RISK PROPERTIES AND PARAMETER ESTIMATION ON MEAN REVERSION AND GARCH MODELS

by

ROELF SYPKENS

submitted in accordance with the requirements
for the degree of

MASTER OF SCIENCE

in the subject

APPLIED MATHEMATICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISER: DR E M RAPOO

SEPTEMBER 2010

List of Acronyms

ACF	- Autocorrelation Function
AIC	- Akaike Info Criterion
AR	- Autoregressive
ARCH	- Autoregressive Conditional Heteroscedasticity
ARMA	- Autoregressive Moving Average
BHHH	- Berndt-Hall-Hall-Hausman
CIR	- Cox-Ingersoll-Ross
CSS	- Conditional Sum of Squares
GARCH	- Generalised Autoregressive Conditional Heteroscedasticity
GBM	- Geometric Brownian Motion
GRG	- Generalized Reduced Gradient
JB	- Jarqu-Bera
Kurt	- Kurtosis
LM	- Lagrange Multiplier
MA	- Moving Average
MLE	- Maximum Likelihood Estimation
OLS	- Ordinary Least Squares
OUMR	- Ornstein-Uhlenbeck Mean Reverting
PACF	- Partial Autocorrelation Function
QQ-Plot	- Quintile-Quintile Plot

Introduction

Most of the notations and terminological conventions used in this thesis are Statistical. The aim in risk management is to describe the risk factors present in time series. In order to group these risk factors, one needs to distinguish between different stochastic processes and put them into different classes. The risk factors discussed in this thesis are fat tails and mean reversion. The presence of these risk factors fist need to be found in the historical dataset. I will refer to the historical dataset as the original dataset. The Ljung-Box-Pierce test will be used in this thesis to determine if the distribution of the original dataset has mean reversion or no mean reversion.

The Ljung-Box-Pierce test is explained in paragraph 2.9, page 15. The second risk property is found by applying a QQ-plot on the original dataset to determine if this dataset has normal (no fat tails) or fat tail present. The first Q stands for the quantiles of the original datasets or the simulated dataset and will be assigned as the x-coordinates. For example let the dataset used to construct the first Q be dataset 1. The second Q stands for the quantiles of a second dataset, for example dataset 2 are used as the yardstick to dataset 1. Dataset 2 is normally distributed with theoretical normal mean and theoretical normal variance with the same number of observations as dataset 1 and will be assigned as the y-coordinates. The QQ-plot will represent coordinates (x_i, y_i) for $i = 1,2,...,n$ where n is the number of observation for dataset 1 and dataset 2. If a QQ-plot is applied on a dataset 1 and the QQ-plot appears to be linear it indicates that dataset 1 has the same normal (no fat tails) properties as dataset 2. If a large amount of coordinates on both ends of the scatter plot generated by the QQ-plot bends away (deviates) from the straight line also in the QQ-plot, then this indicates that dataset 1 has fat tails.

After the risk properties for the original dataset has been found one need to simulate a model that has the same risk properties as in the original dataset. A summary of the simulated models I will be using and the risk properties associated with them are described in the following table.

Table 1.1: Stochastic Processes Used in Simulation

Distribution	**Tail of Distribution**	
Mean Reversion	**Normal**	**Fat**
No	GBM	GARCH, GBM + Jumps, One-Factor Logarithmic Vasicek Model + Jumps
Yes	Vasicek, *AR(p)*	One-Factor Logarithmic Vasicek Model, CIR Models

Key: GBM = Geometric Brownian motion;
GARCH = Generalised Autoregressive Conditional Heteroscedasticity
CIR = Cox-Ingersoll-Ross
AR(p) = autoregressive model of order p

In order to simulate a model one needs to determine the optimal estimators for these models. The risk properties for the simulated model can then be determined which should have the same risk properties as the original dataset. Interest rate processes are often

considered as mean reverting, fat tails are often present in foreign exchange processes and both risk factors (fat tails and mean reversion) are usually present in credit spreads. To test whether a model has the mean reversion property one needs to test for first-order stationarity. If a series is stationary then mean reversion is present, if not, mean reversion is not present. The presence of stationarity is often called the autoregressive (*AR*) test. The *AR* test involves calculations of the autocorrelations and partial autocorrelations to see whether there is lag in the regression.

Ordinary least squares for the initial estimators and the maximum likelihood estimation procedures to find the optimal estimators for the simulated models are discussed in chapter 1. These methods will be used in all the other chapters except chapter 3 where the optimal estimators for these simulated models are determined by minimising the conditional sum of squares.

In chapter 2 the geometric Brownian motion is discussed. The Brownian motion is the ceaseless, irregular random motion of small particles immersed in a liquid or gas as observed by R. Brown in 1827. The stochastic process associated with the Brownian motion is called the Brownian process or Wiener Process. The geometric Brownian motion (GBM) is widely used in explaining prices at equity and foreign exchange markets. I will first generate two datasets with the same characteristics as the GBM model. Risk properties for the original dataset (FTSE100 monthly closing price indices from the 2nd of April 1984 up to the 2nd of January 2001) are determined and a simulated GBM model is fitted on the original dataset. Risk properties for the simulated model are then determined and based on the similarities between the risk properties of the original dataset compared with the risk properties of the simulated model one can decide whether the simulated model is a good choice.

The autoregressive processes are useful in describing situations in which the present value of a time series depend on its preceding values and a random shock. This process is known as a mean reverting model and is discussed in chapter 3. Risk properties for the original dataset (the daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant) are determined and a simulated AR(p) model is fitted on the original dataset. Risk properties for the simulated model are then determined and based on the similarities between the risk properties of the original dataset compared with the risk properties of the simulated model one can decide whether the simulated model is a good choice.

For the modelling of electricity prices researchers prefer to use single-factor, two-factor, multifactor and hybrid models, which include factors like demand and supply. The one-factor Vasicek model on the log spot price is theoretically strong but has little practical implication. This model is discussed in chapter 4. Risk properties for the original dataset (the logarithms of Sweden's Electrical spot prices in terms of Euro per Megawatt from 1st of January 2008 to 31st of December 2008) are determined and a simulated Logarithmic Vasicek simulation model is fitted on the original dataset. Risk properties for the simulated model are then determined and based on the similarities

between the risk properties of the original dataset compared with the risk properties of the simulated model one can decide whether the simulated model is a good choice.

Chapter 5 discusses the Vasicek model. The Vasicek or Ornstein-Uhlenbeck process is the most common model used for the pricing of bonds. The main advantage of the Vasicek model is that it has an explicit solution. The Vasicek and CIR models are two important models for short rate interest rate modelling. The CIR model is discussed in chapter 7. Risk properties will also be investigated in chapters 5, 7 and 8 as what was done in chapters 2, 3 and 4. In chapters 4, 5 and 7 I also investigated the differences in accuracy between the OLS and MLE procedures influenced by the tails of the QQ-plots.

Interest rates are stochastic in nature and there are at least two factors needed in order to explain adequately their behavior thus a two-factor Vasicek is presented in chapter 6.

The *GARCH* models rely in the assumption that the volatility changes with time and the past information. The Geometric Brownian motion assumes the volatility to be constant. For the *GARCH* models it has been observed that when the volatility at a specific point in time abnormally increase or decrease (spike) in terms of the other volatilities around the vicinity of that specific point the actual data from that specific point in time will subsequently increase or decrease more as compare to the actual data around that specific point in time. For this reason, this observation is very important for forecasting purposes. An investor will now identify that at a specific point in time a particular stock will increase or decrease more rapidly as usual. The investor will stand the chance of making more profit as usual or make a bigger loss at that point in time for that particular stock. In chapter 8 the *ARCH* and *GARCH* estimators.

Contents

Introduction

Chapter 1: ESTIMATION PROCEDURES **1**

1.1 Ordinary Least Squares (OLS) Estimation In Time Series 1
1.2 Maximum Likelihood Estimation (MLE) 2

Chapter 2: NO MEAN REVERSION – GEOMETRIC BROWNIAN MOTION 3

2.1 Basic Properties – Geometric Brownian Motion (GBM) 3
2.2 Two Types of Geometric Brownian Motion Datasets 5
2.3 Example 2.1 – Generate Two Geometric Brownian Motion Datasets 6
2.4 Maximum Likelihood Estimation (MLE) – Geometric Brownian Motion 7
2.5 Confidence Intervals for Parameters of the GBM 8
2.6 Example 2.2 - Maximum Likelihood Estimation Procedure 8
2.7 Simulation for Sample Mean and Variance of GBM 11
2.8 Estimated Sample ACF and PACF 14
2.9 Ljung-Box-Pierce-test Procedure 15
2.10 Testing for Mean Reversion and Fat Tails 16
2.11 Notes and Remarks 17

Chapter 3: MEAN REVERSION – THE AUTOREGRESSIVE ORDER ONE: *AR*( ) MODEL **18**

3.1 Basic Properties - *AR(1)* Model 18
3.2 Estimation - *AR(1)* Model 20
3.3 Fitting *AR(1)* Model to Truck Manufacturing Defects Data 20
3.4 Testing for Mean Reversion and Fat Tails 23
3.5 Notes and Remarks 24

Chapter 4: MEAN REVERSION – THE ONE-FACTOR VASICEK MODEL ON LOG SPOT PRICES **25**

4.1 Basic Properties - One-Factor Vasicek Model Based On Log Spot Price 25
4.2 Explicit Solution - One-Factor Vasicek Model on Log Spot Price 27

4.3 Ordinary Least Squares (OLS) Estimation - One-Factor Vasicek Model on Log Spot Prices 29
4.3.1 Deriving the Estimators In Order to Minimise the Variance of the Errors 30
4.3.2 Calculation for the Minimum Variance of the Errors 32
4.3.3 Properties of Ito Calculus 33
4.4 Maximum Likelihood Estimation (MLE) 36
4.5 Simulation of the One-Factor Vasicel Model on the Log Spot Price 39
4.6 Example 4.1 - Estimators for Electricity Spot Prices 40
4.7 Testing for Mean Reversion and Fat Tails 44
4.8 Testing for Differences on OLS Estimation and MLE 45
4.9 Notes and Remarks 46

Chapter 5: MEAN REVERSION – THE VASICEK MODEL 47

5.1 Basic Properties - Vasicek Model 47
5.2 Maximum Likelihood Estimate (Method 1) - Vasicek Model 49
5.3 Simulation - Vasicek Model 51
5.4 Example 5.1 - Generating Original Dataset Using Vasicek Model 51
5.5 Ordinary Least Squares Estimation - Vasicek Model 53
5.6 Example 5.2 - Ordinary Least Squares Estimation 54
5.7 Maximum Likelihood Estimate (Method 2) - Vasicek Model 55
5.8 Example 5.3 - Maximum Likelihood Estimate (Method 2) 57
5.9 Maximum Likelihood Estimate (Method 3) – Vasicek Model 58
5.10 Example 5.4 – Maximum Likelihood Estimate (Method 3) 58
5.11 Testing for Mean Reversion and Fat Tails for Simulated Model 59
5.12 Notes and Remarks 61

Chapter 6: THE TWO-FACTOR VASICEK MODEL 62

6.1 Basic Properties of the Two-Factor Vasicek Model 62
6.2 Maximum Likelihood Estimation for the Two-Factor Vasicek Model 63
6.3 Notes and Remarks 65

Chapter 7: MEAN REVERSION – THE CIR MODEL 66

7.1 Basic Properties – CIR Model 66
7.2 Maximum Likelihood Estimate - CIR Model 67
7.3 The Ordinary Least Squares Estimation Procedure for CIR-Model 68
7.4 The MLE for CIR-Model Using MATLAB Command - *besseli* 69
7.5 The MLE for CIR-Model Using MATLAB Command - *ncx2pdf* 70
7.6 Example 7.1 –Estimators for the CIR Model 72
7.7 Testing for Mean Reversion and Fat Tails 73
7.8 Notes and Remarks 74

Chapter 8: NO MEAN REVERSION – AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (*ARCH*) MODELS **75**

8.1 General Properties for the *ARCH(q)* Process 75
8.2 General Properties for the *ARCH(1)* Process 76
8.3 Generalized *GARCH* Models 77
8.4 Estimation of *ARCH-GARCH* Models 78
8.5 The Berndt-Hall-Hall-Hausman Iteration Method 80
8.6 Test for Normality of Dataset 81
8.7 Application on the *ARCH* and *GARCH* Models 82
8.8 Testing the *ARCH/GARCH* Effects 85
8.9 Lagrange Multiplier (LM) Test 86
8.10 Testing *ARCH/GARCH* Effects on S&P500 87
8.11 Model Selection 88
8.12 Volatility Estimation of *ARCH/GARCH* Models 89
8.13 Notes and Remarks 90

Chapter 9: General Conclusion **91**

Appendix A: Selected Datasets **93**

A.1 FTSE100 Monthly Closing Prices 93

A.2 Daily Average Number of Truck Manufacturing Defects 93

A.3 The Daily Electricity Prices for Sweden in Euros per Megawatt (Read across) 94

Appendix B: Programs for Implementing Estimation **95**

B.1 The MATLAB program used to determine the MLE for the CIR Model using *besseli* command. 95

B.2 The MATLAB program used to determine the MLE for the CIR Model using *ncx2pdf* command. 96

B.3 The Microsoft Excel Macro used to determine the ACF 97

B.4 The MATLAB program used to determine the MLE for the one-factor logarithmic Vasicek Model using (4.42) 99

References **100**

List of Tables

1.1: Risk Properties for Simulated Stochastic Processes I

2.1: Generate Two Geometric Brownian Motions 6

2.2: Calculations for GBM Estimators 9

3.1: The $AR(\cdot)$ Model Fitting to Data, $(-0.428863)\varepsilon_t - 1.78866667) = \varepsilon_t$ 22

4.1: Ordinary Least Squares Estimation on Sweden's Dataset 41

4.2: Summary of One-Factor Logarithmic Vasicek Optimal Estimators 43

4.3: Ordinary Least Squares Estimation on Standard Normal Random Dataset 45

4.4: Summary of One-Factor Logarithmic Vasicek Optimal Estimators 45

5.1: Generating Original Dataset Using Vasicek Model 51

5.2: Calculations of Equations (5.17) 54

5.3: Optimisation of Equation (5.20) 58

5.4: Summary of Vasicek Optimal Estimators 59

6.1: Estimators for α^r, β^r and σ^r 65

6.2: Summary of the Estimators for Two-Factor Vasicek Model 65

7.1: Estimators for the CIR Model 72

8.1: $Q(\cdot)$ Using the Estimated Sample Autocorrelation of the Squared Returns 87

8.2: Calculation for $LM(\cdot)$ Before $ARCH(\cdot)$ or $GARCH(\cdot, p)$ Transformation 87

8.3: Calculation for $LM(\cdot)$ After $ARCH(\cdot)$ or $GARCH(\cdot, p)$ Transformation 88

8.4: Model Selection Test 89

8.5: Volatility Estimation of *ARCH/GARCH* Models 89

List of Figures

2.1: Generate Two Types of Geometric Brownian Motion Datasets 6

2.2: Simulation for Eight Geometric Brownian Motions 7

2.3: QQ-Plot for FTSE100 monthly closing price indices 9

2.4: FTSE100 Closing Monthly Price Indices versus Simulated GBM 11

2.5: Frequency Histogram for 10000 Sample Means 13

2.6: Adjusted Chi-Square Frequency Histogram on 10000 Sample Standard Deviations 13

2.7: QQ-plot for the original dataset: FTSE100 Monthly Indicates 16

2.8: QQ-plot for the Simulated GBM for the FTSE100 Monthly Indicates 16

3.1: The Estimated Autocorrelation for Average Number of Defects per Truck 21

3.2: The Estimated Partial Autocorrelation for Average Number of Defects per Truck 21

3.3: The $AR(1)$ Model Fitting to Data, $(-0.428863)x_t - 1.78866667 = \varepsilon_t$ 23

3.4: QQ-plot for the Simulated *AR(1)* Model for the Truck Manufacturing Defects Data 23

4.1: QQ-plot on the Logarithmic Sweden's Daily Electricity Prices Indicates 41

4.2: Simulated Electricity Prices for Sweden versus the Real Prices 44

4.3: QQ-plot for Simulated One-Factor Vasicek Model on Sweden's Daily Electricity Prices Indicates 44

4.4: QQ-plot for Simulated One-Factor Vasicek Model on Standard Normal Random Variables 46

5.1: Plot of Generated Vasicek Model 52

5.2: QQ-plot for Generated Vasicek Model 52

5.3: QQ-Plot for Simulated Vasicek Model 59

5.4: Estimated Autocorrelations for Simulated Vasicek Model 60

5.5: Estimated Partial Autocorrelations for Simulated Vasicek Model 60

7.1: QQ-Plot for 30-day AA financial interest rates 72

7.2: A Simulated CIR-Model, $x_{t_{i+1}} = \hat{\alpha}\hat{\mu}\Delta t + (1-\hat{\alpha}\Delta t)x_{t_i} + \hat{\sigma}\sqrt{x_{t_i}\Delta t}\,\varepsilon_{t_i}$ 73

7.3: QQ-plot for the Simulated CIR Model 73

8.1: The QQ-Plot of the sample variances of S&P500 Returns - $y_t = \ln(x_t / x_{t-1})$ 82

8.2: Plot of S&P500 Returns - $y_t = \ln(x_t / x_{t-1})$ 83

8.3: Plot of S&P500 Square Returns - $y_t^2 = (\ln(x_t / x_{t-1}))^2$ 83

8.4: Plot of S&P500 Absolute Returns - $|y_t| = |\ln(x_t / x_{t-1})|$ 84

8.5: Estimated Autocorrelation of $y_t = \ln(x_t / x_{t-1})$ 84

8.6: Estimated Autocorrelation $y_t^2 = (\ln(x_t / x_{t-1}))^2$ 85

8.7: Estimated Autocorrelation $|y_t| = |\ln(x_t / x_{t-1})|$ 85

Acknowledgements

I would like to thank Dr. E. Rapoo, my promoter, for her unfailing help, advice and encouragement during the course of my study. She was always friendly, helpful and supportive.

Also I want to thank my wife, Natie for all her encouragement, support and assistance while doing my script.

My mother, brother and sister, thank you for everything and for always being there for me.

My Heavenly Father for His unfailing blessing and mercy.

Chapter 1

1 ESTIMATION PROCEDURES

Before the mean reversion and no mean reversion for example the *GARCH* models are investigated the two main methods for estimation discussed in this thesis will be introduced. The true parameters will be estimated and are revered to as parameter estimates or equivalently estimators. Two procedures of calculating the estimators are ordinary least squares (OLS) estimation and maximum likelihood estimation (MLE).

