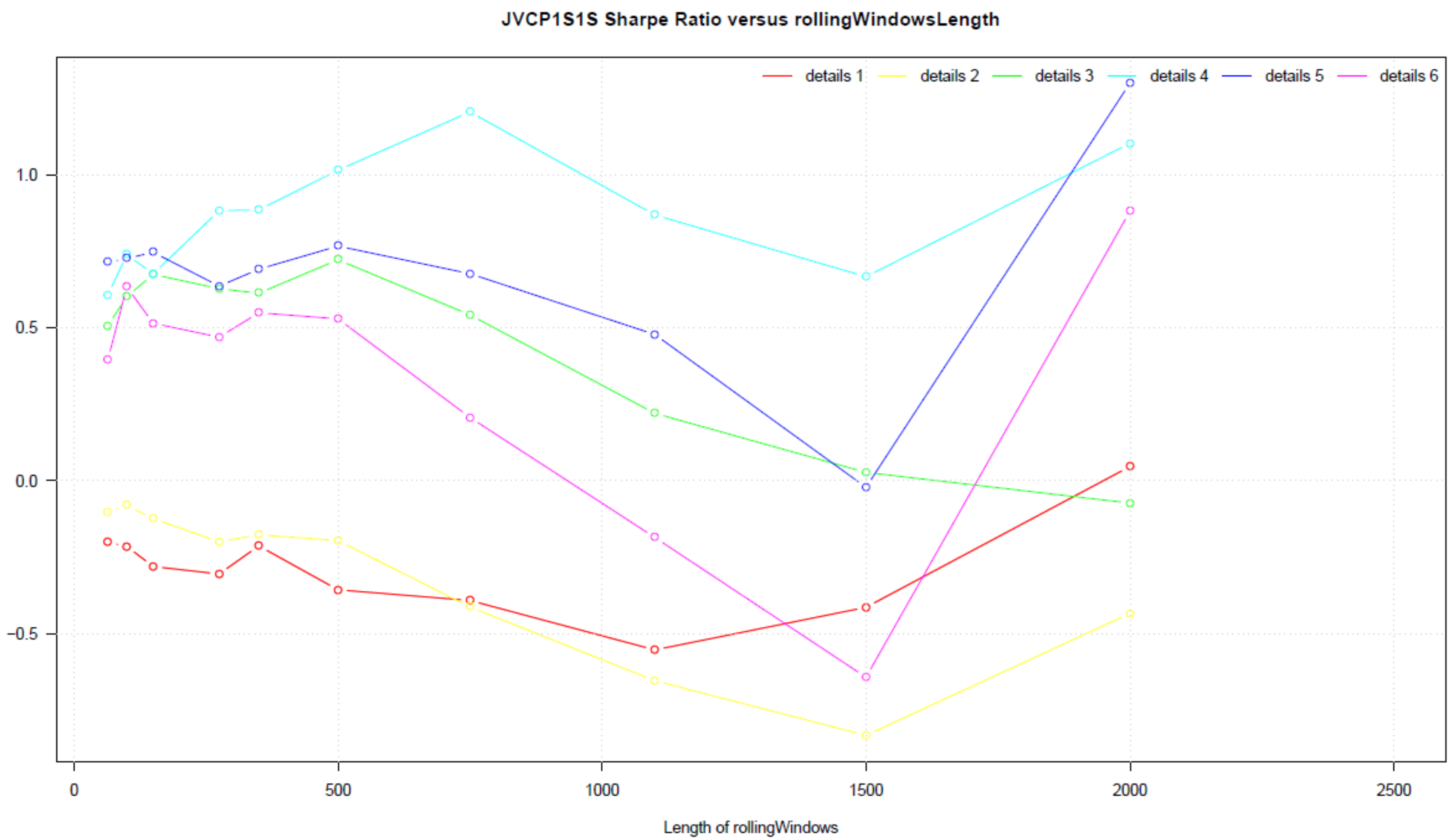

Figure 4.13: Sharpe Ratio with different historical sample length for Gasoline Crude Oil spread front month wavelet-schwartz Sensitivity Analysis with cumulative details

4.2.2 State Space Model

Considering a spread consists of two pillars called CL2S and CL5S, Figure 4.14

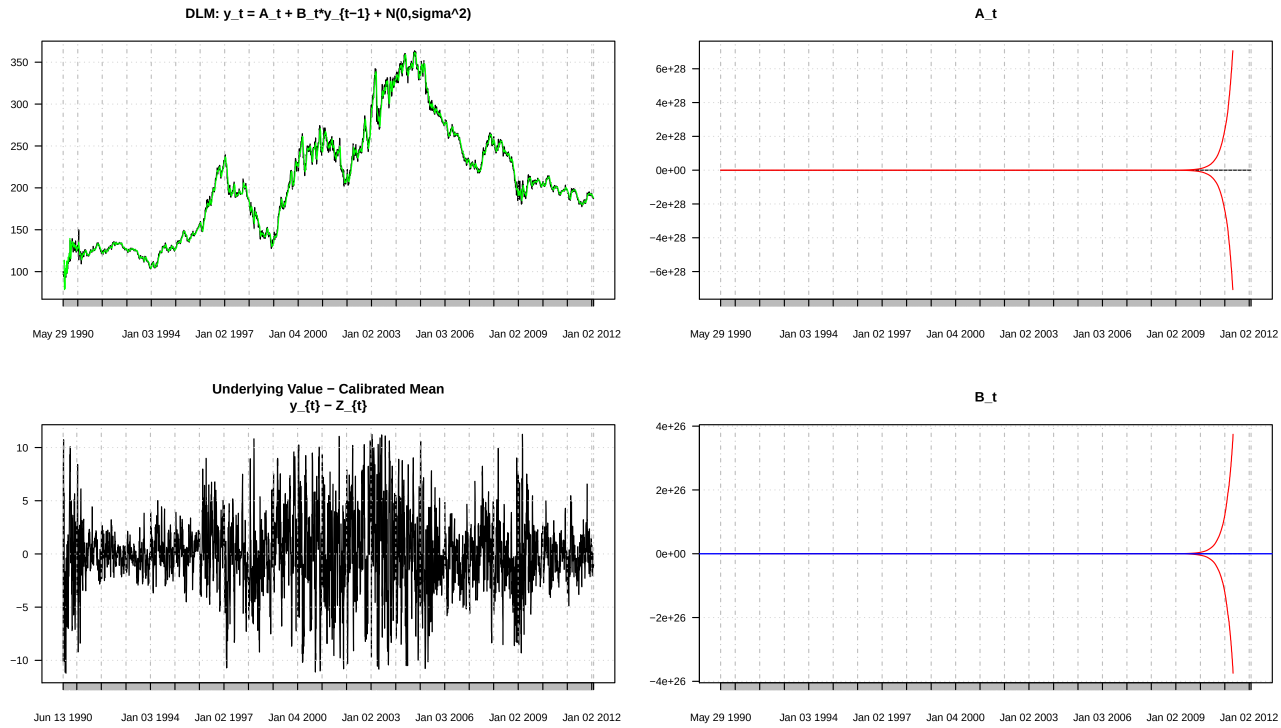

Figure 4.14: State space calibration on energy spread

has four analytical graphs showing the calibration results of state space time-varying mean-reversion model respectively.

1 The upper left corner has the original black time series and the time-varying mean Z_t given by $\frac{A_t}{(1-B_t)}$, which is smoother compared to the price time series.

2 The upper right corner is a control chart for coefficient A_t r.v. , the two diverging lines is the $\pm 2\sigma_{A_t}$, while the middle one is the expectation A_t. As it shows, this specific choice of parameters $(\phi_1, \phi_2, \delta_1, \delta_2)$ does not lead to a convergence of A_t.

3 This is the diffusion or the difference between observed data and moving mean, it is $y_t - Z_t$.

4 Similar to the upper right corner, the lower right is the control chart for B_t.

The state variable θ_t has been evolving through time, with more information arrives, the Bayesian estimation of this two parameters are less accurate. Therefore, the sensitivity analysis on the choice of $(\phi_1, \phi_2, \delta_1, \delta_2)$ is inevitable. In the following subsection, a detailed discussion on the choice of $(\phi_1, \phi_2, \delta_1, \delta_2)$ is presented.

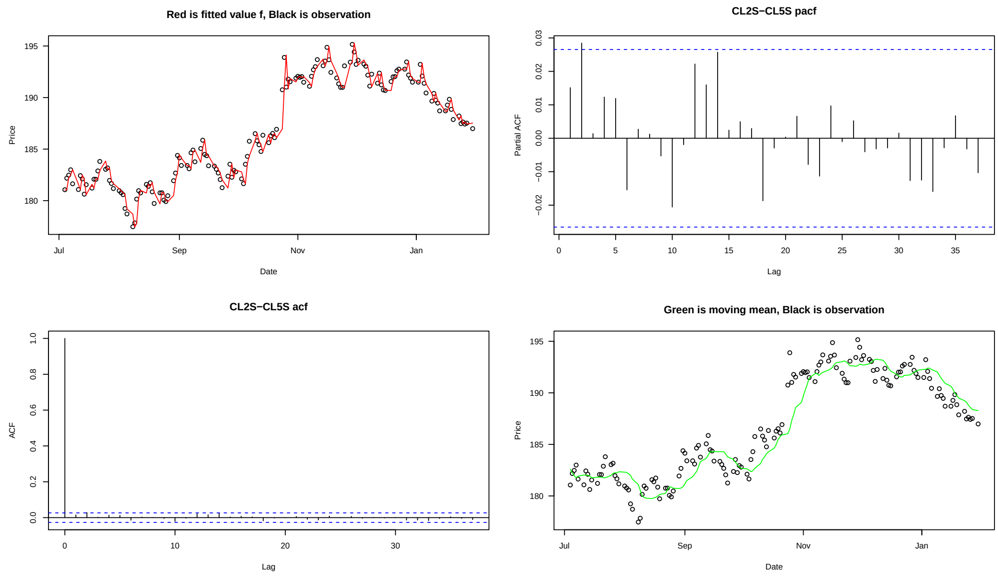

Figure 4.15: ACF and PACF on energy spread

Figure 4.15 has four analytical graphs.

1 The upper left corner has the original time series and the fitted $\hat{y}_t := f_t$, the two evolves similarly through time.

2 The upper right corner is a partial autocorrelation function for $\hat{y}_t - y_t := f_t - y_t$ random variable.

3 The lower left corner is the autocorrelation function for $\hat{y}_t - y_t := f_t - y_t$ random variable.

4 The lower right corner is the moving average z_t.

An inappropriate choice of $(\phi_1, \phi_2, \delta_1, \delta_2)$ leads an unstable calibration which makes the $\hat{y}_t - y_t := f_t - y_t$ unstationary, in the following section, the choice of $\phi_1, \phi_2, \delta_1, \delta_2$ has affected the measuring Sharpe ratios differently.