1.1 Ordinary Least Squares (OLS) Estimation In Time Series

The first method is the ordinary least squares estimation procedure. According to W. S. Wei (1994)[1] the ordinary least squares (OLS) estimation procedure was developed from regression analysis. Regression analysis is probably the most common statistical method in statistical data analysis and the ordinary least squares the most common used estimation procedure in statistics. The following model is considered:

$$Y_t = \theta X_t + e_t, \qquad t = 1,2,...,n \tag{1.1}$$

The following assumptions on the error term, e_t, are made:

- The expectation of the error term is: $E[e_t] = 0$.
- The variance of the error term is: $Var(e_t) = E[e_t^2] - (E[e_t])^2 = E[e_t^2] = \sigma_e^2$, where σ_e^2 is a constant. The variance of the error is the expected squares deviation of Y_t from the line θX_t. One thus needs to minimise $Var(e_t)^2$ in order to obtain the most accurate OLS estimate.
- The autocorrelation between the error terms is: $E[e_t e_k] = 0$ for $t \neq k$.
- The error term and the explanatory variable X_t is uncorrelated: $E[X_t e_t] = 0$.

The ordinary least squares estimator for coefficient θ in equation (1.1) is:

$$\hat{\theta} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2}. \tag{1.2}$$

This estimator, $\hat{\theta}$, is the best linear unbiased estimator of θ under the assumptions of the error term. If one considers the following model:

$$Y_t = \theta Y_{t-1} + e_t, \qquad t = 1,2,...,n \tag{1.3}$$

1 WWS Wei *Time Series Analysis:* "*Univariate and Multivariate Methods*" (1990) 147.

then the ordinary least squares estimator for coefficient θ in equation (1.3) is:

$$\hat{\theta} = \frac{\sum_{t=2}^{n} Y_{t-1}Y_t}{\sum_{t=2}^{n} Y_{t-1}^2} = \frac{\sum_{t=2}^{n} Y_{t-1}(\theta Y_{t-1} + e_t)}{\sum_{t=2}^{n} Y_{t-1}^2} = \theta + \frac{\sum_{t=2}^{n} Y_{t-1}e_t}{\sum_{t=2}^{n} Y_{t-1}^2}. \quad (1.4)$$

In regression analysis, the OLS estimator of an explanatory variable will be inconsistent and **biased** except if the error term is uncorrelated with the explanatory variable. The second method is the maximum likelihood estimation procedure. This procedure is more efficient and commonly used in time series analysis.

1.2 Maximum Likelihood Estimation (MLE)[12]

Suppose we observe an original dataset $Z = z_{t_1}, z_{t_2}, ..., z_{t_n}$ with n observations. We need to apply the MLE procedure on this original dataset. Assume the true unknown conditional probability density function derived from the original dataset is $g(\cdot \mid Z, \theta_0)$ where θ_0 is the true parameter of the original dataset. One needs to choose the most appropriate simulation model for the original dataset. The simulation model will belong to some kind of conditional density function say $f(\cdot \mid Y, \theta)$ that should have the same distribution as $g(\cdot \mid Z, \theta_0)$. One needs to generate a new dataset *X* that must be used to determine the likelihood function defined as:

$$L(\theta) = \prod_{i=1}^{n} f(X \mid Y; \theta)$$

where $f(X \mid Y; \theta)$ is the conditional probability density function of X given Y and θ. Let $X = x_{t_1}, x_{t_2}, ..., x_{t_n}$ be generated from $Y = x_{t_0}, x_{t_2}, ..., x_{t_{n-1}}$ which in tern is generated by the appropriate simulation model. The initial starting value for the simulation model is the same as the starting value of the original dataset such that $x_{t_0} = z_{t_1}$. The simulation model x_{t_0} is used to simulate x_{t_1}, up until $x_{t_{n-1}}$ is used to simulate x_{t_n}. Note that parameter θ can be a vector that consists of more than one parameter. To maximise the likelihood function one can take the logarithm of the likelihood function and if the observations $X = x_{t_1}, x_{t_2}, ..., x_{t_n}$ are independent then:

$$\ln L(\theta) = \ln\left[\prod_{i=1}^{n} f(X \mid Y; \theta)\right] = \ln\left[\prod_{i=1}^{n} f(x_{t_1}, x_{t_2}, ..., x_{t_n} \mid Y; \theta)\right]$$

$$= \ln\left[\prod_{i=1}^{n} f(x_{t_1} \mid Y; \theta) f(x_{t_2} \mid Y; \theta) .. f(x_{t_n} \mid Y; \theta)\right] = \sum_{i=1}^{n} \ln f(X_{t_1} \mid Y; \theta)$$

The maximum likelihood estimator for θ_0 is defined as

$$\hat{\theta} = \arg\max_{\theta} \ln L(\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \ln f(X_{t_1} \mid Y; \theta) \quad (1.5)$$

[12] Carnegie Mellon University "Manual for the Sphinx-III recognition system" Technical version 1" http://www.speech.cs.cmu.edu/sphinxman/HMM.pdf (Accessed 21 September 2010).

Chapter 2

2 NO MEAN REVERSION – GEOMETRIC BROWNIAN MOTION

The main purposes for this chapter are:

- generate two GBM datasets
- simulate the sample mean and variance of GBM
- finding the risk properties for the original dataset (FTSE100 monthly closing price indices)
- choosing the simulated model which should have the same risk properties as the original dataset. Table 1.1 is used to find the most appropriate simulated model with the same risk properties as the original dataset
- finding the estimators for this simulated model. The method used to find the estimators used in the simulation model is called the MLE
- finding the 95% confidence interval for an unknown parameters of the original dataset
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

2.1 Basic Properties – Geometric Brownian Motion (GBM)

The Geometric Brownian Motion (GBM) is a fundamental example of a stochastic process that neither has fat tails nor mean reversion properties. In this section, the basic properties of the GBM will be stated and then a simulated GBM will be fitted on historical data to find the risk properties of the dataset. By looking at the risk properties one can determine if the GBM will be a good model to choose to fit the historical dataset. According to D.G. Luenberger (1998)[3], the GBM is the underlying process from which is derived to form the Black and Scholes formula for pricing European options. Let the continuous-time of stock prices be assigned as x_t where $\ln(x_t)$ obeys the following defined equation:

$$d \ln x_t = v dt + \sigma dW_t \tag{2.1}$$

Here v and $\sigma \geq 0$ are constants and W_t is a standard Brownian motion. In ordinary calculus one may derive that

$$d \ln x_t = \frac{dx_t}{x_t} \tag{2.2}$$

[3] DG Luenberger *Investment Science* (1998) 308.

If equation (2.2) is substituted into (2.1) one will obtain:

$$\frac{dx_t}{x_t} = vdt + \sigma dW_t \tag{2.3}$$

According to ordinary differential calculus equation (2.3) is correctly derived from (2.1) and (2.2) but due to the fact that the Wiener processes are not differentiable functions and do not follow the rules of ordinary calculus, equation (2.3) must be replaced by Ito calculus which accommodates the Wiener processes correctly. If we use Ito's Lemma (as mentioned in J.C. Hull, (2000)[12]) and which is explained in more detail in paragraph 4.2, page 27 of this thesis) the right hand side of equation (2.3) will become:

$$\frac{dx_t}{x_t} = \left(v + \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t . \tag{2.4}$$

For convenience let $v = \mu - \frac{1}{2}\sigma^2$ in equation (2.4), the Geometric Brownian motion of the asset price, x_t, over time becomes:

$$\frac{dx_t}{x_t} = \mu dt + \sigma dW_t \tag{2.5}$$

Here dW_t represents an increment to a standard Brownian motion W_t. According to Z Brzezniak (1999)[16], W_t which is also known as a noise term for $t \in (0, \infty]$ has the following properties:

- for $0 = t_0 \le t_1 \le \cdots \le t_n$, the increments $W_{t_1} - W_{t_0}, ..., W_{t_n} - W_{t_{n-1}}$ are independent.
- $\{W_t\}$ has stationary increments.
- $E[W_t] = 0$ for all t

If W_t is observed as a white noise process dW_t is quantifiable. By replacing v in equation (2.1) with $v = \mu - \frac{1}{2}\sigma^2$, equation (2.1) can be written as:

$$d \ln x_t = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t . \tag{2.6}$$

[12] JC Hull *Options Futures & Other Derivatives* 4 ed (2000) 229.
[16] Z Brzezniak *et al* "Basic Stochastic Processes"

This means that $\ln(x_t)$ is an Arithmetic Brownian Motion. By integrating equation (2.6) between, t and u according to Damiano Brigo *et al* (2007)[4], gives:

$$\ln x_u - \ln x_t = \left(\mu - \frac{1}{2}\sigma^2\right)(u-t) + \sigma(W_u - W_t) \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)(u-t), \sigma^2(u-t)\right) \quad (2.7)$$

According to Damiano Brigo *et al* (2007)[4], by letting $u = T$, $t = 0$ and taking the exponent on equation (2.7) leads to:

$$x_T = x_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right). \quad (2.8)$$

The mean and the variance of x_T according to Damiano Brigo *et al* (2007)[4] are:

$$E[x_T] = x_0 e^{\mu T} \qquad \text{and} \qquad Var[x_T] = e^{2\mu T} x_0^2 \left(e^{\sigma^2 T} - 1\right). \quad (2.9)$$

2.2 Two Types of Geometric Brownian Motion Datasets

One needs to represent a mathematical model (simulated model) that present an approximation of the historical dataset. In order to find an appropriate simulated model one needs to look at the risk properties of the original dataset. After the appropriate simulated model has been chosen, assumptions can be made of how the dataset might react after the last available time. According to D.G. Luenberger (1998)[3] let $\Delta t = t_{i+1} - t_i$ and for $t_0 < t_1 < \cdots < t_n$ where $t = t_0$ is the time for the initial price. The simulation equation derived from equation (2.5) is:

$$x_{t_{i+1}} = \left(1 + \mu\Delta t + \sigma Z_i \sqrt{\Delta t}\right) x_{t_i}. \quad (2.10)$$

Here Z_i is a random variable from an independent identically standard normal distribution, with a mean of zero and a variance of one. Another version of a simulation equation for the GBM can be obtained by using equation (2.1) and the fact that $v = \mu - \frac{1}{2}\sigma^2$:

$$\ln\left(x_{t_{i+1}}\right) - \ln\left(x_{t_i}\right) = v\Delta t + \sigma Z_i \sqrt{\Delta t}. \quad (2.11)$$

---

[4] D Brigo *et al* "A Stochastic Processes Toolkit for Risk Management" (2007) *SSRN* http://ssm.com/abstract=1109160 (Accessed 9 September 2009).
[3] DG Luenberger *Investment Science* (1998) 308.

By taking the exponent of both sides of equation (2.11), it yields:

$$x_{t_{i+1}} = x_{t_i} \exp\left(v\Delta t + \sigma Z_i \sqrt{\Delta t}\right). \tag{2.12}$$

In practice the two different simulation models are equally good.

In the following example two methods are shown to generate approximately the same Geometric Brownian motion.

2.3 Example 2.1 – Generate Two Geometric Brownian Motion Datasets

Let $\Delta t = 1/52$, the initial price $x_{t_0} = 100$, $v = 0.3$, $\mu = 0.425$, and the volatility $\sigma = 0.5$. The second last column on the right of the following table illustrated below is the simulation of the spot price based on equation (2.10) and the last column the simulation of the spot price according to equation (2.12). Both simulations depend on the same $Z_i \sim N(0,1)$.

Table 2.1: Generate Two Geometric Brownian Motions

i	t_i	Z_i	x_{t_i} (2.10)	x_{t_i} (2.12)
0	0		100	100
1	0.019230769	0.114289378	101.6097619	101.3787962
⋮	⋮	⋮	⋮	⋮
51	0.980769231	1.116391104	100.7722344	103.2251215
52	1	0.2270815	103.1825395	105.4700207

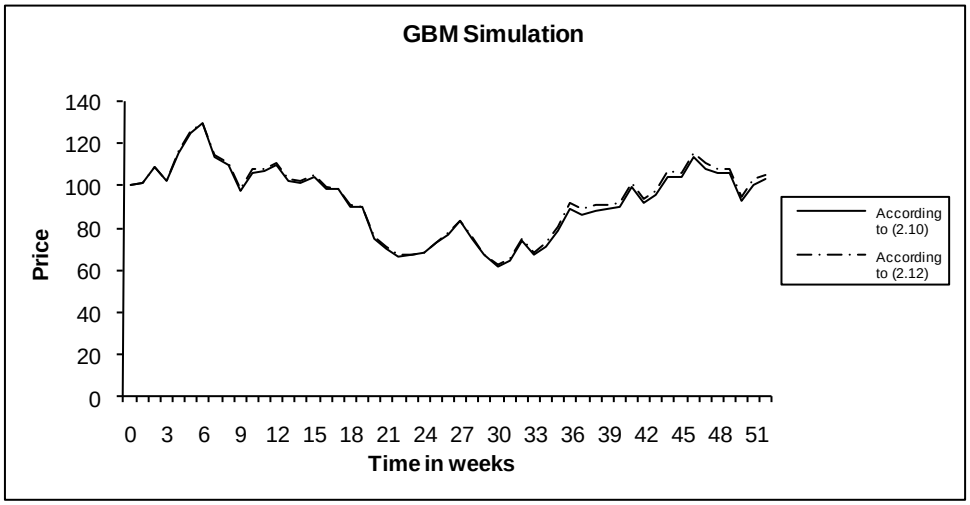

Figure 2.1: Generate Two Types of Geometric Brownian Motion Datasets

The generated Geometric Brownian Motions according to equations (2.10) and (2.12) are simultaneous plotted in Figure 2.1. The following Figure 2.2 illustrates eight simulations

of prices all with initial spot price of $x_{t_0} = 100$ and $\Delta t = 1/365,\ v = 0.15$, and the volatility $\sigma = 0.7$. Note that $t_0 < \cdots < t_{500}$ and simulations are based on equation (2.12).

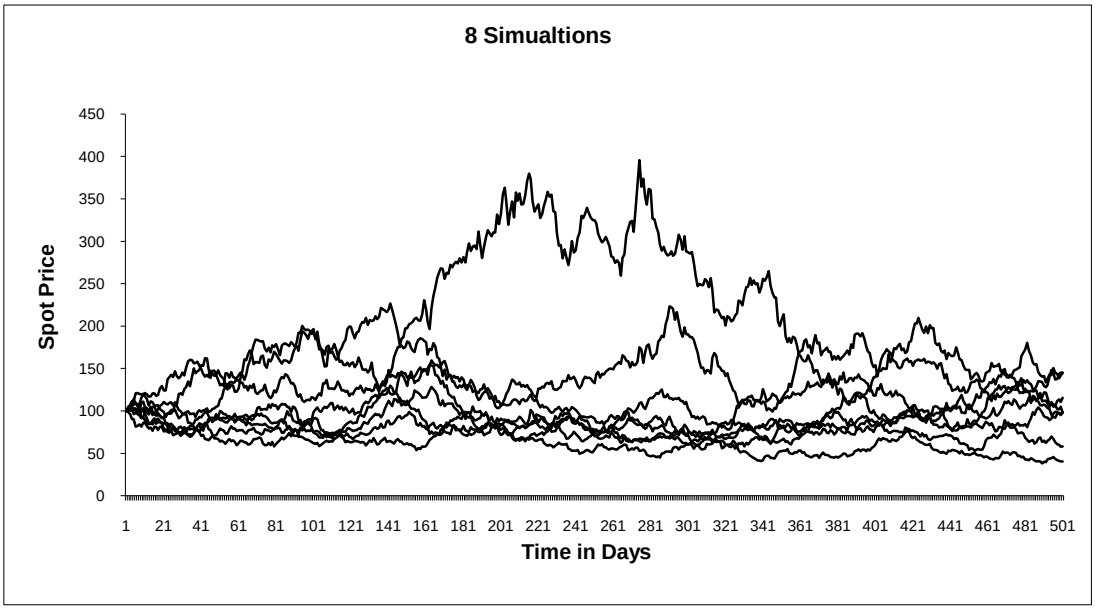

Figure 2.2: Simulation for Eight Geometric Brownian Motions

By observing Figure 2.2 it is noted that not any of the eight simulated Geometric Brownian Motions show evidence of mean reversion.

2.4 Maximum Likelihood Estimation (MLE) – Geometric Brownian Motion

According to Damiano Brigo *et al* (2007)[6], the parameters that need to be optimised are $\theta = (\mu, \sigma)$ for the GBM. Let the log return be given as:

$$y_{t_i} = \ln x_{t_i} - \ln x_{t_{i-1}}. \tag{2.13}$$

According to equation (2.7) $y_{t_i} = \ln x_{t_i} - \ln x_{t_{i-1}}$ is normally distributed for all $y_{t_1}, y_{t_2}, ..., y_{t_n}$ and assume independence for $y_{t_1}, y_{t_2}, ..., y_{t_n}$. The likelihood function will be denoted as:

$$\begin{aligned} L(\theta) &= f_\theta\left(y_{t_1}, y_{t_2}, ..., y_{t_n}\right) \\ &= \prod_{i=1}^{n} f_\theta\left(y_{t_i}\right) = \prod_{i=1}^{n} f\left(y_{t_i} \mid \theta\right). \end{aligned}$$

[6] Brigo *et al* "A Stochastic Processes Toolkit for Risk Management".

Here f_θ is the probability density function, $y_{t_1}, y_{t_2}, ..., y_{t_n}$ are the log returns according to equation (2.13). Let $\theta = (\mu, \sigma)$, then the probability density function f_θ is:

$$f_\theta(y_{t_i}) = \frac{1}{x_{t_i}\sigma\sqrt{2\pi t}} \exp - \frac{\left(\left(\frac{y_{t_i}}{y_{t_0}}\right) - \left(\mu - \frac{1}{2}\sigma^2\right)t\right)^2}{2\sigma^2 t}.$$

The likelihood function needs to be maximised to obtain the optimal estimators $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$. The natural logarithm of the likelihood function must be differentiated in terms of μ and σ then equated to zero which will yield two equations and must be solved simultaneously to obtain:

$$\hat{w} = \left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2\right)\Delta t \tag{2.14}$$

$$\hat{v} = \hat{\sigma}^2 \Delta t \tag{2.15}$$

where

$$\hat{w} = \sum_{i=1}^{n} \frac{y_{t_i}}{n} = \frac{\ln x_{t_n} - \ln x_{t_0}}{n} \tag{2.16}$$

$$\hat{v} = \sum_{i=1}^{n} \frac{\left(y_{t_i} - \hat{w}\right)^2}{n}. \tag{2.17}$$

First one needs to determine $\hat{w}$ and $\hat{v}$ then the MLE are: $\hat{\sigma}^2 = \frac{\hat{v}}{\Delta t}$ and $\hat{\mu} = \frac{1}{2}\hat{\sigma}^2 + \frac{\hat{w}}{\Delta t}$.

2.5 Confidence Intervals for Parameters of the GBM

The 95% confidence interval for the parameter w according to Damiano Brigo *et al* (2007)[7] is:

$$\hat{w} - 1.96\frac{\sqrt{\hat{v}}}{\sqrt{n}} \le w \le \hat{w} + 1.96\frac{\sqrt{\hat{v}}}{\sqrt{n}} \tag{2.18}$$

and the 95% confidence for the sample variance of the GBM v is:

$$\frac{n\hat{v}}{\chi^2_{n,0.025}} \le v \le \frac{n\hat{v}}{\chi^2_{n,0.975}} \tag{2.19}$$

Here $\chi^2_{n,0.025}$ and $\chi^2_{n,0.975}$ are the upper and lower percentiles of the chi-squared distributions with n degrees of freedom. Confidence intervals are used to find the upper and lower regions of the true parameters w and v. The reason why one needs to estimate

[7] D Brigo *et al* "A Stochastic Processes Toolkit for Risk Management".

parameters w and v first, is because $\hat{w}$ and $\hat{v}$ are used to estimate σ^2 and μ in order to determine MLE.