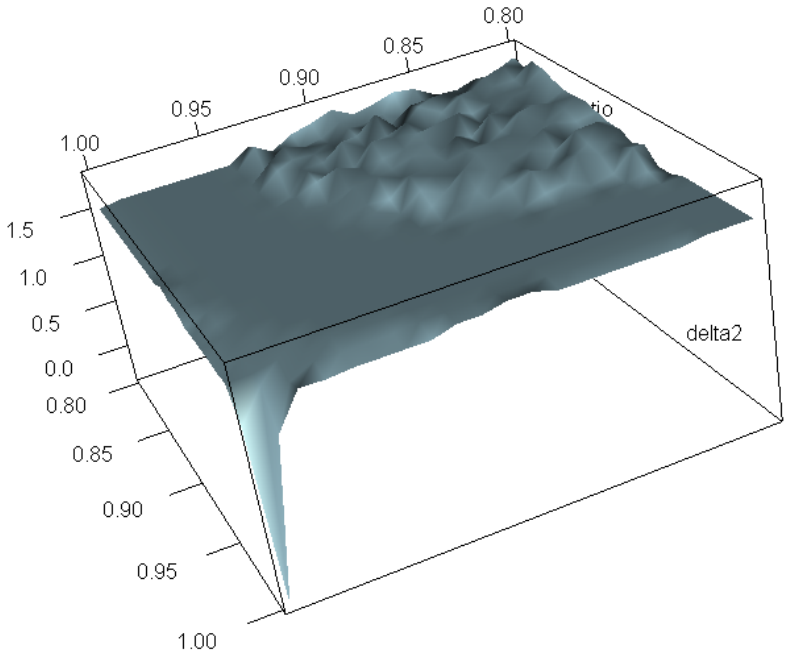

Figure 4.16: Sharpe Ratio sensitivity analysis on δ_1 and δ_2

Figure 4.16 is the sensitivity analysis with fixed value of (ϕ_1, ϕ_2) as $(1, 1)$, by varying the ϕ_1,ϕ_2. As we can see, the ratio does not demonstrate an impressive improvement until (δ_1, δ_2) has moved out from (0.95,0.95) to (1,1) region. However, (0.95,0.95) to (1,1) region has an average Sharpe Ratio around 1.5, this shows that, the time-varying model has a relatively good Ratio if (δ_1, δ_2) is not one, which approves the assumption that variance of state equation depends on time.

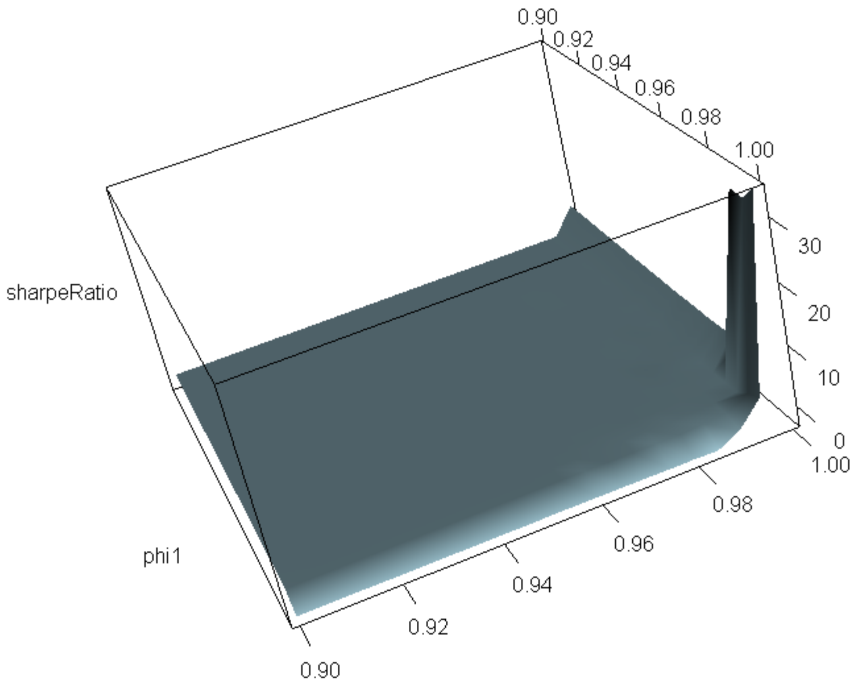

Figure 4.17: Sharpe Ratio sensitivity analysis with ϕ_1 and ϕ_2

Moreover, Figure 4.17 fixed (δ_1, δ_2) as $(0.95, 0.95)$, which is smaller than previous test, as the graph shows the Sharpe Ratio is negative and flat for most of the (ϕ_1, ϕ_2) except $(1, 1)$. This is to say, the A_t and B_t with fixed coefficient $(1, 1)$ can bring along a better Sharpe Ratio as we have observed in the first sensitivity tests.

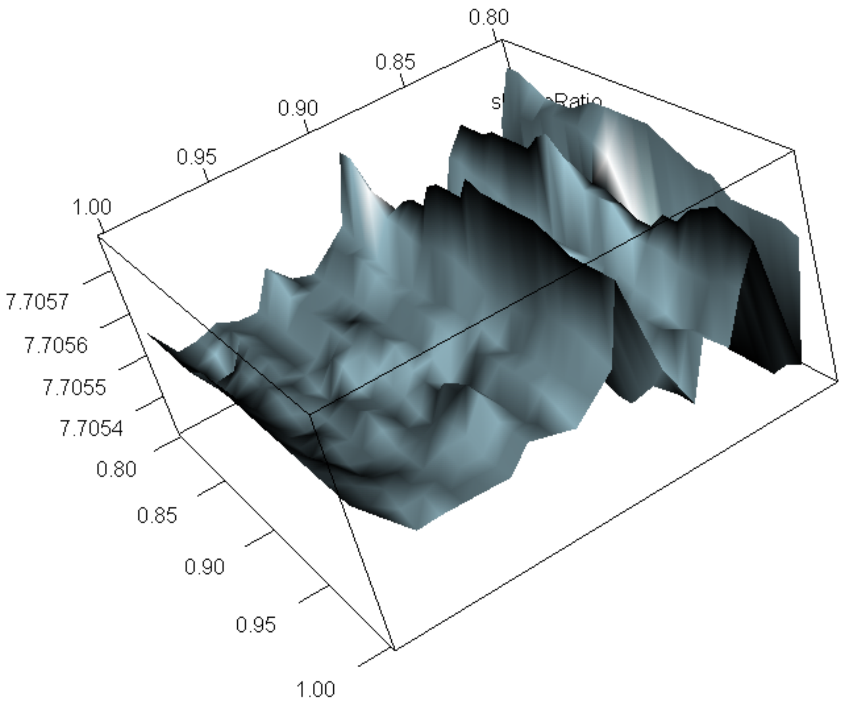

Figure 4.18: Sharpe Ratio sensitivity analysis with δ_1 and δ_2 two days

To deal with the potential lookup bias in the sensitivity analysis. Figure 4.18 has chosen the larger interval between the sample points, i.e. from one day to two days sampling. The sensitivity analysis of the Sharpe Ratio shows that the stationarity of the time series has lost, as more and more peaks occur with the range of δ_1 and δ_2 from 0.80 to 1.00.

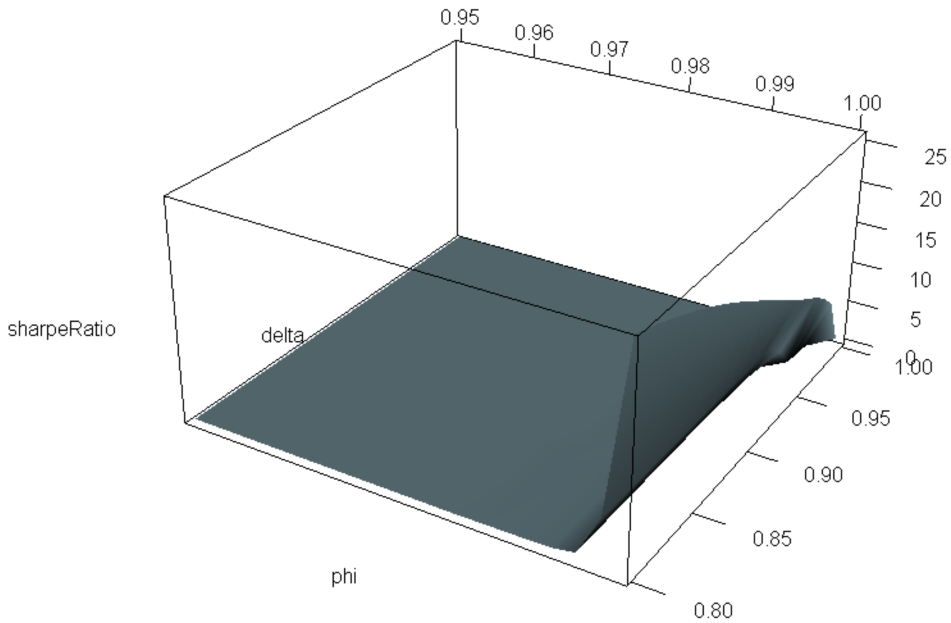

Figure 4.19: Sharpe Ratio sensitivity analysis with δ and ϕ

While in previous sensitivity analysis, we have observed an obvious symmetry between ϕ_1 (δ_1) and ϕ_2 (δ_2), it is rational to make them both equal. Figure 4.19 shows that the fixing $\delta_1 = \delta_2$, $\phi_1 = \phi_2$, the Ratio has a skew with respect to δ when ϕ is high, or close to 1, and negative ratio occurs with smaller choice of ϕ. The key parameter to determine the mean-reversion is therefore ϕ_1 and ϕ_2.