2.6 Example 2.2 - Maximum Likelihood Estimation Procedure

First the risk properties of the dataset which in this example is the FTSE100 monthly closing price indices from the 2nd of April 1984 up to the 2nd of January 2001 must be calculated. The Ljung-Box-Pierce test defined in paragraph 2.8 page 14 is:

$$Q = 3784.976 > \chi^2_{50,0.05} = 67.504$$

The corresponding $p-value$ is 2.2e-16 which is less than 0.05. The null hypothesis (H_0) will be rejected indicating that the distribution of the original dataset is not mean reverting.

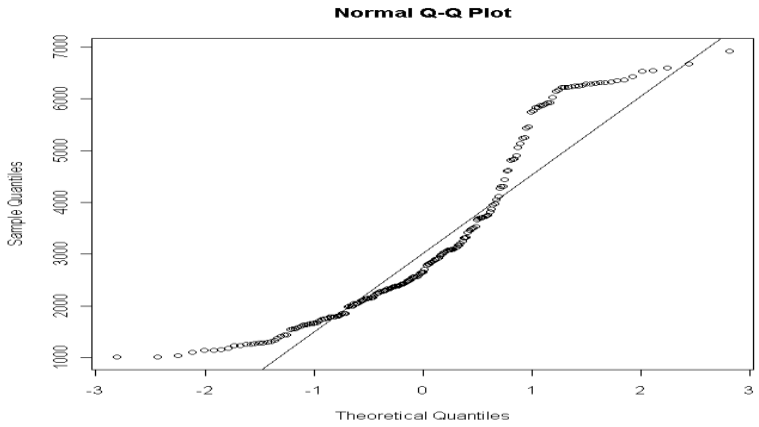

Figure 2.3: QQ-Plot for FTSE100 monthly closing price indices

Fat tails are observed in the QQ-plot for the FTSE100 monthly closing price indices. According to Table 1.1 the most appropriate simulated models to choose for similar risk properties of the original dataset are GBM + Jumps or One-Factor Logarithmic Vasicek Model + Jumps. The simulated model that I choose for this dataset is the GBM which will not be the most appropriate simulation model. I choose this model for example purposes only. In this example the estimators of the FTSE100 monthly closing price indices from the 2nd of April 1984 up to the 2nd of January 2001 are calculated by using the maximum likelihood estimation procedure. The equations used are (2.16) and (2.17).

Table 2.2: Calculations for GBM Estimators

Date	i	t_i	x_{t_i}	$\ln(x_{t_i})$	y_{t_i}
1984/04/02	0	0.000000	1138.3	7.0372912	
1984/05/01	1	0.083333	1016.6	6.924219005	-0.1130722
⋮	⋮	⋮	⋮	⋮	⋮
2001/01/02	201	16.750000	6297.5	8.747908008	0.011980974

According to equation (2.16):

$$\hat{w} = \sum_{i=1}^{n} \frac{y_i}{n} = \frac{1.710617}{201} = 0.008510531 \tag{2.20}$$

and

$$\hat{w} = \frac{\ln x_{t_n} - \ln x_{t_0}}{n} = \frac{(8.747908008 - 7.0372912)}{201} = 0.008510531.$$

By using Table 2.2 and applying equation (2.17) one will obtain:

$$\hat{v} = \sum_{i=1}^{n} \frac{(y_i - \hat{w})^2}{n} = \frac{0.476079261}{201} = 0.002360452. \tag{2.21}$$

The 95% confidence interval for the parameter w (sample mean of the GBM) is:

$$0.008510531 - 1.96\frac{\sqrt{0.002360452}}{\sqrt{201}} \le w \le 0.008510531 + 1.96\frac{\sqrt{0.002360452}}{\sqrt{201}}$$
$$0.001793836 \le w \le 0.015227227$$

and the 95% confidence for the sample variance of the GBM v is:

$$\frac{201(0.002360452)}{242.1559} \le v \le \frac{201(0.002360452)}{163.63}$$
$$0.001959279 \le v \le 0.002899535.$$

We have 95% confidence that the true parameter estimated with w will lie between 0.001793836 and 0.015227227. The true variance of the GBM will lie between 0.001959279 and 0.002899535 with a 95% level of confidence.

Solving equations (2.14) and (2.15) yield

$$\hat{\sigma}^2 = \frac{\hat{v}}{\Delta t} = \frac{0.002360452}{(1/12)} = 0.02832543 \tag{2.22}$$

$$\hat{\sigma} = \sqrt{\frac{\hat{v}}{\Delta t}} = 0.168301603 \tag{2.23}$$

$$\hat{\mu} = \frac{\hat{w}}{\Delta t} + \frac{1}{2}\hat{\sigma}^2 = \frac{0.008510531}{(1/12)} + \frac{1}{2}(0.02832543) = 0.116289091. \tag{2.24}$$

Fitting the GBM to the FTSE100 monthly closing prices will involve equation (2.10), $\hat{\mu}$ and $\hat{\sigma}^2$ such that:

$$x_{t_{i+1}} = \left(1 + \hat{\mu}\Delta t + \hat{\sigma} Z_i \sqrt{\Delta t}\right) x_{t_i} . \tag{2.25}$$

Let $\Delta t = \frac{1}{12}$, and $Z_i \sim N(0,1)$ independent identical random variables and let the initial price index be $x_{t_0} = 1138.3$. The following Figure 2.4 will plot the GBM simulation (2.25) together with the original FTSE100 closing monthly price indices.

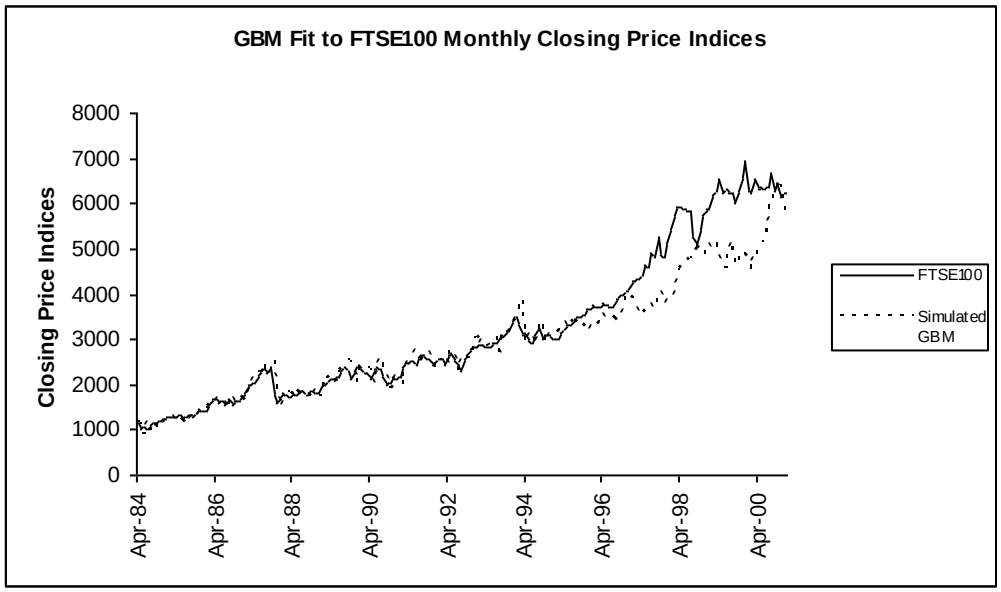

Figure 2.4: FTSE100 Closing Monthly Price Indices versus Simulated GBM

By looking at Figure 2.4 it is clear that neither the FTSE100 closing monthly price indices nor the simulated GBM model shows mean reversion.

2.7 Simulation for Sample Mean and Variance of GBM

From a total number of 201 logarithmic returns for the FTSE100 monthly closing prices the maximum likelihood estimates for the mean and the standard deviation are given in equations (2.20) and (2.22). Suppose we generate 10000 datasets each containing 201 normal random values with mean $\hat{w} = 0.008510531$ and variance $\hat{\sigma}^2 = 0.0283254$3.Let the datasets be denoted as D_{ij}, where $i = 1,2,...10000$ (indicating the datasets) and $j = 1,2,3,...201$ (amount of random values in each set).

Ten thousand sample mean estimates for each dataset are determined as:

$$\hat{W}_1 = \frac{1}{201}\sum_{j=1}^{201} D_{1j}$$
$$\vdots$$
$$\hat{W}_{10000} = \frac{1}{201}\sum_{j=1}^{201} D_{10000\,j}$$

for random values D_{ij} that are normally distributed with mean 0.008510531 and variance 0.02832543. The estimated variances for these datasets are:

$$\hat{V}_1 = \sum_{j=1}^{201} \frac{\left(D_{1j} - \hat{W}_1\right)^2}{201}$$
$$\vdots$$
$$\hat{V}_{10000} = \sum_{j=1}^{201} \frac{\left(D_{10000\,j} - \hat{W}_{10000}\right)^2}{201}$$

Let $\Delta t = 1/12$ for the FTSE100 monthly closing prices and substitute it into a similar equation as equation (2.23). The sample standard deviations estimates for 10000 datasets are:

$$\hat{S}_1 = \sqrt{\frac{\hat{V}_1}{\Delta t}}$$
$$\vdots$$
$$\hat{S}_{10000} = \sqrt{\frac{\hat{V}_{10000}}{\Delta t}}\,. \qquad (2.26)$$

Similar equations as (2.24) are used to derive 10000 sample mean estimates defined as:

$$\hat{A}_1 = \frac{\hat{W}_1}{\Delta t} + \frac{1}{2}\left(\hat{S}_1\right)^2$$
$$\vdots$$
$$\hat{A}_{10000} = \frac{\hat{W}_{10000}}{\Delta t} + \frac{1}{2}\left(\hat{S}_{10000}\right)^2. \qquad (2.27)$$

The frequency histogram in Figure 2.4 below for equations (2.27) indicate that the sample means for 10000 datasets do not deviate much from the first estimated sample mean $\hat{\mu} = 0.116289091$.

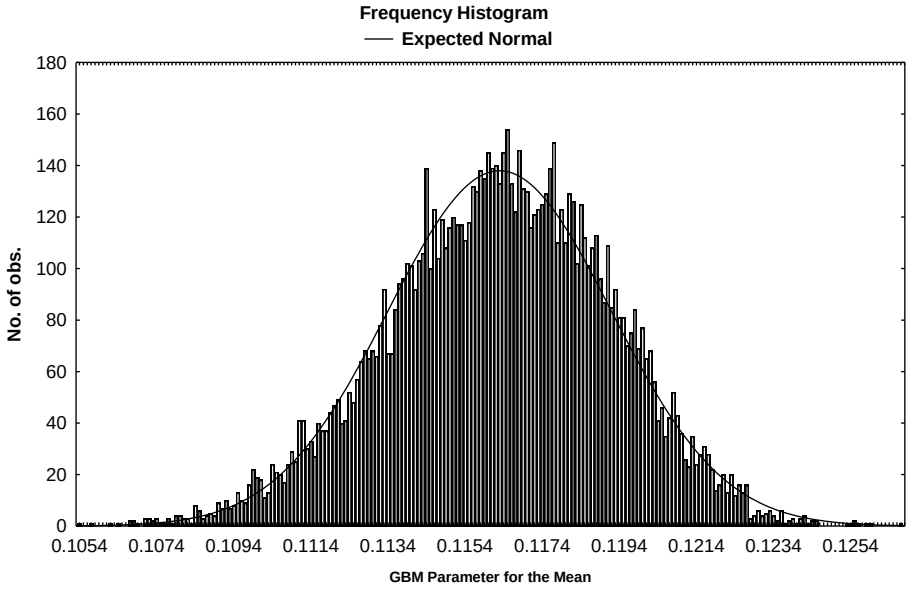

Figure 2.5 Frequency Histogram for 10000 Sample Means

The frequency histogram, for 10000 sample standard deviations derived in equations (2.26), using an adjusted chi-squared distribution is displayed in Figure 2.6 below. The adjusted residuals are used to determine the adjusted chi-squared distribution. The reason for using the adjusted chi-squared distribution rather than the chi-squared distribution is that a shape which is less skew can be fitted on this frequency histogram indicating where the sample standard deviations cluster.

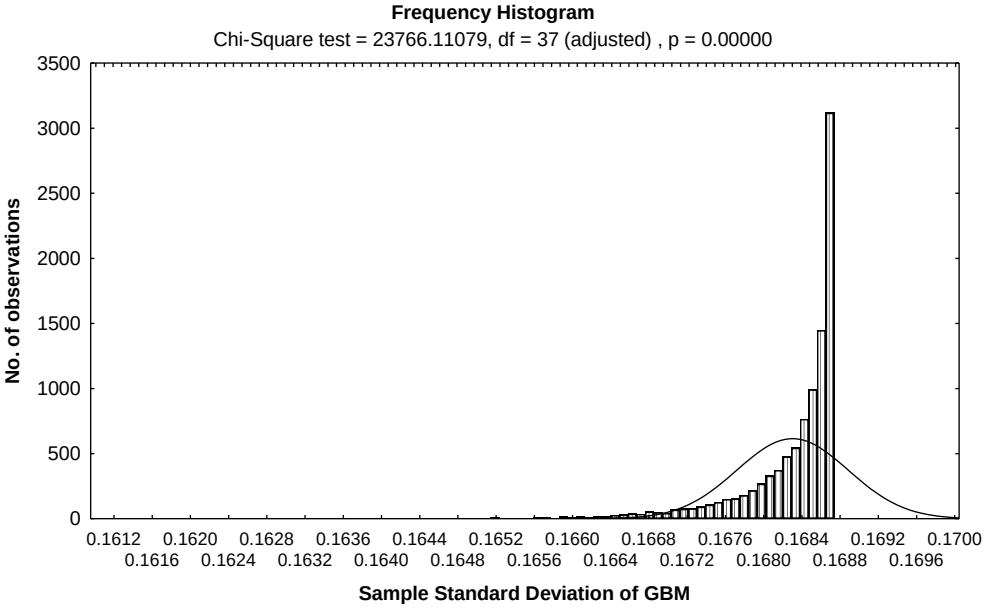

Figure 2.6: Adjusted Chi-Square Frequency Histogram on 10000 Sample Standard Deviations

In both cases the two estimators are very close to their theoretical limits. Estimated sample standard deviations indicated in equations (2.26) seems to cluster around the estimated sample standard deviation $\hat{\sigma} = 0.168301603$ indicated in (2.23). Before one needs to determine the mean reversion in the simulated models one needs to introduce the following test.

2.8 Estimated Sample ACF and PACF

The autocorrelation functions (ACF or ρ_k) and partial autocorrelation functions (PACF or ϕ_{kk}) are estimated by the sample autocorrelation function, r_k, and the sample partial autocorrelation, f_{kk}.[9] For a given time series, $x_1, x_2, ..., x_n$, the autocorrelation at lag k is estimated by:

$$r_k = \frac{c_k}{c_0} \tag{2.28}$$

where c_k is known as the estimated autocorrelation at lag k defined as:

$$c_k = \begin{cases} \dfrac{1}{n-|k|}\displaystyle\sum_{t=1}^{n-|k|}(x_t - \bar{x})(x_{t+k} - \bar{x}) & if \quad |k| < n \\ 0 & otherwise \end{cases}$$

where c_0 is known as the estimated variance at lag 0 defined as:

$$c_0 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2 \ .$$

The mean of the time series is defined as:

$$\bar{x} = \frac{1}{n}\left(\sum_{t=1}^{n} x_t\right).$$

The estimated partial autocorrelation, f_{kk} at lag k is given as:

$$f_{kk} = \frac{|P_k^*|}{|P_k|} \tag{2.29}$$

where $|P_k|$ denotes the determinant of the matrix:

[9] DG Nel *Time Series Analysis Module* (1994) 7.

$$P_k = \begin{bmatrix} 1 & r_1 & r_2 & r_3 & \cdots & r_{k-2} & r_{k-1} \\ r_1 & 1 & r_1 & r_2 & \cdots & r_{k-3} & r_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k-1} & r_{k-2} & r_{k-3} & r_{k-4} & \cdots & r_1 & 1 \end{bmatrix}$$

and where matrix, P_k^*, is:

$$P_k^* = \begin{bmatrix} 1 & r_1 & r_2 & r_3 & \cdots & r_{k-2} & r_1 \\ r_1 & 1 & r_1 & r_2 & \cdots & r_{k-3} & r_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k-1} & r_{k-2} & r_{k-3} & r_{k-4} & \cdots & r_1 & r_k \end{bmatrix}.$$

2.9 Ljung-Box-Pierce-test Procedure

The Ljung-Box-Pierce-test procedure explained by G.M. Ljung and G.E.P. Box (1979)[5] is used to test for the significance of the ACF and PACF. A set of m autocorrelations may be tested at once by using the Ljung-Box-Pierce test. The hypothesis test is used when testing whether the true autocorrelations, ρ_k, are significant:

$H_0: \rho_k = 0,\ k \le m$ (dataset follows mean reversion/random/white noise process)

versus

$H_1: \rho_k \neq 0,\ k \le m$ (dataset does not follow mean reversion/random/white noise process).

If H_0 is not rejected then it indicates that the time series is random and follows a white noise process. On the other hand, if H_0 is rejected it indicates that the time series shows a clear pattern and is not random indicating serial correlation is present.

The test statistic is:

$$Q = n(n-2)\sum_{k=1}^{m} \frac{r_k^2}{(n-k)} \tag{2.30}$$

where m is the number of sample autocorrelation function, r_k, be tested. The sample autocorrelation function, r_k is defined at equation (2.30). The number of sample

[5] GM Ljung and GEP Box "The likelihood function of stationary autoregressive-moving average models" (1979) *Biometrika* 66 265 – 270.

autocorrelation functions is, $m \approx \frac{n}{4}$, where n is the total number of observations of the original dataset . We reject H_0 if $Q > \chi^2_{m;\alpha}$ where $\chi^2_{m;\alpha}$ is a chi-square distribution with m degrees of freedom and α level of significance.

2.10 Testing for Mean Reversion and Fat Tails

The FTSE100 monthly stock index will be modeled by making use of the GBM. The FTSE100 monthly closing prices from the 2nd of April 1984 up to the 2nd of January 2001 are used. The FTSE100 dataset is available in Appendix A1. The QQ-plot for the original dataset (FTSE100 monthly closing prices) is found to have fat tails (Figure 2.7) as well as the QQ-plot for the simulated GBM model. The simulated GBM is derived from equation (2.10). The QQ-plot for the original dataset (FTSE100 monthly closing prices) and the simulated GBM is plotted in Figure 2.7 and Figure 2.8 respectively.