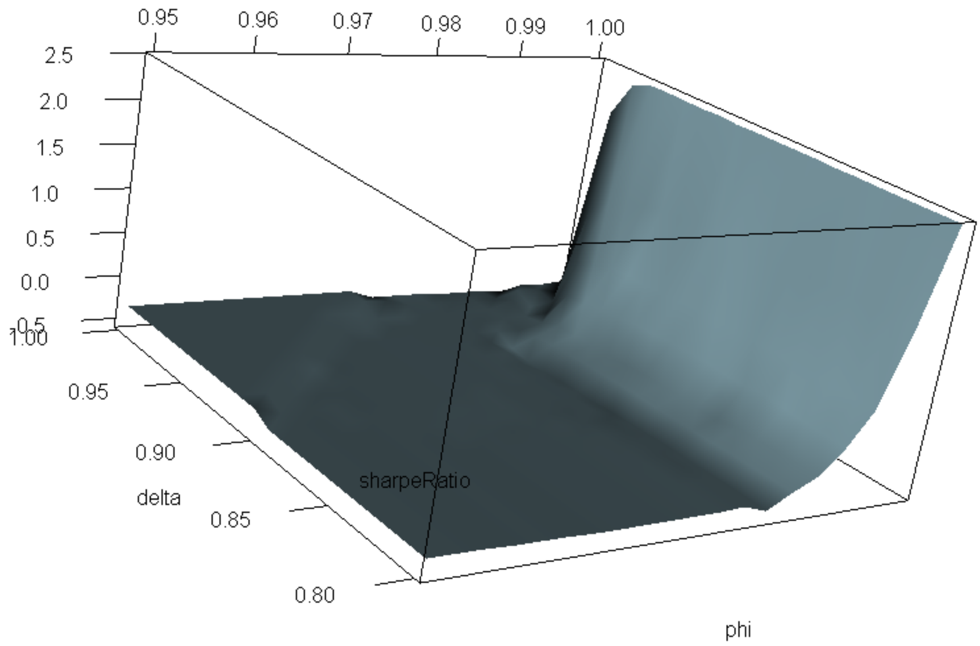

Figure 4.20: Sharpe Ratio sensitivity analysis with δ and ϕ Whole

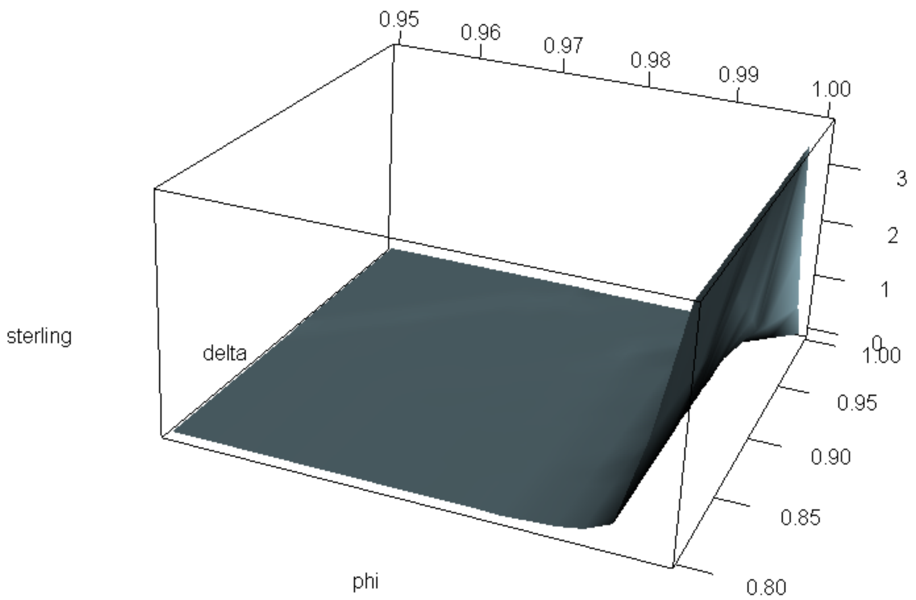

Figure 4.21: Sterling Ratio sensitivity analysis with δ and ϕ Whole

Similarly, by choosing another measure of goodness or trading profits, i.e. sterling ratio, and annualized returns. Figure 4.20, Figure 4.21 and Figure 4.22 show quite similar skewed sensitivity output as to last sens-

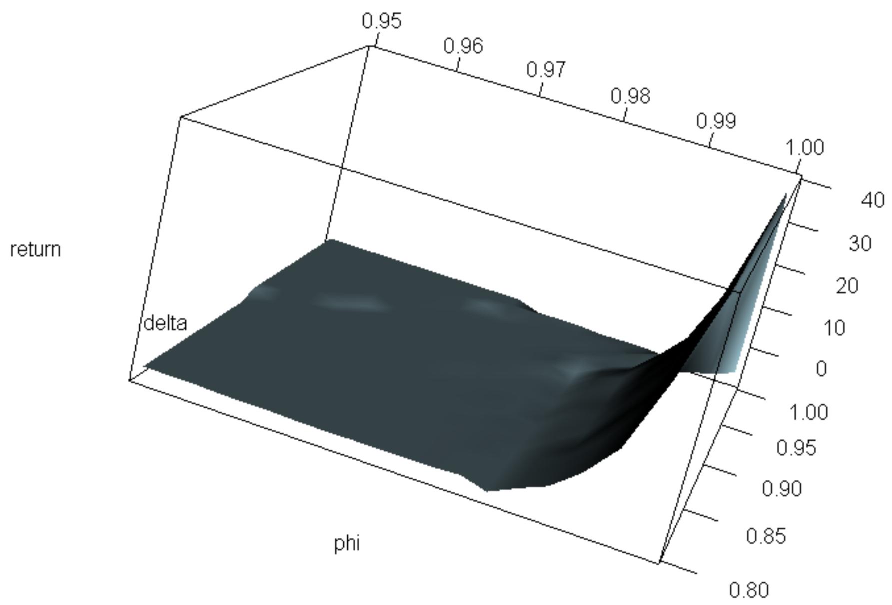

Figure 4.22: Return sensitivity analysis with δ and ϕ Whole

itivity analysis by fixing $\phi_1 = \phi_2, \delta_1 = \delta_2$.

In a nutshell, the sensitivity analysis in this section shows that the ϕ and δ parameters as the control of state coefficients and variance has shown symmetry on parameter choices. The analysis was conducted on different choice of goodness measure of trading profit and loss, which shows the similar result.