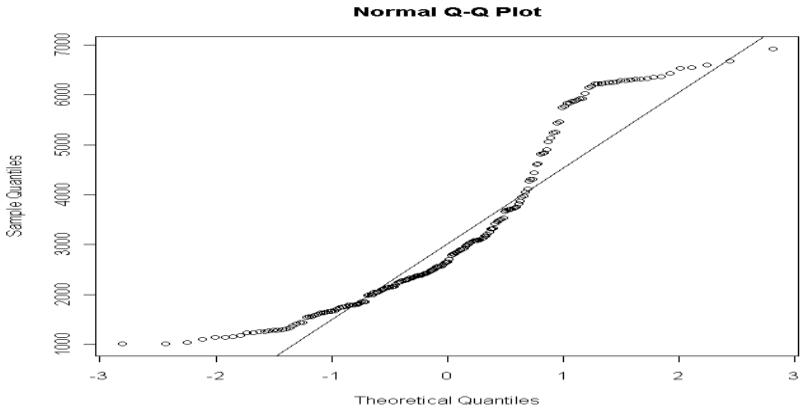

Figure 2.7: QQ-plot for the original dataset: FTSE100 Monthly Indicates

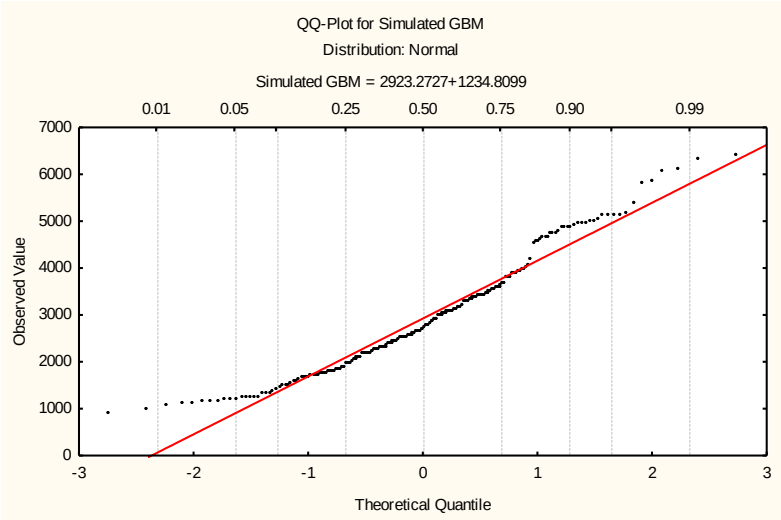

Figure 2.8: QQ-plot for the Simulated GBM for the FTSE100 Monthly Indicates

Let $\Delta t = \frac{1}{12}$, and $Z_i \sim N(0,1)$ independent identical random variable and the initial price index be $x_{t_0} = 1138.3$. The general formula for the estimated autocorrelation function if we use a time series $x_{t_1}, x_{t_2}, ..., x_{t_n}$ is:

$$r_k = \frac{1}{(n-k)\hat{v}} \sum_{i=1}^{n-k} \left(x_{t_i} - \hat{w}\right)\left(x_{t_{i+k}} - \hat{w}\right), \quad k = 1,2,... \tag{2.31}$$

where $\hat{v}$ and $\hat{w}$ are defined in equation (2.14) up to equation (2.17).

2.11 Notes and Research

Firstly I noticed that the FTSE100 monthly closing prices produced risk properties similar to the GBM + Jump and the One-Factor Logarithmic Vasicek Model + Jump models. The GBM will be an inappropriate simulation model to use. I used the GBM for example purposes only. After the GBM is simulated on the original dataset the risk properties of the simulated model are investigated.

If H_0 is rejected for the Ljung-Box-Pierce test on the simulated model it implies that there is mean reversion of the simulated GBM model. For fifteen sample autocorrelations, the test statistic $Q = 2067 > \chi^2_{15,0.05} = 24.9957$ indicates H_0 will be rejected, meaning the simulated GBM model does not follow a white noise process and significant autocorrelations are present thus mean reversion is not present (no stationarity present).

Both the original dataset as well as the simulated GBM model have fat tails and no mean reversion properties, meaning that the FTSE100 monthly closing prices roughly follows the risk profile of the GBM. One may choose a more appropriate simulation model based on the risk properties of the original dataset based on Table 1.1 and based on these risk properties one may choose the GBM with jumps, which have fat tails but no mean reversion.

In paragraph 2.3 if standard normal random variable were used to generate a GBM as indicated in the generation of equation (2.10) and plotted in Figure 2.1, the QQ-plot indicate short tails (normal tails) and the test for mean reversion of the 52 generated GBM model is $Q = 214.0 > \chi^2_{15,0.05} = 24.9957$. The dataset generated from equation (2.10) fits the risk properties of the simulated GBM based on Table 1.1. For this reason the generated GBM will be well fitted with a simulated GBM.

Chapter 3

3 MEAN REVERSION – THE AUTOREGRESSIVE ORDER ONE: $AR(1)$ MODEL

The main purposes for this chapter are:

- finding the risk properties for the original dataset (daily average number of defects per truck)
- choosing the simulated model which should have the same risk properties as the original dataset. Table 1.1 is used to find the most appropriate simulated model with the same risk properties as the original dataset
- using the conditional sum of squares method of finding the optimal estimator to be used for this simulated model
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

3.1 Basic Properties - *AR*(1) Model

The property of mean reversion means that the time series process always reverts to a certain constant with a limited variance (variation) or deviation around the reverted constant. The stationary autoregressive model of order one has this property.

Suppose x_t is a time series (not necessarily stationary) such that x_t is dependent on x_{t-1} and a random shock, ε_t, where $E(\varepsilon_t)=0$, $Var(\varepsilon_t)=\sigma_e^2$ and $E(\varepsilon_t\varepsilon_s)=0$ for $t \neq s$. Then the model of x_t can be written as:

$$x_t = \alpha(x_{t-1}) + \varepsilon_t \tag{3.1}$$

This is called a simple autoregressive model of the first order and is denoted as $AR(1)$. This model is also known as the Markov model. The $AR(1)$ as in (3.1), is centered around zero but the model can also be centered around μ indicated as:

$$(x_t - \alpha(x_{t-1})) - \mu = \varepsilon_t \tag{3.2}$$

This process will be stationary and invertible if $|\alpha| < 1$ such that $(1-\alpha B)^{-1} < \infty$. A stationary or a mean reverting dataset will have an overall pattern that is parallel with the $x-$axis. The inverted model from (3.1) is:

$$x_t = (1-\alpha B)^{-1}\varepsilon_t \tag{3.3}$$

where B is a back shift operator such that $x_t B = x_{t-1}$ and for a constant $\alpha B = \alpha$. Equation (3.3) can be extended to:

$$\begin{aligned} x_t &= \left(1 + \alpha B + \alpha^2 B^2 + \cdots\right)\varepsilon_t \qquad \text{for } |B\alpha| < 1 \\ &= \left(1 + \psi_1 B + \psi_2 B^2 + \cdots\right)\varepsilon_t \\ &= \Psi(B)\varepsilon_t \end{aligned}$$

with $\psi -$weights: $\psi_j = \alpha^j$. The autocovariances for the $AR(1)$ model are:

$$\begin{aligned} \sigma_k &= \sigma_e^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \\ &= \sigma_e^2 \sum_{j=0}^{\infty} \alpha^j \alpha^{j+k} \\ &= \sigma_e^2 \alpha^k \sum_{j=0}^{\infty} \alpha^{2j} \\ &= \frac{\sigma_e^2 \alpha^k}{1-\alpha^2}. \end{aligned}$$

The variance of the $AR(1)$ model is then:

$$Var(x_t) = \sigma_0 = \frac{\sigma_e^2}{1-\alpha^2}.$$

The aurocorrelation at lag k for the *AR(1)* model is:

$$\rho_k = \frac{\sigma_k}{\sigma_0} = \alpha^k.$$

The partial autocorrelations are:

$$\phi_{kk} = \begin{cases} \alpha & if \quad k = 1 \\ 0 & if \quad k > 1 \end{cases}$$

Equation (3.1) can be modified in such a way that it is centered to its mean μ to become equation (3.2). Substitute $\varepsilon_t = \sigma_e^2 \zeta_t$, where ζ_t are independent uncorrelated normal $N(0,1)$ random variables, into equation (3.2) then:

$$\left(x_t - \alpha\left(x_{t-1}\right)\right) - \mu = \sigma_e^2 \zeta_t \qquad (3.4)$$

Equation (3.4) can now be expressed by:

$$\left(x_t - \alpha\left(x_{t-1}\right)\right) = \mu + x_{t-1} - x_{t-1} + \sigma_e^2 \zeta_t \tag{3.5}$$

then by rearranging (3.5) gives:

$$\begin{aligned}\Delta x_t = \left(x_t - x_{t-1}\right) &= \mu - x_{t-1} + \alpha\left(x_{t-1}\right) + \sigma_e^2 \zeta_t \\ &= \left(1-\alpha\right)\left(\frac{\mu}{\left(1-\alpha\right)} - x_{t-1}\right) + \sigma_e^2 \zeta_t .\end{aligned}$$

3.2 Estimation - *AR*(1) Model

By investigating the significance of the autocorrelation functions (ACF's) and partial autocorrelation functions (PACF's) of x_t we choose the order for p and q. The method of testing the significance of the ACF's and PACF's is known as the Ljung-Box-Pierce-test procedure explained by W. S. Wei (1994).[8] If one draw the ACF and PACF for different lags of the $AR(p)$ model one will observe that the ACF will tails off as exponential decay or damped sine waves and the PACF will cuts off after lag p.

By using the Box-Jenkins estimation procedure, which is explained by W. S. Wei (1994),[8] one will obtain the conditional sum of squares, $CSS = \sum_{t=1}^{n} \varepsilon_t^2$, which will be minimised by the ordinary least squares method to find the optimal value for α. One can choose any initial value for α such that $|\alpha| < 1$ and use equation (3.2) to determine $\varepsilon_1, \ldots, \varepsilon_n$. Determine CSS for different values of α. The α value for the smallest CSS will be optimal value for α assigned as $\hat{\alpha}$ and used in the simulated *AR*(1) model.

3.3 Fitting *AR(1)* Model to Truck Manufacturing Defects Data

The dataset in this section are available in Appendix A2. This dataset is the daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. The data consists of 45 daily observations of consecutive business days from the 4th of November to the 10th of January. One needs to determine the risk properties of this dataset. The Ljung-Box-Pierce test on this dataset is:

$$Q = 15.7287 < \chi^2_{11,0.05} = 19.67514$$

The corresponding $p-value$ is 0.1515 which is larger than 0.05. The null hypothesis (H_0) will not be rejected indicating that the distribution of the original dataset is mean

[8] Wei *Time Series Analysis* 106.

reverting. Indecisive tails are observed in the QQ-plot. Based on the risk properties calculated for the average number of defects per truck if one assume normal tails one may choose the $AR(p)$ simulated model.

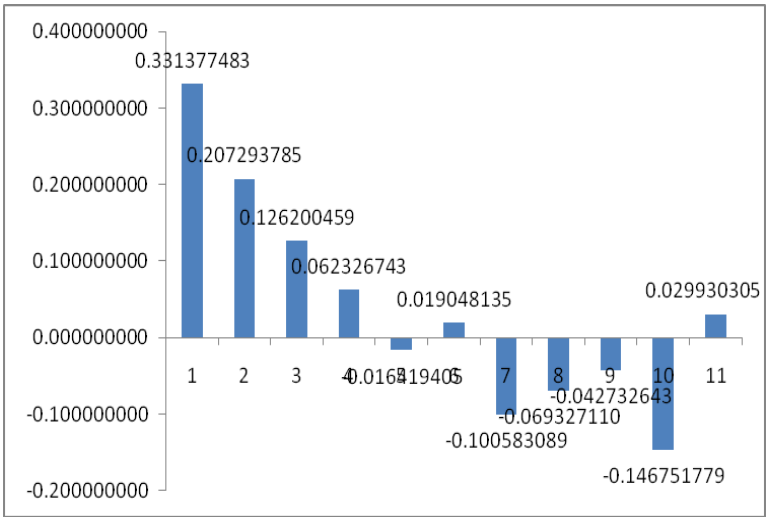

Figure 3.1: The Estimated Autocorrelation for Average Number of Defects per Truck

Figure 3.1 presents the estimated ACF for eleven lags. By using paragraph 3.2 one can observe that these ACF values indicate damped sine waves.

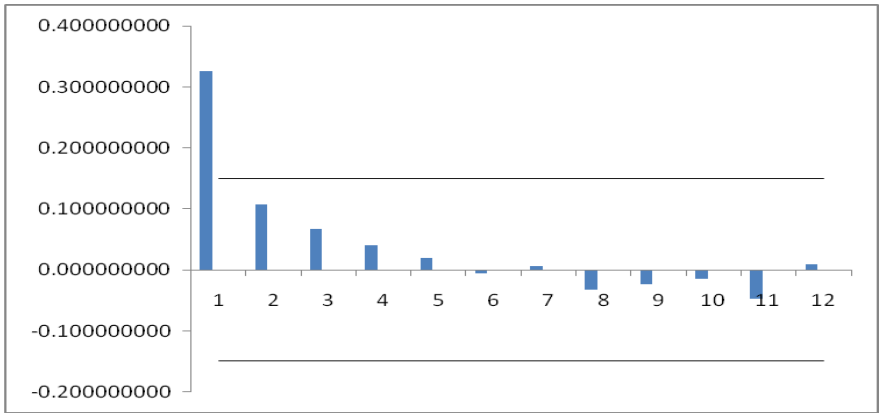

Figure 3.2: The Estimated Partial Autocorrelation for Average Number of Defects per Truck

Figure 3.2 presents the estimated PACF for eleven lags. By using paragraph 3.2 one can observe that these PACF values will cuts off after lag order $p=1$. For this reason I will choose an $AR(1)$ simulation model.

The $AR(1)$ series $x_t = \alpha x_{t-1} + \varepsilon_t$ will be fitted to the dataset. One needs to center this dataset around its mean.

Let $\mu = \frac{1}{45}\left(\sum_{t=1}^{45} x_t\right) = 1.78866667$. Now let $x_t - \mu = z_t$ such that equation (3.1) becomes:

$$z_t - \alpha(z_{t-1}) = \varepsilon_t .$$

Let $\varepsilon_1 = 0$, then $\varepsilon_2 = z_2 - \alpha z_1$, $\varepsilon_3 = z_3 - \alpha z_2$,..., $\varepsilon_{45} = z_{45} - \alpha z_{44}$. One needs to take the conditional sum of squares (*CSS*), which will be minimised by choosing the optimal value for α. One may use Microsoft Excel's solver add-inn functionality to minimise the conditional sum of squares with respect to α. The following table gives the optimal value of α assigned as $\hat{\alpha}$ where ε_t depend on $\hat{\alpha}$ in order to obtain a minimum conditional sum of squares.

Table 3.1: The $AR(1)$ Model Fitting to Data, $(1-0.428863)(x_t - 1.78866667) = \varepsilon_t$

t	x_t	$z_t = x_t - \mu$	ε_t	ε_t^2
1	1.2	-0.58867	0	0
2	1.5	-0.28867	-0.03621	0.001311
3	1.54	-0.24867	-0.12487	0.015592
4	2.7	0.911333	1.017977	1.036278
5	1.95	0.161333	-0.2295	0.052672
⋮	⋮	⋮	⋮	⋮
45	1.84	0.051333	0.05505	0.003031

where

$$\hat{\alpha} = 0.428863 \qquad CSS = \sum_{t=1}^{n} \varepsilon_t^2 = 9.251275 \qquad \sigma_e^2 = \frac{CSS}{n-1} = 0.21025625.$$

The fitted model is:

$$(1 - 0.428863B)(x_t - 1.78866667) = \varepsilon_t$$
$$(x_t - 1.78866667) - 0.428863(x_{t-1} - 1.78866667) = \varepsilon_t$$

which gives:

$$x_t = 1.78866667 + 0.428863x_{t-1} - 0.7670929541 + \varepsilon_t . \qquad (3.8)$$

Let $x_0 = 1.2$ and ζ_t are independent uncorrelated normal $N(0,1)$ random variables such that $\sigma_e^2 = 0.21025625$ and $\varepsilon_t \sim N(0, 0.21025625)$ then the fitted equation (3.8) is plotted against the actual data:

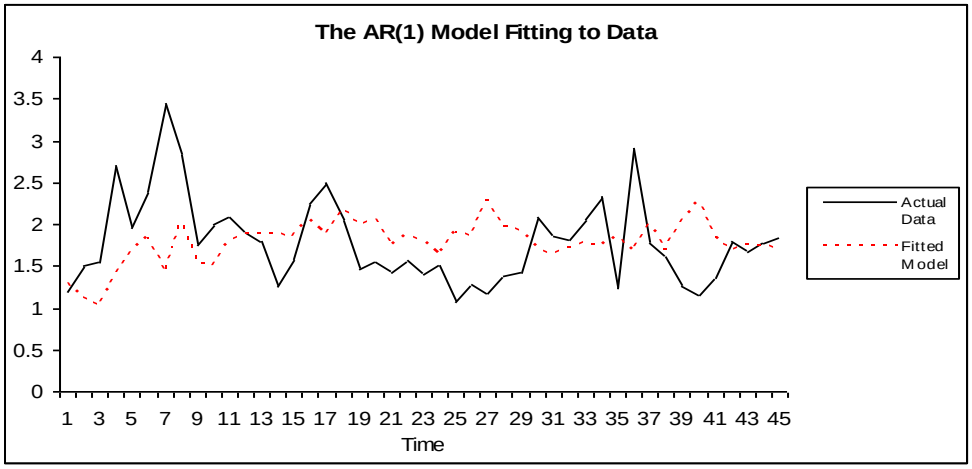

Figure 3.3: The $AR(1)$ Model Fitting to Data, $(1-0.428863)(x_t - 1.78866667) = \varepsilon_t$

3.4 Testing for Mean Reversion and Fat Tails

The QQ-plot for the simulated model according to equation (3.8) indicated in Figure 3.3 is plotted in Figure 3.4.

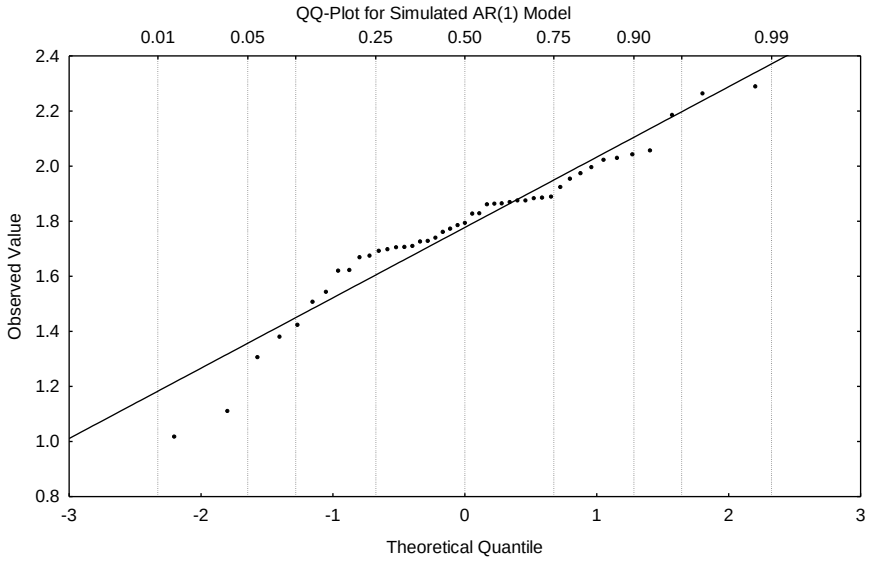

Figure 3.4: QQ-plot for the Simulated *AR(1)* Model for the Truck Manufacturing Defects Data

According to Figure 3.4 indecisive tails are present in the simulated dataset indicated by the QQ-plot. The QQ-plot on the original dataset also indicates indecisive tails are present. Testing the mean reversion of the simulated model involve the Ljung-Box-Pierce-test.

3.5 Notes and Research

By investigating the original dataset and apply the Ljung-Box-Pierce test on this dataset one observed that mean reversion is present. Indecisive tails are observed from the QQ-plot on the original dataset. Based on the these risk properties one may choose the *AR(p)* simulation model.

In order to determine the order of p for the *AR(p)* simulation model one may use paragraph 3.2 and conclude that this is a *AR(1)* simulation model. This conclusion is made due to the fact that the estimated PACF's cut off after lag 1 and the estimated ACF's produce damped sine waves.

The Ljung-Box-Pierce-test for the simulated *AR(1)* model is $Q = 22.42 > \chi^2_{11,0.05} = 19.67$ indicates H_0 will be rejected, meaning the simulated *AR(1)* model does not follow a white noise process and significant autocorrelations are present thus mean reversion is not present (stationarity not present). From these results one can conclude that the simulated *AR(1)* will not be the best model to use for this particular dataset. This could have been the result of the indecisive tails observed for the original dataset and the simulated model.

The simulated *AR(1)* should have mean reversion properties that are the same as the risk properties for the original dataset. Rather choose either One-Factor Logarithmic Vasicek or CIR simulation models where mean reversion and fat tails are present.

Chapter 4

4 MEAN REVERSION – THE ONE-FACTOR VASICEK MODEL ON LOG SPOT PRICES

For electricity price modelling some academics, like M. Davidson, C. Markus and B. Anderson are strong proponents for "hybrid" models. Other academics such as A. Lavassani, A. Sadeghi and A. Ware prefer to use the single-factor and multi-factor models. The main purposes for this chapter are:

- finding the risk properties for the original dataset (the historical model involve electricity prices)
- choosing the simulated model which should have the same risk properties as the original dataset. Table 1.1 is used to find the most appropriate simulated model with the same risk properties as the original dataset
- finding different methods of calculating the estimators for this simulated model. These methods used to find the estimators are called OLS estimation and two different MLE methods
- comparing the accuracy of the OLS estimation and the two MLE by observing whether the model have fat or normal tails. This is done by using the QQ-plots
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

4.1 Basic Properties - One-Factor Vasicek Model Based On Log Spot Price

An initial one-factor Vasicek model can be defined by:

$$dP_t = \alpha(\mu - \ln P_t)P_t dt + \sigma P_t dW_t . \tag{4.1}$$

We call α the mean reversion rate, μ the related long-term mean level of the natural logarithm of the Electricity spot prices, σ the related volatility in the electricity spot prices and W_t the standard Brownian/Wiener process.

A natural logarithm transformation on the spot price, the log-price process $\ln P_t$ will be:

$$x_t = \ln P_t \tag{4.2}$$

where P_t is the electricity spot prices at time t. According to P. Bjerksund et al. (2008)[10], the generalised one-factor Vasicek process has a state variable defined as:

[10] P Bjerksund *et al* "Gas Storage Valuation: Price Modelling v. Optimization Methods" (2008) *Journal of Enconometric Literature* http://www.nhh.no/Admin/Public/Download.aspx?file=/Files/Filler/institutter/for/dp/2008/2008.pdf (Accessed 09 September 2009).

$$x_t = e^{-\alpha t}\left(x_0 + \frac{\sigma^2}{2\alpha}\right) - \frac{\sigma^2}{2\alpha} + \alpha e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u}\mu du + \sigma e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u} dW_u \qquad (4.3)$$

According to P. Bjerksund et al. (2008)[10], the conditional expectation for the natural logarithm of the spot price x_t, given that time $t = 0$ is defined as:

$$E_0[x_t] = e^{-\alpha t}\left(x_0 + \frac{\sigma^2}{2\alpha}\right) - \frac{\sigma^2}{2\alpha} + \alpha e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u}\mu du \qquad (4.4)$$

and the conditional variance of the spot price x_t, given that time $t = 0$ is defined as:

$$\begin{aligned} Var_0[x_t] &= Var_0\left(\sigma e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u} dW_u\right) \\ &= \sigma^2 e^{-2\alpha t}\int_{u=0}^{t} e^{2\alpha u} du \\ &= \sigma^2 e^{-2\alpha t}\left[\frac{1}{2\alpha} e^{2\alpha u}\right]_{u=0}^{t} \\ &= \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right). \end{aligned} \qquad (4.5)$$

According to P. Bjerksund et al. (2008)[10] the conditional expectation for the spot price x_t, given that the time is s is defined as:

$$E_s[x_t] = e^{-\alpha(t-s)}\left[x_s + \frac{\sigma^2}{2\alpha}\right] - \frac{\sigma^2}{2\alpha} + \alpha e^{-\alpha t}\int_{u=s}^{t} e^{\alpha u}\mu du \qquad (4.6)$$

and the conditional variance for the spot price x_t, given that the time is s is defined as:

$$\begin{aligned} Var_s[x_t] &= Var_s\left(\sigma e^{-\alpha t}\int_{u=s}^{t} e^{\alpha u} dW_u\right) \\ &= \sigma^2 e^{-2\alpha t}\int_{u=0}^{t} e^{2\alpha u} du \\ &= \sigma^2 e^{-2\alpha t}\left[\frac{1}{2\alpha} e^{2\alpha u}\right]_{u=s}^{t} = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha(t-s)}\right). \end{aligned} \qquad (4.7)$$

[10] P Bjerksund *et al* "Gas Storage Valuation: Price Modelling v. Optimization Methods" (2008) *Journal of Enconometric Literature* http://www.nhh.no/Admin/Public/Download.aspx?file=/Files/Filler/institutter/for/dp/2008/2008.pdf (Accessed 09 September 2009).

4.2 Explicit Solution - One-Factor Vasicek Model on Log Spot Price

According to S.M. Sheikh (2007)[11] we need an initial model for the stochastic differential equation (SDE). As mentioned in the previous section, let:

$$x_t = \ln P_t \tag{4.8}$$

$$dx_t = d(\ln P_t). \tag{4.9}$$

If we use Ito's Lemma as mentioned in J.C. Hull, (2000)[12], such that if one has a variable P_t it followings an Ito process:

$$dP_t = a(P_t, t)dt + b(P_t, t)dW_t \tag{4.10}$$

where dW_t is a Wiener process. Let a and b be functions of P_t and t. When we have a function, x_t, of P_t and t such as equation (4.9). We can subsequently express equation (4.9) as:

$$dx_t = \left(a\frac{\partial x_t}{\partial P_t} + \frac{\partial x_t}{\partial t} + \frac{1}{2}b^2\frac{\partial^2 x_t}{\partial P_t^2} \right)dt + b\frac{\partial x_t}{\partial P_t}dW_t \tag{4.11}$$

If one compares equation (4.1) with equation (4.10) it is observed that:

$$a(P_t, t) = \alpha(\mu - \ln P_t)P_t,\ b(P_t, t) = \sigma P_t, \tag{4.12}$$

and by using (4.8) and (4.11) we obtain:

$$\frac{\partial x_t}{\partial P_t} = \frac{1}{P_t},\ \frac{\partial x_t}{\partial t} = 0,\ \frac{\partial^2 x_t}{\partial P_t^2} = -\frac{1}{P_t^2} \tag{4.13}$$

By substituting (4.12) and (4.13) into (4.11), equation (4.11) can be expressed as:

$$d(\ln P_t) = \left(\frac{1}{P_t}(\alpha(\mu - \ln P_t)P_t) + 0 - \frac{1}{2P_t^2}(\sigma P_t)^2 \right)dt + \frac{1}{P_t}\sigma P_t dW_t$$

which simplify to:

$$dx_t = \alpha(\mu - \ln P_t)dt - \frac{\sigma^2}{2}dt + \sigma dW_t$$

$$dx_t = \alpha\left(\mu - \frac{\sigma^2}{2\alpha} - x_t \right)dt + \sigma dW_t. \tag{4.14}$$

---

[11] SM Sheikh "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed 9 September 2009).

[12] JC Hull *Options Futures & Other Derivatives* 4 ed (2000) 229.

Now consider,

$$U_t = e^{\alpha t} x_t .$$

Appling Ito's Lemma to previous equation gives,

$$dU_t = d\left(e^{\alpha t} x_t\right) = e^{\alpha t} dx_t + \alpha e^{\alpha t} x_t dt + \frac{1}{2}\alpha^2 e^{\alpha t} (dt)^2 . \tag{4.15}$$

We know that $(dt)^2$ equals zero and if we use equation (4.14) and substitute it into (4.15), then:

$$dU_t = d\left(e^{\alpha t} x_t\right) = e^{\alpha t}\left[\alpha\left(\mu - \frac{\sigma^2}{2\alpha} - x_t\right)dt + \sigma dW_t\right] + \alpha e^{\alpha t} x_t dt$$

$$d\left(e^{\alpha t} x_t\right) = e^{\alpha t}\left[\alpha\left(\mu - \frac{\sigma^2}{2\alpha}\right)dt + \sigma dW_t\right] + \alpha e^{\alpha t} x_t dt - \alpha e^{\alpha t} x_t dt$$

$$d\left(e^{\alpha t} x_t\right) = \alpha\left(\mu - \frac{\sigma^2}{2\alpha}\right)e^{\alpha t} dt + \sigma e^{\alpha t} dW_t . \tag{4.16}$$

Integrating equation (4.16) on both sides with regards to time from $t_i \to t_{i+1}$ one obtains:

$$e^{\alpha t_{i+1}} x_{t_{i+1}} - e^{\alpha t_i} x_{t_i} = \alpha\left(\mu - \frac{\sigma^2}{2\alpha}\right)\int_{t_i}^{t_{i+1}} e^{\alpha s} ds + \sigma \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s . \tag{4.17}$$

Use $\Delta t = t_{i+1} - t_i$ and multiply both sides of equation (4.17) with $e^{-\alpha t_{i+1}}$ and reorder then:

$$x_{t_{i+1}} = x_{t_i} e^{-\alpha \Delta t} + \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha \Delta t}\right) + \sigma e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s . \tag{4.18}$$

4.3 Ordinary Least Squares (OLS) Estimation - One-Factor Vasicek Model on Log Spot Price

With reference to S.M. Sheikh (2007)[11], I derive from equation (4.18) a system of linear equations. First we need to convert equation (4.18) into the following structure.

$$Y = mX + c + \varepsilon \tag{4.19}$$

From equation (4.19) we need to minimise the variance of the error, $Var_{\min}(\varepsilon)$. The minimising of the variance of the error term is the second assumption on the error term, $e_t = \varepsilon$, made in paragraph 1.1, page 1 of this thesis.

By minimising the variance of the error one can obtain the estimators $\hat{\sigma}$, $\hat{\mu}$ and $\hat{\alpha}$. Subtracting x_{t_i} on both sides of equation (4.18) yields,

$$x_{t_{i+1}} - x_{t_i} = x_{t_i}\left(e^{-\alpha\Delta t} - 1\right) + \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha\Delta t}\right) + \sigma e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s \,. \tag{4.20}$$

By comparing equation (4.19) with equation (4.20) one will obtain:

$$\begin{aligned} m &= \left(e^{-\alpha\Delta t} - 1\right) \\ c &= \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha\Delta t}\right) \\ \varepsilon &= \sigma e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s \end{aligned} \tag{4.21}$$

Let

$$Y = \begin{pmatrix} x_{t_1} - x_{t_0} \\ x_{t_2} - x_{t_1} \\ \vdots \\ x_{t_{n-1}} - x_{t_{n-2}} \\ x_{t_n} - x_{t_{n-1}} \end{pmatrix}, \quad X = \begin{pmatrix} x_{t_0} \\ x_{t_1} \\ \vdots \\ x_{t_{n-2}} \\ x_{t_{n-1}} \end{pmatrix}$$

such that $Y = x_{t_{i+1}} - x_{t_i}$, is a $[n \times 1]$ matrix, containing the differences of the log spot prices, X is a $[n \times 1]$ matrix, the intercept, c, and the slope of the regression line, m. Let the noise or residual term be denoted as α and let the noise term be a $[n \times 1]$ matrix. As indicated above, the equation for the slope of the regression line is, $m = \left(e^{-\alpha\Delta t} - 1\right)$. By using this equation one will obtain α, which yields:

[11] SM Sheikh "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed 9 September 2009).

$$\alpha = -\frac{\ln(m+1)}{\Delta t}. \tag{4.22}$$

The equation for the intercept of the regression line also indicated above is, $c = \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha\Delta t}\right)$. By using this equation we can find an expression for μ:

$$\mu = \frac{c}{\left(1 - e^{-\alpha\Delta t}\right)} + \frac{\sigma^2}{2\alpha}. \tag{4.23}$$

Notes that the parameters α, μ and σ^2 must be estimated.

4.3.1 Deriving the Estimators In Order to Minimise the Variance of the Errors

In this section we will determine the variance of the error, ε, and then determine the estimaors which will minimise this variance. First I will state some rules of expectation according to J.S. Milton et al (1995)[13] and R.E. Walpole (1990)[14] ed. 3:

- $E[c] = c$ (The expected value of any constant is that constant).
- $E[cX] = cE[X]$ (Constants can be excluded from expectations).
- $E[X + Y] = E[X] + E[Y]$ (The expected value of the sum is equal to the sum of the expected values).

Using equation (4.19), we can write an expression for the error ε as:

$$\varepsilon = Y - mX - c \tag{4.24}$$

The general formula for the variance of the error ε is given as:

$$Var(\varepsilon) = E[\varepsilon^2] - (E[\varepsilon])^2. \tag{4.25}$$

In order to apply the Ordinary Least Squares Regression according to J. A. Rice (1988)[15], one needs to assume $E[\varepsilon] = E[Y - mX - c]$, is zero. The assumption, $E[\varepsilon] = E[e_t] = 0$, is also derived from paragraph 1.1 of page 1 of this thesis. It follows that:

$$\begin{aligned} Var(\varepsilon) &= E[(Y - mX - c)^2] \\ &= E[Y^2 - mXY + m^2X^2 - mXY + m^2X^2 + cmX - cY + mcX + c^2] \\ &= E[(Y - mX)^2 - 2c(Y - mX) + c^2] \\ &= E[Y^2 - 2mXY + m^2X^2 - 2cY + 2mcX + c^2]. \end{aligned}$$

[13] JS Milton and JC Arnold *Introduction to Probability and Statistics: Principles and Applications for Engineers and the Computing Sciences* 3 ed (1995) 53.
[14] RE Walpole *Introduction to Statistics* 3 ed (1990) 145.
[15] JA Rice *Mathematical Statistics and Data Analysis* (1988) 459.

By using the rules of expectation stated above we can say that:

$$Var(\varepsilon) = E[Y^2] - 2mE[XY] + m^2E[X^2] - 2cE[Y] + 2mcE[X] + c^2. \quad (4.26)$$

In order to minimise the variance of the residuals we need to take partial derivatives of the variance with respect to m and c. Subsequently a zero will be equated to these two equations as follows:

$$\frac{\partial Var(\varepsilon)}{\partial m}|_{\hat{m}} = -2E[XY] + 2mE[X^2] + 2cE[X] = 0 \quad (4.27)$$

$$\frac{\partial Var(\varepsilon)}{\partial c}|_{\hat{c}} = -2E[Y] + 2mE[X] + 2c = 0 \quad (4.28)$$

Observe that if the second equation (4.28) is simplified we obtain:

$$E[Y] - mE[X] - c = 0$$
$$E[Y - mX - c] = E[\varepsilon] = 0.$$

This proves the assumption that the expectation of the error, $E[\varepsilon]$, is zero. The first equation (4.27) needs to be simplified:

$$mE[X^2] + cE[X] = E[XY]. \quad (4.29)$$

Subsequently equation (4.28) must be simplified and multiplied with $E[X]$ which give:

$$mE[X]E[X] + cE[X] = E[X]E[Y] \quad (4.30)$$

Subtracting (4.30) from (4.29) yield:

$$m[\underbrace{E(X^2) - (E(X))^2}_{Variance(X)}] = \underbrace{E[XY] - E[X]E[Y]}_{Covariance(X,Y)}.$$

Divide both sides of the previous equation with $Var(X)$ then the optimal value of *m* becomes $\hat{m}$:

$$\hat{m} = \frac{Cov(X,Y)}{Var(X)}. \quad (4.31)$$

Substitute equation (4.31) into (4.28) one will obtain the optimal value *c* indicated as $\hat{c}$ which yields:

$$\hat{c} = E[Y] - \frac{Cov(X,Y)E[X]}{Var(X)}. \quad (4.32)$$

The minimum estimators for parameters m and c will be respectively $\hat{m}$ and $\hat{c}$.

4.3.2 Calculation for the Minimum Variance of the Errors

With reference to S.M. Sheikh (2007)[11] the optimal intercept, c, and the slope, m, of the regression line derived in equations (4.31) and (4.32) are substituted into equation (4.26) to minimise the variance of the residuals, ε. First let me define some abbreviations:

$$Var(X) = \sigma_X^2$$
$$Var(Y) = \sigma_Y^2$$
$$Cov(X,Y) = \sigma_{XY}.$$

The minimum variance will be:

$$Var_{\min}(\varepsilon) = E(Y^2) - 2\frac{\sigma_{XY}}{\sigma_X^2}E(XY) + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 E(X^2) - 2E(Y) - \left(\frac{\sigma_{XY}E(X)}{\sigma_X^2}\right)E(Y) +$$
$$2\frac{\sigma_{XY}}{\sigma_X^2}\left[E[Y] - \frac{\sigma_{XY}E(X)}{\sigma_X^2}\right]E(X) + \left[E(Y) - \frac{\sigma_{XY}E(X)}{\sigma_X^2}\right]^2$$
$$= E(Y^2) - 2[E(Y)]^2 + [E(Y)]^2 - \underbrace{2\frac{\sigma_{XY}}{\sigma_X^2}E(XY)}_{A} + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 E(X^2) + \underbrace{2E(X)E(Y)\frac{\sigma_{XY}}{\sigma_X^2}}_{A}$$
$$\underbrace{+2E(X)E(Y)\frac{\sigma_{XY}}{\sigma_X^2}}_{B} - 2\left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 E(X)^2 - \underbrace{2E(X)E(Y)\frac{\sigma_{XY}}{\sigma_X^2}}_{B} + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 E(X)^2.$$

From the last expression one needs to merge the marked symbols A and merge the marked symbols B to obtain:

$$Var_{\min}(\varepsilon) = \underbrace{E(Y^2) - [E(Y)]^2}_{\sigma_Y^2} - 2\frac{\sigma_{XY}}{\sigma_X^2}\underbrace{[E(XY) - E(X)E(Y)]}_{\sigma_{XY}^2} + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 \underbrace{[E(X^2) - [E(X)]^2]}_{\sigma_X^2}.$$

The minimum variance will become:

$$Var_{\min}(\varepsilon) = \sigma_Y^2 - \left(\frac{\sigma_{XY}}{\sigma_X}\right)^2. \qquad (4.33)$$

The minimum variance of the errors equation (4.33) must be equal to the variance of the residuals derived in equation (4.21) which yield:

$$Var(\varepsilon) = E(\varepsilon^2) = E\left(\sigma e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s\right)^2 = \sigma_Y^2 - \left(\frac{\sigma_{XY}}{\sigma_X}\right)^2. \qquad (4.34)$$

[11] SM Sheikh "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed 9 September 2009).