CHAPTER 5

Conclusion

5.1 Summary

In this thesis, first of all a review of three main forms of mean reversion is done and a formal mathematical definition of what most investment practitioners seem to mean by mean reversion, based on the correlation of returns between disjoint intervals is studied. Then a collection of existing mean reversion tests are categorized. This thesis proposes two new models: Time-Invariant Wavelet-Schwartz Model and Time-Varying State Space Model. The first model, by incorporating wavelet decomposition techniques into the Schwartz model, is capable of describing any time series. The second model, adopting a Bayesian approach, can be used to describe mean reverting processes where the mean is constantly changing. Apart from the theoretical development, this thesis applies the proposed two models to historical energy price data to test their profitability in trading activity. The simulation results generated by the two models are compared and discussed. Last chapter is the conclusion of the thesis and some further works of this research topic.

5.2 Contribution

The main contributions of this thesis include an improvement of time-invariant mean reversion model based on wavelet decomposition. This model makes the mean-reversion modeling independent from testing. The mean-reversion nature of asset price can be divided and conquered on different cycles extracted trough wavelet decomposition techniques, called details. A calibration methodology is proposed to the time-invariant mean-reversion model. Besides, another important contribution is the proposal of time-varying model based on linear state space analysis. As an adaptive process, calibration methodology based on Bayesian approach is also proposed to the time-varying model, which makes the modeling complete. Lastly, an application of these two mean reversion models is conducted on the energy future contracts. The application has shown that, when different measure are taken, for instance sterling Ratio and return, they demonstrated similar sensitivity on the choice of (ϕ, δ) by fixing $\phi_1 = \phi_2, \delta_1 = \delta_2$. But the performance of state-space model is very sensitive.

5.3 Future Directions

In the future, researchers could relax some constraints in both proposed models, like introducing time-varying variance to the price observation equation in state-space model, i.e. make the σ^2 time-dependent as in V_t. By doing so, it might introduce two more coefficients to the whole model, which makes the calibration more difficult in practice, i.e. higher degree of freedom make the optimization of the log likelihood function less appealing. Besides, this state space model does not take into account the regime switching effect, like the most recent European debt crisis compared to the the pre-2008 roaring of global economics. On the state evolution equation, a Markov switching modification could be applied such that the resulted model could capture better he mean reversion patterns. In addition, for the first wavelet-Schwartz model, a different choice of decomposition filtering function could be applied to explore deeper the essence of mean reversion on cycles extracted. Moreover, seasonality in storage plays a prominent role in Energy products like natural gas and heating oil. An improved wavelet Schawartz model could be take account of seasonality of commodity futures. Last but not least, co-integration plus mean reversion model will improve the performance to a higher degree, topics related to co-integration deserves being investigated more in the future.

Bibliography

[1] R.J. Balvers, Y. Wu. Momentum and mean reversion across national equity markets. Journal of Empirical Finance, Vol. 13: 24-48, 2006

[2] P.M. Bentley, J.T.E. McDonnel. Wavelet transforms: an introduction, Electronic and Communication Engineering. Journal, Volume 6, issue 4, August 1994

[3] H. Bessembinder, J. Coughenour, P. Seguin, M. Smoller. Mean reversion in equilibrium asset price: Evidence from the futures term structure. Journal of Finance, Vol.50:361-375, 1995

[4] F. Black and M. Scholes. The pricing of options and corporate liabilities. Journal of Political Economy, Vol.81:637-659, 1973

[5] G.E.P. Box, G.M. Jenkins. Time series analysis: Forecasting and control, 2nd edition, San Francisco, Holden-Day, 1976

[6] P.J. Brockwell, R.A. Davis. Introduction to Time Series and Forecasting, second edition, Springer-Verlag, New York. 2002

[7] G. Carcano, P. Falbo, S. Stefani. Speculative trading in mean reverting markets. European Journal of Operational Research, Vol. 163:132-144, 2005

[8] L. Clewlow, C. Strickland. A Multi-Factor Model for Energy Derivatives. Research Paper Series 28, Quantitative Finance Research Centre, University of Technology, Sydney.

[9] G. Cortazar and E.S. Schwartz. The Valuation of Commodity Contingent Claims. Journal of Derivatives, 1:4, 27-39, 1994

[10] J.C. Cox, J.E. Ingersoll, S.A. Ross. A Theory of the Term Structure of Interest Rates. Econometrica, Vol. 53:, 385-407, 1985

[11] J. Crosby. A Multi-Factor Jump-Diffusion Model for Commodities. Quantitative Finance, Vol. 8:181-200, 2008

[12] Econometric Theory and Methods. R. Davidson, J.G. MacKinnon, page 623, 2004

[13] S. Deng. Stochastic Models of Energy Commodity Prices and Their Applications: Mean-reversion with Jumps and Spikes. Unpublished manuscript, Georgia Institute of Technology

[14] A.K. Dixit, R.S. Pindyck. Investment under Uncertainty. Princeton University Press, 1994, page 76

[15] A. Eydeland, H.German. Fundamentals of Electricity Derivatives. Energy Modelling and the Management of Uncertainty, Risk Books, 35-43, 1999

[16] E.F. Fama, K.R. French. Permanent and Temporary Components of Stock Prices. Journal of Political Economy, Vol. 96: 246-273, 1988

[17] J.A. Frankel, A.K. Rose. A panel project on purchasing power parity: Mean reversion within and between countries. Journal of International Economics, Vol. 40: 209-224, 1996

[18] K.R. French. Why and when do spot prices of crude oil revert to futures price levels? Working Paper, Finance and Economics Discussion Series 2005-30, Board of Governors of the Federal Reserve System.

[19] H. Geman, A. Roncoroni. Understanding the Fine Structure of Electricity Prices. Journal of Buisness, Vol. 79: 1225-1261, 2006

[20] H. German. Mean Reversion versus Random Walk in Oil and Natural Gas Prices. Seminar of Financial Engineering, Columbia University, 2006

[21] C. W. J. Granger, R. Joyeux. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1: 15–30, 1980

[22] P.J. Harrison, C. Stevens. Bayesian Forecasting with discussion. Journal of the Royal Statistical Society Series B, Vol.38: 205-247, 1976

[23] R.Z. Hasminskii. Stochastic Stability of Differencial Equations. Sitjthoff and Noordhoff, 1980

[24] S. Heston. A closed form solution for options with stochastic volatility with Applications to Bond and Currency Options. Review of Financial Studies, Vol. 6: 327-343, 1993

[25] J.E. Hilliard, J. Reis. Valuation of Commodity Futures and Options under Stochastic Convenience Yields, Interest Rates, and Jump Diffusions in the Spot. Journal of Financial and Quantitative Analysis, Vol. 33: 61-86, 1998

[26] J. Hull. Options Futures and Other derivatives. 8th edition. Person

[27] S.H. Irwin, C.R. Zulauf, T.E. Jackson. Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices. American Journal of Agricultural Economics, Vol. 78(2):387-399,1996

[28] N. Kaldor. Speculation and economic stability. Review Economics Study 7:1-27, 1939

[29] M.J. Kim, C.R. Nelson, R. Startz. Mean reversion in stock prices? A reappraisal of the empirical evidence. Review of Economic Studies 58, 515–528, 1991

[30] F.C. Klebaner. Introduction to Stochastic Calculus with applications. Second Edition, Imperial College Press, 2005

[31] P.E. Kloeden, E. Platen. Numerical solution of stochastic differential equations. Springer-Verlag 1992, Berlin, chapter 4.4.

[32] R.H. Litzenberger, N. Rabinowitz. Backwardation in oil futures markets: Theory and empirical evidence. Journal of Finance, Vol.50(5):1517-1545, 1995

[33] A.W. Lo, C. MacKinlay. The Size and Power of the Variance Ratio Tests in Finite Samples: A Monte Carlo Investigation. Journal of Econometrics Vol. 40: 203-38, 1988

[34] S. Mallat. A wavelet tour of signal processing, 2nd Edition, Academic Press, 1999

[35] B.B. Mandelbrot, J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Review Vol. 10(4): 422-437, 1968

[36] J. Morlet, G. Arens, E. Fourgeau and D. Giard. Wave propagation and sampling theory - Part I: complex signal and scaterring in multiayered media. Geophysics, 47, 203-221, 1982

[37] G.P. Nason. Wavelet Methods in Statistics with R. Springer 2008

[38] G.E.Uhlenbeck, L.S.Ornstein. On the theory of Brownian Motion, Physics Review, 36:823–841, 1930

[39] D.B. Percival and A.T. Walden. Wavelet Methods for Time Series Analysis, Cambridge University Press, 2000

[40] P.C.B. Phillips, P. Perron. Testing for a unit root in time series regression. Biometrika, 75, 335–346, 1998

[41] P.C.B. Phillips. Jackknifing Bond Option Prices. Review of Financial Studies, Oxford University Press for Society for Financial Studies, vol. 18(2), pages 707-742, 2005

[42] R. Pindyck. The dynamics of commodity spot and futures market: A primer. Energy Journal, Vol. 22(3):1-29,2001

[43] J.Poterba, L.Summers. Mean Reversion in Stock Returns: Evidence and Implications. Journal of Financial Economics, Vol. 22: 27-59, 1988

[44] M.Richardson, J. Stock. Drawing inferences from statistics based on multi-year asset returns. Journal of Financial Economics 25, 323–348, 1989

[45] M.Richardson. Temporary components of stock prices: A skeptic's view. Journal of Business and Economic Statistics 11, 199–207, 1993

[46] S.A. Ross. Hedging Long-Run Commitments: Exercises in Incomplete Market Pricing. Economic Notes by Banca Monte dei Paschi di Siena SpA Vol. 26: 385-420

[47] B. Routeldge, D. Seppi, C. Spatt. Equilibrium forward curves for commodities. Journal of Finance Vol.55(3):1297-1338, 2000

[48] E.S. Schwartz. The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging. Journal of Finance, Vol. 52:, 923-973, 1997

[49] S. Grobkinsky, C. Klingenberg, K.Oelschlager. A rigorous derivation of Smoluchowski's equation in the moderate limit. Stochastic Analysis and Applications. Vol. 22, No.1 pp. 113-141, 2004

[50] M. West, J. Harrison. Bayesian Forecasting and Dynamic Models, 2nd edition, Springer-Verlag, 1997

[51] H. Working. The theory of the price of storage. American Economics Review 39:1254-1262, 1949

[52] P.C.B. Phillips, J. Yu. Simulation-Based Estimation of Contingent-Claims Prices, Review of Financial Studies, Oxford University Press for Society for Financial Studies, vol. 22(9), pages 3669-3705, September, 2009

APPENDIX A

Stochastic Calculus

A.1 Brownian Motion

Lemma A.1.1. *If B_t is a standard Brownian Motion, then $\mathbb{E}(B_t) = 0$ and $\mathbb{E}(B_t B_s) = min(s,t)$ [30]*

$$\mathbb{E}(B_t|\mathcal{F}_s) = \mathbb{E}(B_t - B_s + B_s|\mathcal{F}_s) = B_s + \mathbb{E}(B_t - B_s|\mathcal{F}_s) = B_s$$