According to B. Øksendal (2000)[16] the integral needs to be evaluated by using the following properties.

4.3.3 Properties of Ito Calculus

Assume that $E[W_t] = 0$, W_t is stationary for all t, for $t_1 \neq t_2$, W_{t_1} and W_{t_2} are independent Wiener processes then let $f \in V(S,T)$ such that the Ito integral is defined by:

$$I[f](\omega) = \int_S^T f(t,\omega) dW_t .$$

Let function $f \in V(S,T)$ be approximated by $\phi \in V(S,T)$, and ϕ is known as the elementary function:

$$\phi(t,\omega) = \sum_j e_j(\omega) \chi_{[t_i, t_{j+1})}(t).$$

Let χ be denoted as the characteristic (indicator) function and e_j must be F_{t_j}-measurable in such a way that F_{t_j} be the history of W_S up to time t. The integral of the elementary function ϕ is:

$$\int_S^T \phi(t,\omega) dW_t = \sum_{j \geq 0} e_j(\omega) \left[W_{t_{j+1}} - W_{t_j} \right].$$

Now we can define the following Ito isometry.

[16] B Øksendal, *Stochastic Differential Equation: An Introduction with Applications,* 5 ed. (2000) 32.

Theorem 1. If $\phi(t,\omega)$ is bounded such that $E\left[\int_S^T (\phi-\phi_n)^2\,dt\right]\to 0$, $\phi_n \in V(S,T)$ and $\phi(t,\omega)$ is an elementary function then

$$E\left[\left(\int_S^T \phi(t,w)dW_t\right)^2\right] = E\left[\int_S^T \phi^2(t,w)dt\right]$$

$$E\left[\left(\int_S^T \phi(t,w)dW_t\right)\right] = 0$$

Proof. Let $\Delta W_j = W_{t_{j+1}} - W_{t_j}$. Then

$$E\left[e_i e_j \Delta W_i \Delta W_j\right] = \begin{cases} 0 & if \quad i \neq j \\ E\left[e_j^2\right](t_{j+1}-t_j) & if \quad i = j \end{cases}$$

And we know that ΔW_i and ΔW_j are independent if $i < j$. It follows that

$$E\left[\left(\int_S^T \phi(t,w)dW_t\right)^2\right] = E\left[\int_S^T \phi^2(t,w)dt\right]$$

and

$$E[e_j \Delta W_j] = E[e_j]E[\Delta W_j] = 0, \text{ then}$$

$$E\left[\left(\int_S^T \phi(t,w)dW_t\right)\right] = 0.$$

□

Based on these conditions we can say that since $E\left[\int_S^T (\phi-\phi_n)^2\,dt\right]\to 0$ if follows that $E\left[\int_S^T (\phi-\phi_n)^2\,dt\right]\to 0$ or $\left[\int_S^T E(\phi)^2\,dt < \infty\right]$. It follows that from equation (4.34):

$$\int_0^t E\left(e^{\alpha t}\right)^2 ds = \frac{1}{2\alpha}\left[e^{2\alpha t} - 1\right] < \infty.$$

The solution of equation (4.34) by using Ito's isometry will yield:

$$\begin{aligned} Var(\varepsilon) &= E\left(\sigma e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s \right)^2 \\ &= \sigma^2 e^{-2\alpha t_{i+1}} E\left(\int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s \right)^2 \\ &= \sigma^2 e^{-2\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} e^{2\alpha s} ds \\ &= \sigma^2 e^{-2\alpha t_{i+1}} \frac{1}{2\alpha} \left[e^{2\alpha s} \right]_{s=t_i}^{s=t_{i+1}} \\ &= \sigma^2 e^{-2\alpha t_{i+1}} \frac{1}{2\alpha} \left[e^{2\alpha t_{i+1}} - e^{2\alpha t_i} \right] \\ &= \frac{\sigma^2}{2\alpha} \left(1 - e^{2\alpha \Delta t} \right). \end{aligned} \tag{4.35}$$

Therefore by setting equation (4.35) equal to (4.34) it will enable us to determine the optimal estimator $\hat{\sigma}$ for the parameter σ. Before the optimal estimators $\hat{m}$ and $\hat{c}$ are substituted the parameter σ will yield:

$$\begin{aligned} \sigma_Y^2 - \left(\frac{\sigma_{XY}}{\sigma_X} \right)^2 &= \frac{\sigma^2}{2\alpha} \left(1 - e^{2\alpha \Delta t} \right) \\ \sigma^2 &= \frac{2\alpha}{\left(1 - e^{2\alpha \Delta t} \right)} \left(\sigma_Y^2 - \frac{\left(\sigma_{XY} \right)^2}{\sigma_X^2} \right). \end{aligned} \tag{4.36}$$

By first solving the optimal estimators $\hat{m}$ and $\hat{c}$ by using (4.51) and (4.52) and substituting these estimators into equations (4.22), (4.23) and (4.36) yield the OLS estimates:

$$\begin{aligned} \hat{\alpha} &= -\frac{\ln\left(\hat{m} + 1 \right)}{\Delta t} \\ \hat{\mu} &= \frac{\hat{c}}{\left(1 - e^{-\hat{\alpha} \Delta t} \right)} + \frac{\hat{\sigma}^2}{2\hat{\alpha}} \\ \hat{\sigma}^2 &= \frac{2\hat{\alpha}}{\left(1 - e^{2\hat{\alpha} \Delta t} \right)} \left(\sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \right). \end{aligned}$$

These three OLS estimates above are obtained from applying the first two assumption of the error term in paragraph 1.1 page 1 on this model.

4.4 Maximum Likelihood Estimation (MLE)

With reference to S.M. Sheikh (2007)[11] one may use equation (4.18), the conditional expectation and conditional variance for the logarithm of the electricity spot prices are respectively given as:

$$v_{t_{i+1}} = E\left(x_{t_{i+1}} \mid x_{t_i}\right) \tag{4.37}$$

and

$$w^2_{t_{i+1}} = \sigma^2 e^{-2\alpha t_i} E\left[\int_{t_i}^{t_{i+1}} e^{\alpha s} dW_s\right]^2.$$

Since each x_{t_i} term is known the conditional variance for the logarithm of the electricity spot prices can be further extended with $x_{t_0} \rightarrow x_{t_i}$ such that :

$$\begin{aligned} v_{t_i} &= E\left(x_{t_{i+1}} \mid x_{t_i}\right) \\ &= E\left(x_{t_i} e^{-\alpha t_i} + \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha t_i}\right) + \sigma e^{-\alpha t_i} \int_0^{t_i} e^{\alpha s} dW_s \right) \\ &= E\left[x_{t_i} e^{-\alpha t_i}\right] + E\left[\left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha t_i}\right)\right] + \sigma e^{-\alpha t_i} E\left[\int_0^{t_i} e^{\alpha s} dW_s\right] \\ &= x_{t_0} e^{-\alpha t_i} + \left(\mu - \frac{\sigma^2}{2\alpha}\right)\left(1 - e^{-\alpha t_i}\right) \end{aligned} \tag{4.38}$$

because of the fact that $E\left[\int_0^{t_i} e^{\alpha s} dW_s\right]$=0.

The variance $w^2_{t_i}$ is:

$$\begin{aligned} w^2_{t_i} &= \sigma^2 e^{-2\alpha t_i} E\left[\int_0^{t_i} e^{\alpha s} dW_s\right]^2 \\ &= \sigma^2 e^{-2\alpha t_i} \int_0^{t_i} e^{2\alpha s} ds \\ &= \frac{\sigma^2 e^{-2\alpha t_i}}{2\alpha}\left[e^{2\alpha t_i} - 1\right] \\ &= \frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha t_i}\right]. \end{aligned} \tag{4.39}$$

[11] SM Sheikh "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed 9 September 2009).

We may conclude that the log-prices are $x_{t_i} \sim N\left(v_{t_i}, w_{t_i}^2\right)$ for $i = 1,2,3,...,n$. From chapter 1 the likelihood function becomes:

$$L\left(\alpha, \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi w_i^2}} \exp\left(-\frac{1}{2w_{t_i}^2} \left(x_{t_i} - v_{t_i}\right)^2 \right). \tag{4.40}$$

The natural logarithm of (4.40) is taken and the derivatives in terms of α, μ and σ^2 for which three equations will be obtained and each substituted to zero.

The natural logarithm of equation (4.41) is:

$$\ln L\left(\alpha, \mu, \sigma^2\right) = -\frac{n}{2} \ln 2\pi - \sum_{i=1}^{n} \left(\frac{\ln w_{t_i}^2}{2} + \frac{1}{2w_{t_i}^2} \left(x_{t_i} - v_{t_i}\right)^2 \right). \tag{4.41}$$

One needs to express equation (4.41) in terms of the functions of α, which will make equation (4.41) a one-variable maximisation problem. Taking the derivative of (4.41) in terms of μ and set the derivative to zero yields:

$$\begin{aligned}
\frac{\partial(\ln L)}{\partial \mu} &= 0 \\
-\frac{\partial}{\partial \mu} \left(\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i}\right)^2}{2w_{t_i}^2} \right) &= 0 \\
\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i}\right)}{w_{t_i}^2} \frac{\partial v_{t_i}}{\partial \mu} &= 0 \\
\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i}\right)}{w_{t_i}^2} \left(1 - e^{-\alpha t_i}\right) &= 0 .
\end{aligned} \tag{4.42}$$

Substituting equation (4.38) into (4.42) gives:

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{\left(x_{t_i} - x_{t_0} e^{-\alpha t_i} - \left(\mu - \frac{\sigma^2}{2\alpha} \right)\left(1 - e^{-\alpha t_i}\right) \right)}{w_{t_i}^2} \left(1 - e^{-\alpha t_i}\right) = 0 \\
&\sum_{i=1}^{n} \frac{\left(x_{t_i} - x_{t_0} e^{-\alpha t_i}\right)\left(1 - e^{-\alpha t_i}\right)}{w_{t_i}^2} - \left(\mu - \frac{\sigma^2}{2\alpha} \right) \sum_{i=1}^{n} \frac{\left(1 - e^{-\alpha t_i}\right)^2}{w_{t_i}^2} = 0 .
\end{aligned}$$

Subsequently we may assign the first function in terms of an optimal α assigned as $\hat{\alpha}$ which yields:

$$f(\hat{\alpha}) = \left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}} \right) = \sum_{i=1}^{n} \frac{\left(x_{t_i} - x_{t_0} e^{-\hat{\alpha} t_i} \right)}{\left(1 - e^{-\hat{\alpha} t_i} \right)}. \tag{4.43}$$

where the optimal values of α, μ and σ assigned $\hat{\alpha}$, $\hat{\mu}$ and $\hat{\sigma}$ (also known as the MLE's) are used to solve the last equation on page 37. The derivative of the natural logarithm of the likelihood function, $\frac{\partial L}{\partial \alpha}$, is complicated to derive, subsequently by looking at $\frac{\partial L}{\partial \sigma^2} = 0$, all the other functions in terms of α can be determined. First let:

$$\frac{\partial L}{\partial \sigma^2} = 0$$

$$-\frac{1}{2}\sum_{i=1}^{n} \frac{\partial \ln\left(w_{t_i}^2\right)}{\partial \sigma^2} - \frac{1}{2}\sum_{i=1}^{n} \frac{1}{w_{t_i}^2} \frac{\partial \left(x_{t_i} - v_{t_i} \right)^2}{\partial \sigma^2} - \frac{1}{2}\sum_{i=1}^{n} \left(x_{t_i} - v_{t_i} \right)^2 \frac{\partial w_{t_i}^{-2}}{\partial \sigma^2} = 0$$

$$-\frac{n}{2\sigma^2} - \frac{1}{2}\sum_{i=1}^{n} \frac{1}{w_{t_i}^2} 2\left(x_{t_i} - v_{t_i} \right) \frac{1}{2\alpha} \left(1 - e^{-\alpha t_i} \right) + \frac{1}{2}\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i} \right)^2 \frac{1}{2\alpha} \left(1 - e^{-2\alpha t_i} \right)}{w_{t_i}^4} = 0. \tag{4.44}$$

If equation (4.39) is substituted into (4.44) and if σ^2 is multiplied throughout, then:

$$-\frac{n}{2} - \frac{\sigma^2}{2}\sum_{i=1}^{n} \frac{1}{\left(\frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha t_i} \right] \right)} 2\left(x_{t_i} - v_{t_i} \right) \frac{1}{2\alpha} \left(1 - e^{-\alpha t_i} \right) + \frac{\sigma^2}{2}\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i} \right)^2 \frac{1}{2\alpha} \left(1 - e^{-2\alpha t_i} \right)}{\left(\frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha t_i} \right] \right)^2} = 0$$

$$-\frac{n}{2} - \sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i} \right)\left(1 - e^{-\alpha t_i} \right)}{\left(1 - e^{-2\alpha t_i} \right)} + \frac{1}{2}\sum_{i=1}^{n} \frac{\left(x_{t_i} - v_{t_i} \right)^2}{\left(\frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha t_i} \right] \right)} = 0.$$

Subsequently we may assign the second function in terms of an optimal α assigned as $\hat{\alpha}$ which yields:

$$g(\hat{\alpha}) = \frac{\hat{\sigma}^2}{2\hat{\alpha}} = \frac{\sum_{i=1}^{n} \frac{\left(x_{t_i} - \hat{v}_{t_i} \right)^2}{\left(1 - e^{-2\hat{\alpha} t_i} \right)}}{2\left(n + \sum_{i=1}^{n} \frac{\left(x_{t_i} - \hat{v}_{t_i} \right)}{\left(1 + e^{-\hat{\alpha} t_i} \right)} \right)}. \tag{4.45}$$

where the optimal values of α, μ and σ assigned $\hat{\alpha}$, $\hat{\mu}$ and $\hat{\sigma}$ (also known as the MLE's) are used to solve the equation above equation (4.45). After substituting $\hat{\alpha}$, $\hat{\mu}$

and $\hat{\sigma}$ into equation (4.38) and substitute equation (4.41) into the changed equation (4.38) one may say:

$$\hat{v}_{t_i} = x_{t_0} e^{-\hat{\alpha}t_i} + f(\hat{\alpha})\left(1 - e^{-\hat{\alpha}t_i}\right) = h_i(\hat{\alpha}). \tag{4.46}$$

Equation (4.45) can now be expressed by:

$$g(\hat{\alpha}) = \frac{\hat{\sigma}^2}{2\hat{\alpha}} = \frac{\sum_{i=1}^{n} \frac{\left(x_{t_i} - h_i(\hat{\alpha})\right)^2}{\left(1 - e^{-2\alpha t_i}\right)}}{2\left(n + \sum_{i=1}^{n} \frac{\left(x_{t_i} - h_i(\hat{\alpha})\right)}{\left(1 + e^{-\hat{\alpha}t_i}\right)}\right)}.$$

After substituting $\hat{\alpha}$, $\hat{\mu}$ and $\hat{\sigma}$ into equation (4.39) one will obtain:

$$\hat{w}_{t_i}^2 = g(\hat{\alpha})[1 - e^{-2\hat{\alpha}t_i}]. \tag{4.47}$$

If we substitute equation (4.46) and (4.47) into (4.41) then (4.41) becomes:

$$\ln L(\hat{\alpha}, \hat{\mu}, \hat{\sigma}^2) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n}\left(\ln\left(g(\hat{\alpha})[1 - e^{-2\hat{\alpha}t_i}]\right)\right)^2$$
$$-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(x_{t_i} - x_{t_0}e^{-\hat{\alpha}t_i} - f(\hat{\alpha})\left(1 - e^{-\hat{\alpha}t_i}\right)\right)^2}{\left(g(\hat{\alpha})[1 - e^{-2\hat{\alpha}t_i}]\right)}. \tag{4.48}$$

4.5 Simulation of the One-Factor Vasicek Model on the Log Spot Price

As indicated in equation (4.8) let:

$$x_{t_i} = \ln P_{t_i}.$$

where P_{t_i} is the electricity spot price at time t_i. The stochastic differential equation for x_{t_i} is defined in equation (4.14) as:

$$dx_{t_i} = \hat{\alpha}\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}} - x_{t_i}\right)dt + \hat{\sigma}dW_t.$$

With reference to A. Lari-Lavassani *et al* (2001)[17] the numerical simulation for above equation can be expressed by:

$$x_{t_{i+1}} - x_{t_i} = \hat{\alpha}\Delta t\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}} - x_{t_i}\right) + \hat{\sigma}\sqrt{\Delta t}Z_i$$

such that:

$$x_{t_{i+1}} = x_{t_i} + \hat{\alpha}\Delta t\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}} - x_{t_i}\right) + \hat{\sigma}\sqrt{\Delta t}Z_i$$

$$x_{t_{i+1}} = (1 - \hat{\alpha}\Delta t)x_{t_i} + \hat{\alpha}\Delta t\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}}\right) + \hat{\sigma}\sqrt{\Delta t}Z_i \qquad (4.50)$$

where $\Delta t = t_{i+1} - t_i$ and Z_i are independent identically normally distributed random variable with a mean of zero and a variance of one. The actual spot prices P_{t_i} will be obtained by taking the exponent of x_{t_i} such that:

$$P_{t_i} = e^{x_{t_i}}.$$

4.6 Example 4.1 - Estimators for Electricity Spot Prices

Sweden's Electrical spot prices are given in terms of Euro per Megawatt from 1st of January 2008 to 31st of December 2008 in Appendix A3. The first risk property of this original data is the mean reversion. The Ljung-Box-Pierce test on this dataset is:

$$Q = 5062.834 > \chi^2_{90,0.05} = 113.1453$$

The corresponding $p-value$ is 2.2e-16 which is less than 0.05. The null hypothesis (H_0) will be rejected indicating that the distribution of the original dataset is not mean reverting.

[17] A Lari-Lavassani *et al* "Mean reverting models for energy option pricing" (2001) *University of Calgary*.

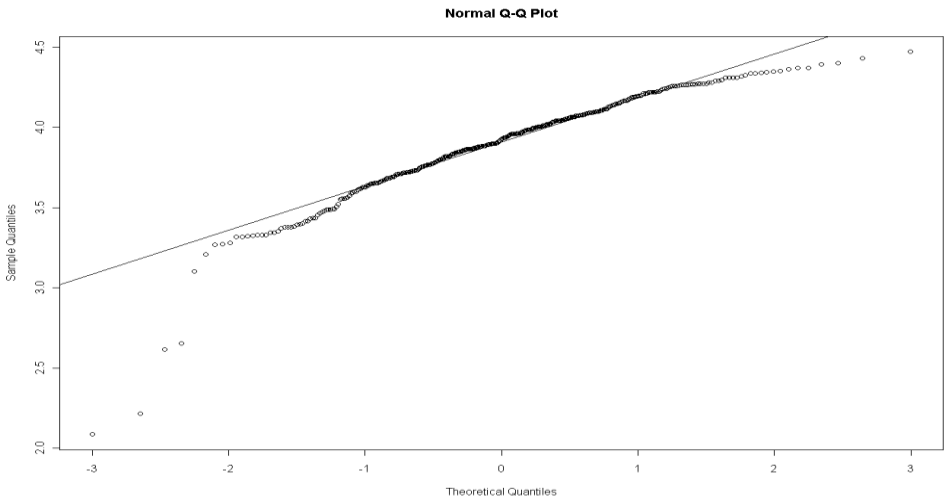

Figure 4.1: QQ-plot on the Logarithmic Sweden's Daily Electricity Prices Indicates

One can see that according to Figure 4.1 the data in the QQ-plot do not follow the straight line which indicates fat tails are present in the original dataset. One may now choose an appropriate simulated model based on Table 1.1. Due to the fact that the Logarithmic Sweden's Daily Electricity Prices Indicates have no mean reversion and fat tail risk properties one may choose the more appropriate One-Factor Logarithmic Vasicek simulation model with jumps to fit on this original dataset. I however choose the One-Factor Logarithmic Vasicek simulation model which is an inappropriate model based on the risk properties of Table 1.1. Equation (4.50) is used to simulate the one-factor Vasicek model on Sweden's electrical log spot prices. Assume that the initial spot price is $P_{t_0} = 41.41$ such that $x_{t_0} = 3.815732$. Let $\Delta t = 1/366$. The vector Y is the difference of the log spot prices and (4.31) and (4.32) will be used to determine the ordinary least square estimates $\hat{m}$ and $\hat{c}$ respectively.

Table 4.1: Ordinary Least Squares Estimation on Sweden's Dataset

Date	i	t	x_{t_i}	$y_{t_i} = x_{t_{i+1}} - x_{t_i}$
	0	0	3.815732	
1/1/2008	1	0.0027322	4.00296	0.187228
1/2/2008	2	0.0054645	3.960242	-0.04272
1/3/2008	3	0.0081967	3.870576	-0.08967
1/4/2008	4	0.0109290	3.850573	-0.02
1/5/2008	5	0.0136612	3.863673	0.0131
1/6/2008	6	0.0163934	4.046554	0.182881
1/7/2008	7	0.0191257	3.894266	-0.15229
⋮	⋮	⋮	⋮	⋮
12/30/2008	365	0.9972678	3.773910	-0.01434
12/31/2008	366	1.0000000	3.759571	-0.02572

By using Table 4.1 the first estimator for this model is:

$$\hat{m} = \frac{Cov(X,Y)}{Var(X)} = \frac{-0.0161}{0.0985} = -0.16343. \tag{4.51}$$

By using equation (4.32) one can solve $\hat{c}$ which yields the second estimator:

$$\hat{c} = E[Y] - \frac{Cov(X,Y)E[X]}{Var(X)} = -0.0002 - \frac{(0.0161)(3.891363)}{0.0985} = 0.635742. \tag{4.52}$$

Then equations on bottom of page 35 are used to yield the following estimators for this model:

$$\hat{\alpha} = -\frac{\ln(\hat{m}+1)}{\Delta t} = 65.31102783$$

$$\hat{\mu} = \frac{\hat{c}}{\left(1-e^{-\hat{\alpha}\Delta t}\right)} + \frac{\hat{\sigma}^2}{2\hat{\alpha}} = 3.98895956$$

$$\hat{\sigma}^2 = \frac{2\hat{\alpha}}{\left(1-e^{2\hat{\alpha}\Delta t}\right)}\left(\sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\right) = 12.92752034.$$

According to Appendix B.4, the negative logarithmic likelihood function given as:

$$\tilde{L}\left(\hat{\alpha},\hat{\mu},\hat{\sigma}^2\right) = \frac{n}{2}\ln 2\pi + \sum_{i=1}^{n}\left(\frac{\ln \hat{w}_{t_i}^2}{2} + \frac{1}{2\hat{w}_{t_i}^2}\left(x_{t_i} - \hat{v}_{t_i}\right)^2\right) \tag{4.53}$$

is minimised using the 'fminsearch' function in Matlab to determine the maximum likelihood estimators. The OLS estimates are used for the initial values of the likelihood function before minimising the likelihood function. The estimates obtained to minimise (4.53) are:

$$\hat{\alpha} = 69.832765\mathbf{51} \quad \hat{\mu} = 3.8364366\mathbf{5} \quad \hat{\sigma}^2 = 0.0000000\mathbf{004023}$$

If one chooses the above estimates as initial values and one calculate the optimal estimates to maximise (4.53) by using the Microsoft Excel solver add-inn functionality, the estimates obtained converge to:

$$\hat{\alpha} = 69.8327655 \quad \hat{\mu} = 3.8364366\mathbf{5} \quad \hat{\sigma}^2 = 0$$

More information on the Microsoft Excel solver add-inn functionality is explained in chapter 5, paragraph 5.9 on page 55. By observing the simulated plot based on these estimates and compared to the OLS, it is clear that $\hat{\sigma}^2 = 0$ is not accurate.

The optimisation of α in terms of maximising the likelihood function (4.53) did not provide an adequate simulated model that represents a similar shape of the original dataset. Alternatively I choose to apply the following two procedures. Firstly I choose $\hat{\alpha}_1 = 69.8327655$ obtained from Matlab and Microsoft Excel solver add-inn when minimising (4.53) and use equations (4.44), (4.47) and (4.46) derived from the MLE procedure to derive the optimal estimators $\hat{\mu}$ and $\hat{\sigma}^2$:

$$f(\hat{\alpha}) = \left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}}\right) = \sum_{i=1}^{n} \frac{\left(x_{t_i} - x_{t_0} e^{-\hat{\alpha} t_i}\right)}{\left(1 - e^{-\hat{\alpha} t_i}\right)} = \frac{1405.93008}{361.243017} = 3.8919232$$

$$g(\hat{\alpha}) = \frac{\hat{\sigma}^2}{2\hat{\alpha}} = \frac{\sum_{i=1}^{n} \frac{\left(x_{t_i} - \hat{v}_{t_i}\right)^2}{\left(1 - e^{-2\hat{\alpha} t_i}\right)}}{2\left(n + \sum_{i=1}^{n} \frac{\left(x_{t_i} - \hat{v}_{t_i}\right)}{\left(1 + e^{-\hat{\alpha} t_i}\right)}\right)} = \frac{\left(\frac{35.89152273}{363.8477}\right)}{2\left(365 + \frac{-3.1E-13}{370.757}\right)} = 0.054204066$$

Solving these equations simultaneously, one will obtain:

$$\hat{\alpha} = 69.8327655 \quad \hat{\mu} = 3.946127262 \quad \hat{\sigma}^2 = 7.570439641$$

Secondly I choose $\hat{\alpha}_2 = 65.31102783$ obtained from the OLE procedure and also used equations (4.44), (4.47) and (4.46) derived from the MLE procedure to derive the optimal estimators $\hat{\mu}$ and $\hat{\sigma}^2$:

$$f(\hat{\alpha}) = \left(\hat{\mu} - \frac{\hat{\sigma}^2}{2\hat{\alpha}}\right) = 3.89199959 \qquad g(\hat{\alpha}) = \frac{\hat{\sigma}^2}{2\hat{\alpha}} = 0.054204571$$

Table 4.2: Summary of One-Factor Logarithmic Vasicek Optimal Estimators

	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\sigma}^2$
OLS	3.98895956	65.3110178	12.92752034
MLE - $\hat{\alpha}_1$	3.94612726	69.8327655	7.57043964
MLE - $\hat{\alpha}_2$	3.94620416	65.3110178	7.080311448

In the following Figure 4.1 the simulated models are: the blue line (MLE - $\hat{\alpha}_1$), the green line as (MLE - $\hat{\alpha}_2$), and the red line (uses the OLS estimators). The black line indicates Sweden's real electricity prices. According to Figure 4.2 the green and blue lines as almost completely the same. In Figure 4.2 the real electricity prices for Sweden in Euro per Megawatt and the simulated One-Factor Vasicek model on the log spot price according to equation (4.50) was transformed to $P_{t_i} = e^{x_{t_i}}$ and simultaneously displayed.

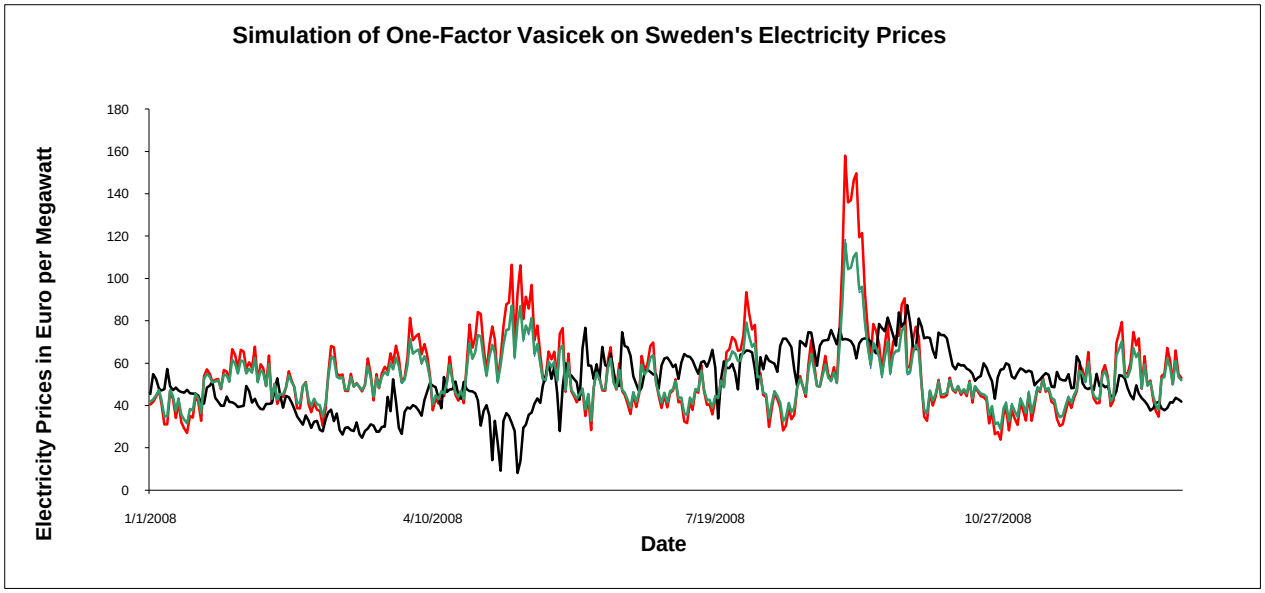

Figure 4.2: Simulated Electricity Prices for Sweden versus the Real Prices

4.7 Testing for Mean Reversion and Fat Tails

Sweden's daily electricity prices in Euro per megawatt are available in Appendix A3. By using the dataset for the simulated green model, the QQ-plot on Sweden's simulated One-Factor Vasicek model indicates the tails of the quantiles are significantly larger than compared to the standard normal quantiles. This factor indicates the presence of fat tails in return of these prices. The result in Figure 4.3, derived from Statistica 8.0, indicates that a Jump process is needed in this model. Due to the topic of this thesis, Jump process will however not be discussed.

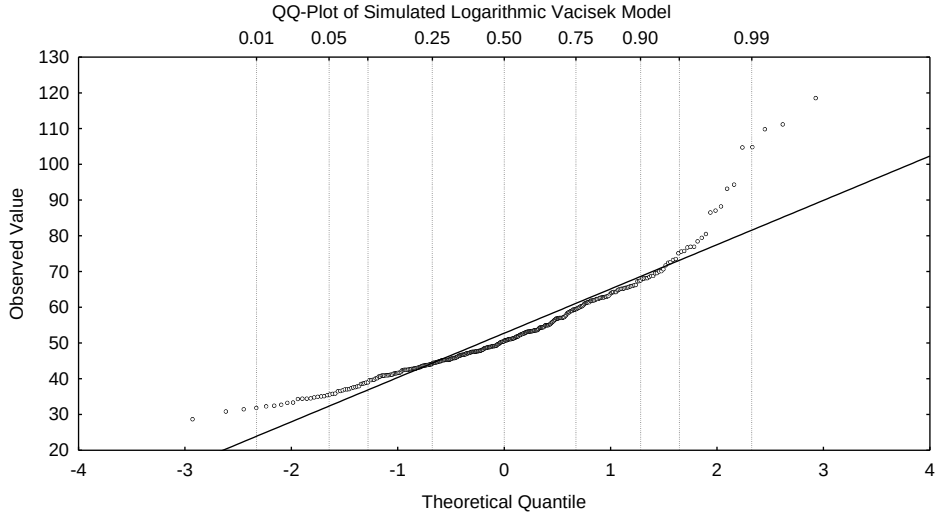

Figure 4.3: QQ-plot for Simulated One-Factor Vasicek Model on Sweden's Daily Electricity Prices Indicates

According to Figure 4.3 fat fails are present in the QQ-plot. The mean reversion property for the simulated One-Factor Vasicek Model according to the Ljung-Box-Pierce-test is

$$Q = 1550.802 > \chi^2_{90,0.05} = 113.1453,$$

indicating H_0 will be rejected, meaning the simulated One-Factor Vasicek model using the estimators obtained from the MLE (using the green model), procedure does not follow a white noise process and significant autocorrelations are present thus mean reversion is not present (no stationarity present).

As observed in Table 4.2, the OLS estimates compared to the two MLE methods for the three estimators are quite different due to the fact that fat tails are present in the QQ-plot indicated in Figure 4.3. Please note that I should have rather fit a one-factor Vasicek model with jumps on the log spot prices for more accuracy. The following paragraph deals with more differences between OLS and MLE for less fat tails.

4.8 Testing for Differences on OLS Estimation and MLE

One may choose specific estimators when generating a dataset from the one-factor Vasicek model on standard normal random variables using equation (4.50). The sufficiency of the OLS and MLE estimators can then be compared to the original parameters. Assume that the initial spot price is $P_{t_0} = 100$ such that $x_{t_0} = 4.60517$. Let $\Delta t = 1/366$. Let the original parameters be $\alpha = 5$, $\mu = 6$ and $\sigma^2 = 3$. A new original dataset is generated from 366 standard normal random values with initial $x_{t_0} = 4.60517$.

Table 4.3: Ordinary Least Squares Estimation on Standard Normal Random Dataset

Date	i	t	x_{t_i}	$y_{t_i} = x_{t_{i+1}} - x_{t_i}$
	0	0	4.60517	
1/1/2008	1	0.0027322	4.557516	-0.04765
⋮	⋮	⋮	⋮	⋮
12/31/2008	366	1.0000000	5.935349	-0.01055

Table 4.4: Summary of One-Factor Logarithmic Vasicek Optimal Estimators

	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\sigma}^2$
OLS	6.372716095	5.419546444	2.744700762
MLE – Solver Add-Inn	6.509520394	3.691574684	1.3819780487

By using Table 4.3 and equations (4.23), (4.24) and (4.36) one will obtain the results for the OLS indicated in Table 4.4. By using the Microsoft Excel solver add-inn functionality, the estimators are given in the second row of Table 4.4. According to Figure 4.4, the tails are not so fat (normal tails) compared to Figure 4.1 and Figure 4.3

and the estimator differences are less for the OLS estimate and the MLE in Table 4.4 as compared to Table 4.2. I have observed that the fatter the tails or equivalently the more outliers present in the original dataset, the larger the divergences for the estimators determined with the OLS estimation compared to the MLE procedure.

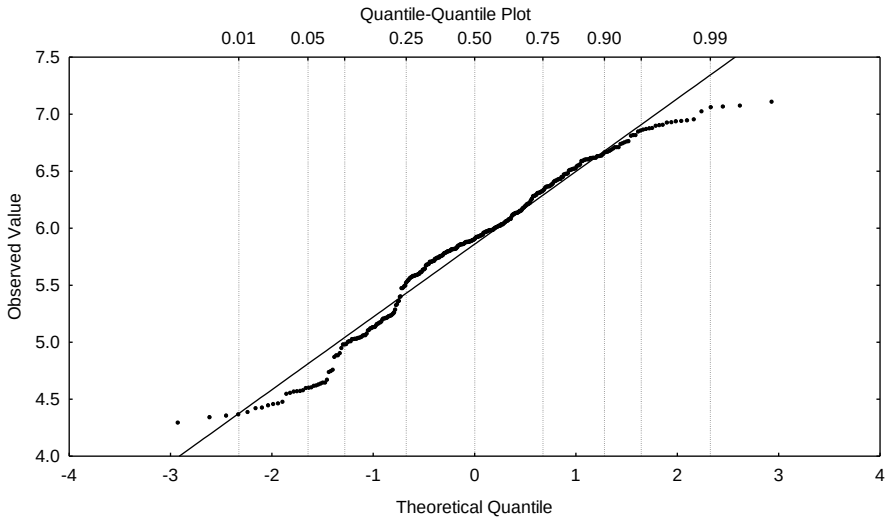

Figure 4.4: QQ-plot for Simulated One-Factor Vasicek Model on Standard Normal Random Variables

4.9 Notes and Research

My aims in this chapter are first to find the risk properties of the historical dataset. One may then choose an appropriate simulation model with the same risk properties as the original dataset. In order to generate a simulation model, the OLS and two methods for the MLE of the Sweden's the log spot price in terms of Euro per Megawatt from 1st of January 2008 to 31st of December 2008 are calculated. After calculating the OLS and the two MLE's I showed via the QQ-plot that fat tails exist in the original dataset and the simulated model.

When fat tails are present it indicates that large differences will occur between the OLS and MLE methods. The OLS and MLE are approximately the same if the tails are not fat, but if the tails get fatter the more inaccurate the OLS procedure become due to the fact that the OLS method is biased. Significant Ljung-Box-Pierce-tests are found in the original dataset and the simulated model such that mean reversion is not present.

Due to these results one should rather use the One-Factor Logarithmic Vasicek simulation model with jumps to fit on this original dataset. I however choose the One-Factor Logarithmic Vasicek simulation model which is an inappropriate model to use based on the risk properties of Table 1.1. The One-Factor Logarithmic Vasicek simulation model with jumps is not discussed in this thesis. I choose the One-Factor Logarithmic Vasicek simulation model for illustration purposes only. As seen in Figure 4.2 both methods of the MLE are approximately the same and give more accurate simulated models than the simulated model derived from the OLS method.