If $t > s$,

$$\mathbb{E}(B_t B_s) = \mathbb{E}((B_t - B_s + B_s)B_s) = \mathbb{E}(B_s^2)$$

Therefore

$$\mathbb{E}(B_s B_t) = min(s,t)$$

A.2 Solving Ornstein-Uhlenbeck SDE

This solving process is based on Kloeden and Platen [31]. If x_t follows an Ornstein-Uhlenbeck process with following stochastic differential equation

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t$$

Let

$$f(t, x_t) = x_t e^{\theta t}$$

Applying Itô' s Lemma, there is

$$\begin{aligned} df(t, x_t) &= x_t\theta e^{\theta t}dt + e^{\theta t}dx_t \\ &= x_t\theta e^{\theta t}dt + \theta e^{\theta t}(\mu - x_t)dt + \sigma e^{\theta t}dW_t \\ &= \mu\theta e^{\theta t}dt + \sigma e^{\theta t}dW_t \end{aligned}$$

Therefore,

$$x_t e^{\theta t} - x_0 = \int_0^t \mu\theta e^{\theta s}\, ds + \int_0^t \sigma e^{\theta s}\, dW_s$$

$$x_t = x_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)}\, dW_s$$

Besides, as $\mathbb{E}[\int_0^t \varphi_s\, dW_s] = 0$ for any definite function φ_s.

$$\mathbb{E}[x_t] = x_0 e^{-\theta t} + \mu(1 - e^{-\theta t})$$

Similar to expectation, the covariance and variance are

$$\begin{aligned}\mathbb{C}\mathrm{ov}[x_s, x_t] &= \mathbb{E}[(x_s - \mathbb{E}[x_s])(x_t - \mathbb{E}[x_t])] \\ &= \mathbb{E}[\int_0^s \sigma e^{\theta(\mu - s)} dW_\mu \int_0^t \sigma e^{\theta(v-t)} dW_v] \\ &= \sigma^2 e^{-\theta(s+t)} \mathbb{E}[\int_0^s e^{\theta \mu} dW_u \int_0^t e^{\theta v} dW_v] \\ &= \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} (e^{2\theta \min(s,t)} - 1)\end{aligned}$$

$$\mathbb{V}\mathrm{ar}[x_t] = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$$

When $t \to \infty$, $\mathbb{E}[x_t] \to \mu$ and $\mathbb{V}\mathrm{ar}[x_t] \to \frac{\sigma^2}{2\theta}$.

A.3 Half-Life log Ornstein-Uhlenbeck process

The original concept of half-life comes from the physics: measuring the rate of decay of a particular substance. Half-life is the time taken by a given amount of the substance to decay to half its mass. It gives the slowness of a mean-reversion process. Derived from the SDE of log Ornstein-Uhlenbeck process

$$\mathbb{E}[dx_t] = \theta(\mu - x_t)dt$$

And the deterministic equation is therefore:

$$\frac{dx_t}{\mu - x_t} = \theta dt$$

Integrating from x_0 to the expected price at the instant t_1, denoted by x_1, then

$$\ln(\frac{x_1 - \mu}{x_0 - \mu}) = -\theta(t_1 - t_0)$$

suppose $H = t_1 - t_0$ is the time elapsed for half-life. then $H = \frac{\ln(0.5)}{-\theta} = \frac{\ln(2)}{\theta}$

APPENDIX B

Finite invariant measure on mean reversion

$\forall \phi : \mathbb{R} \to \mathbb{R}$ in C^2, consider the Smoluchowski equation [49],

$$dX_t = \phi^{'}(X_t)dt + dW_t$$

where W_t is a standard Brownian motion. Then the measure $\mu(dx) = e^{2\phi(x)}dx$ is invariant for the process X_t. A measure μ is said to be invariant for the process (X_t) if and only if

$$\int \mu(dx)P_t f(x) = \int \mu(dx) f(x)$$

for any bounded function f. Consider now an Ornstein-Uhlenbeck process with a standard deviation equal to 1, then

$$dX_t = (a - bX_t dt + dW_t), a, b > 0$$

$$\phi^{'}(x) = a - bx$$

$$\phi(x) = c + ax - \frac{bx^2}{2}$$

and $\mu(dx) = e^{-bx^2+2ax+c}dx$ is an invariant measure which is considered as a Gaussian density. In the general case of an Ornstein Uhlenbeck process reverting to the mean m, standard deviation σ, the invariant measure will be $\mathbb{N}(m, \sigma^2)$

APPENDIX C

Distribution and estimators

C.1 Joint Normal Gaussian Distribution

(X, Y) follows Normal Gaussian Distribution with four parameters, i.e. $(X, Y) \sim \mathcal{NG}(\mu, \Delta, a, b)$ if $X|Y \sim \mathcal{N}(\mu, \Delta/Y)$ and $Y \sim \mathcal{G}(a, b)$.

In other words, the probability density function of Y is

$$f(y; a, b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}, y > 0$$

And it's straightfoward to notice that $X \sim t(\mu, \frac{\Delta b}{a}, 2a)$

C.2 Derivatives of Maximum likelihood Estimator

Akaike Information criterion: $AIC = 8 - 2\log(L)$

Schwartz criterion: $BIC = -2\log(L) + 4\log(T)$