Chapter 5

5 MEAN REVERSION – THE VASICEK MODEL

The main purposes for this chapter are:

- finding the risk properties for the original dataset (the historical model standard normal random data)
- choosing the simulated model which should have the same risk properties as the original dataset. Table 1.1 is used to find the most appropriate simulated model with the same risk properties as the original dataset
- finding different methods of calculating the estimators for this simulated model. These methods used to find the estimators are called OLS estimation and two different MLE methods
- comparing the accuracy of the OLS estimates and the two MLE by observing whether the model have fat or normal tails. This is done by using the QQ-plots.
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

5.1 Basic Properties - Vasicek Model

The Vasicek model, owing its name to Vasicek (1977), is one of the earliest stochastic models of the short-term interest rate. This model is also known as the Gaussian model, the Ornstein-Uhlenbeck mean reverting (OUMR) process or the extended-Vasicek model according to P. Skantze *et al* (2000)[18] which is a suitable model to apply the maximum likelihood estimation (MLE) or the ordinary least squares (OLS) method on. The instantaneous spot rate (or "short rate") used for historical estimation is defined according to J.C.G. Franco [19] as:

$$dx_t = \alpha(\mu - x_t)dt + \sigma dW_t \tag{5.1}$$

where α, μ and x_0 are constants and dW_t represents an increment to a standard Brownian motion W_t. The spot rate x_t will fluctuate randomly but over the long run tends to revert to some level μ. The speed of reversion is known as α and the short-term standard deviation is σ where both influence the reversion.

The non null property of negative rates is a major shortcoming for the Vasicek model. This is an unrealistic property for the modelling of interest rate when using Vasicek

[18] P Skantze *et al* "Stochastic Modeling of Electric Power Prices in a Multi-market Environment" (2000) *IEEE* Cambridge University.
[19] JCG Franco "Maximum Likelihood Estimation of Mean Reverting Processes" (2008) http://www.investmentscience.com/Content/howtoArticles/MLE_for_OR_mean_reverting.pdf (Accessed 09 September 2009).

models. Solving the Ornstein-Uhlenbeck Stochastic Differential Equation includes taking the derivative of $e^{\alpha t} x_t$ which yields:

$$d\left(e^{\alpha t} x_t\right) = x_t \alpha e^{\alpha t} dt + e^{\alpha t} dx_t . \tag{5.2}$$

Rearrange the order of equation (5.2) gives:

$$e^{\alpha t} dx_t = d\left(e^{\alpha t} x_t\right) - x_t \alpha e^{\alpha t} dt . \tag{5.3}$$

Multiply both sides of equation (5.1) with $e^{\alpha t}$ is:

$$e^{\alpha t} dx_t = e^{\alpha t} \alpha\left(\mu - x_t\right) dt + e^{\alpha t} \sigma dW_t . \tag{5.4}$$

By using equation (5.3) and substitute it into equation (5.4), yields,

$$d\left(e^{\alpha t} x_t\right) = \alpha e^{\alpha t} \mu dt + e^{\alpha t} \sigma dW_t . \tag{5.5}$$

If an integral is taken from time $t = 0$ to t gives:

$$e^{\alpha t} x_t = x_0 + \int_{t=0}^{t} \alpha e^{\alpha s} \mu ds + \int_{t=0}^{t} e^{\alpha s} \sigma dW_s . \tag{5.6}$$

Writing equation (5.6) in terms of x_t yields:

$$x_t = x_0 e^{-\alpha t} + \int_{t=0}^{t} \alpha e^{\alpha s} e^{-\alpha t} \mu ds + \int_{t=0}^{t} e^{\alpha s} e^{-\alpha t} \sigma dW_s = x_0 e^{-\alpha t} + \int_{t=0}^{t} \alpha e^{-\alpha(t-s)} \mu ds + \int_{t=0}^{t} e^{-\alpha(t-s)} \sigma dW_s$$

$$= x_0 e^{-\alpha t} + \mu\left(1 - e^{-\alpha t}\right) + \int_{t=0}^{t} e^{-\alpha(t-s)} \sigma dW_s . \tag{5.7}$$

The solution of the stochastic differential equation (5.1) between s and t, if $0 \le s < t$ is:

$$x_t = x_s e^{-\alpha(t-s)} + \mu\left(1 - e^{-\alpha(t-s)}\right) + \sigma e^{-\alpha t} \int_{t=s}^{t} e^{\alpha u} dW_u . \tag{5.8}$$

As indicated in chapter 4, the second integral on the right hand side of equation (5.7) is $\int_{t=0}^{t} e^{-\alpha(t-s)} \sigma dW_s$ which follows a normal distribution with a mean of zero and a variance such that:

$$E\left[\left(\int_{t=0}^{t} e^{-\alpha(t-s)}\sigma dW_s\right)^2\right] = \int_0^t \left(e^{-\alpha(t-s)}\sigma\right)^2 ds$$
$$= \frac{\sigma^2}{2\alpha}\left(1-e^{-2\alpha t}\right).$$

The conditional mean and variance of x_t given x_0 is:

$$E_0[x_t] = \mu + (x_0 - \mu_t)e^{-\alpha t}$$
$$Var_0[x_t] = \frac{\sigma^2}{2\alpha}\left(1-e^{-2\alpha t}\right), \ \alpha > 0. \tag{5.9}$$

The conditional mean and variance of x_t given x_s are:

$$E_s[x_t] = \mu + (x_0 - \mu_t)e^{-\alpha(t-s)}$$
$$Var_s[x_t] = \frac{\sigma^2}{2\alpha}\left(1-e^{-2\alpha(t-s)}\right). \tag{5.10}$$

If time increases the mean tends to the long-term value μ and the variance remains bounded, implying mean reversion. The long-term distribution of the Ornstein-Uhlenbeck process is stationary and is Gaussian with mean μ and variance $\sigma^2/2\alpha$.

5.2 Maximum Likelihood Estimate (Method 1) - Vasicek Model

The discrete time version on the time grid $t_1, t_2, ..., t_n$ with time step $\Delta t = t_i - t_{i-1}$ will be used in this section. If $x_{t_{i-1}}$ is given, the conditional density function f of x_{t_i} by using equations (5.9), yields:

$$f_i\left(x_{t_i};\mu,\alpha,\sigma\right) = \left(2\pi\right)^{-1/2}\left(\frac{\sigma^2}{2\alpha}\left(1-e^{-2\alpha t}\right)\right)^{-1/2} \exp\left[\frac{-\left(x_{t_i} - \mu - \left(x_{t_{i-1}} - \mu\right)e^{-\alpha(t_i - t_{i-1})}\right)^2}{2\left(\frac{\sigma^2}{2\alpha}\right)\left(1-e^{-2\alpha(t_i - t_{i-1})}\right)}\right].$$

The log-likelihood function is given by:

$$\ln L\left(x_t;\mu,\alpha,\sigma\right) = -\frac{n}{2}\ln\left(\frac{\sigma^2}{2\alpha}\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\left(1-e^{-2\alpha(t_i - t_{i-1})}\right)$$
$$= -\frac{n}{\sigma^2}\sum_{i=1}^{n}\frac{\left(x_{t_i} - \mu - \left(x_{t_{i-1}} - \mu\right)e^{-\alpha(t_i - t_{i-1})}\right)^2}{1-e^{-2\alpha(t_i - t_{i-1})}}. \tag{5.11}$$

The log-likelihood function needs to be maximised by taking partial derivatives of equation (5.11) towards μ, α and σ which yield three equations all equal to zero:

$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \mu}|_{\hat{\mu}}=0 \quad \frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \alpha}|_{\hat{\alpha}}=0 \quad \frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \sigma}|_{\hat{\sigma}}=0\,.$$

The aim is to solve these equations simultaneously and write equation (5.11) as a one-dimensional expression.

Looking at the first derivative:

$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \mu}=-\frac{\alpha}{\sigma^2}\sum_{i=1}^{n}\frac{\left(x_{t_i}-\mu-\left(x_{t_{i-1}}-\mu\right)e^{-\alpha(t_i-t_{i-1})}\right)}{1+e^{-\alpha(t_i-t_{i-1})}}$$

which if we assume that $\alpha \neq 0$ and $\sigma \neq 0$ gives:

$$\hat{\mu}=f(\hat{\alpha})=\sum_{i=1}^{n}\frac{\left(x_{t_i}-x_{t_{i-1}}e^{-\hat{\alpha}(t_i-t_{i-1})}\right)}{1+e^{-\hat{\alpha}(t_i-t_{i-1})}}\left(\sum_{i=1}^{n}\frac{\left(1-e^{-\hat{\alpha}(t_i-t_{i-1})}\right)}{\left(1+e^{-\hat{\alpha}(t_i-t_{i-1})}\right)}\right)^{-1}. \tag{5.12}$$

The second derivative gives:

$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \sigma}=-\frac{n}{\sigma}+\frac{2\alpha}{\sigma^3}\sum_{i=1}^{n}\frac{\left(x_{t_i}-\mu-\left(x_{t_{i-1}}-\mu_t\right)e^{-\alpha(t_i-t_{i-1})}\right)^2}{1-e^{-2\alpha(t_i-t_{i-1})}}$$

which simplify to:

$$\hat{\sigma}=g(\hat{\mu},\hat{\alpha})=\sqrt{\frac{2\hat{\alpha}}{n}\sum_{i=1}^{n}\frac{\left(x_{t_i}-\hat{\mu}-\left(x_{t_{i-1}}-\hat{\mu}\right)e^{-\hat{\alpha}(t_i-t_{i-1})}\right)^2}{1-e^{-2\hat{\alpha}(t_i-t_{i-1})}}}\,. \tag{5.13}$$

The log-likelihood function first needs to be expressed by functions of α by using (5.12) and (5.13) and then maximised with respect to α which yields:

$$\hat{V}(\alpha)=-\frac{n}{2}\ln\left[\frac{g(f(\alpha),\alpha)^2}{2\alpha}\right]-\frac{1}{2}\sum_{i=1}^{n}\ln\left(1-e^{-2\alpha(t_i-t_{i-1})}\right)$$
$$-\frac{n}{g(f(\alpha),\alpha)^2}\sum_{i=1}^{n}\frac{\left(x_{t_i}-f(\alpha)-\left(x_{t_{i-1}}-f(\alpha)\right)e^{-\alpha(t_i-t_{i-1})}\right)^2}{1-e^{-2\alpha(t_i-t_{i-1})}}\,. \tag{5.14}$$

According to this method, a likelihood expressed in (5.14) is written in terms of functions of α and then maximised in terms of α. An optimal value $\hat{\alpha}$ derived from maximising (5.14) is substituted into (5.12) and then (5.13) to determine $\hat{\mu}$ and $\hat{\sigma}$.

5.3 Simulation - Vasicek Model

The simulation model for the Ornstein-Uhlenbeck process with time step $\Delta t = t_i - t_{i-1}$ according to M.A. van den Berg (2007) [20] is derived from (5.8) and is given as:

$$x_{t_{i+1}} = x_{t_i} e^{-\hat{\alpha}\Delta t} + \mu\left(1 - e^{-\hat{\alpha}\Delta t}\right) + \hat{\sigma}\sqrt{\frac{1 - e^{-2\hat{\alpha}\Delta t}}{2\hat{\alpha}}} Z_i$$

where Z_i is independent identically distributed and follows a standard normal distribution with a mean of zero and a variance of one.

5.4 Example 5.1 – Generating Original Dataset Using Vasicek Model

Let $\Delta t = 0.25$, the mean reversion rate $\alpha = 7$, the long term mean $\mu = 5.5$ and the volatility $\sigma = 0.15$. Also let the initial spot rate be $x_{t_0} = 15$. Table 5.1 determines the generated values for $x_{t_1}, ...x_{t_{155}}$.

Table 5.1: Generating Original Dataset Using Vasicek Model

i	t_i	Z_i	x_{t_i}
0	0		15
1	0.25	-0.021674	7.144863
2	0.5	0.910547	6.037468
3	0.75	2.316137	6.233474
4	1	-0.537386	5.478949
5	1.25	-0.459722	5.369295
6	1.5	2.956149	6.294233
7	1.75	1.364187	6.015017
8	2	-1.254898	5.242699
9	2.25	-1.747471	4.972366
10	2.5	0.767391	5.620383
⋮	⋮	⋮	⋮
154	38.5	-0.11734	5.45406
155	38.75	-1.93103	4.958368

[20] MA van den Berg "Calibrating the Ornstein-Uhlenbeck model" (2009) http://www.sitmo.com/doc/Calibrating_the_Omstein-Uhlenbeck_model (Accessed 02 June 2009).

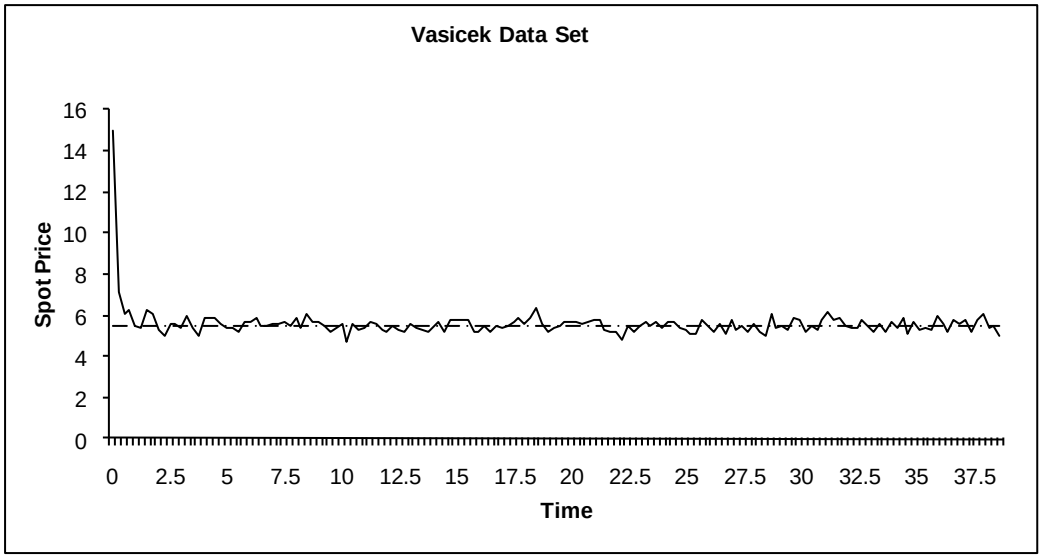

Figure 5.1: Plot for Generated Vasicek Model

Figure 5.1 plots the generated values for $x_{t_1},...x_{t_{155}}$ with $t_1,...,t_{155}$. By observing Figure 5.1 one can clearly see the reversion of the original data that revert to some level $\mu = 5.5$. The risk properties for this dataset $x_{t_1},...x_{t_{155}}$ must be determined. The Ljung-Box-Pierce test on this dataset is:

$$Q = 10.4913 < \chi^2_{38,0.05} = 53.835$$

The corresponding $p-value$ is 1 which is larger than 0.05. The null hypothesis (H_0) will not be rejected indicating that the distribution of the original dataset is mean reverting. The risk property stated above corresponds to Table 1.1 for the Vasicek model.

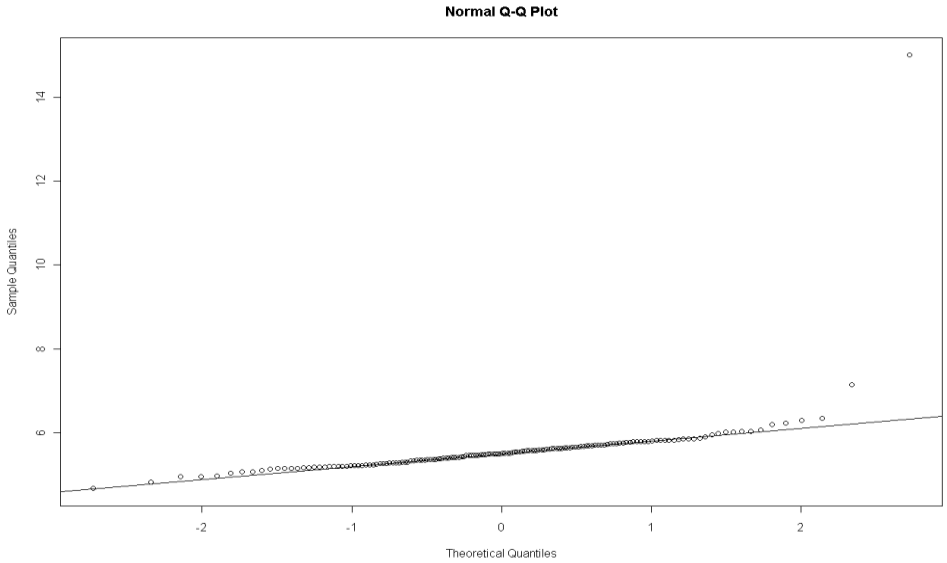

Figure 5.2: QQ-Plot of Generated Vasicek Model

Based on observing Table 5.2, one can see that normal tails are present which also corresponds to Table 1.1 for the simulated Vasicek model. The Vasicek simulated model will thus be an appropriate simulated model to choose based on the similarities between the risk properties of the generated dataset and Table 1.1.

5.5 Ordinary Least Squares Estimation - Vasicek Model

The linear relationship between two consecutive observations $x_{t_{i+1}}$ and x_{t_i} is linear with independent identical random values ε such that:

$$x_{t_{i+1}} = ax_{t_i} + b + \varepsilon$$

where

$$a = e^{-\alpha\Delta t} \qquad b = \mu\left(1 - e^{-\alpha\Delta t}\right) \qquad \varepsilon_{sd} = \sigma\sqrt{\frac{1 - e^{-2\alpha\Delta t}}{2\alpha}} \tag{5.15}$$

Express these equations in terms of the parameters μ, α and σ which yield:

$$\alpha = -\frac{\ln(a)}{\Delta t} \qquad \mu = \frac{b}{1-a} \qquad \sigma = \varepsilon_{sd}\sqrt{\frac{-2\ln(a)}{\Delta t\left(1 - a^2\right)}} \tag{5.16}$$

The following formulas are used to simplify further calculations:

$$S_x = \sum_{i=1}^{n} x_{t_{i-1}} \qquad S_y = \sum_{i=1}^{n} x_{t_i}$$

$$S_{xx} = \sum_{i=1}^{n} x_{t_{i-1}}^2 \qquad S_{yy} = \sum_{i=1}^{n} x_{t_i}^2 \qquad S_{xy} = \sum_{i=1}^{n} x_{t_{i-1}} x_{t_i} \ . \tag{5.17}$$

The estimated coefficients of the parameters *a*, *b* and ε_{sd} of least squares regression line are:

$$\hat{a} = \frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2} \qquad \hat{b} = \frac{S_y - aS_x}{n}$$

$$\hat{\varepsilon}_{sd} = \sqrt{\frac{nS_{yy} - S_y^2 - \hat{a}\left(nS_{xy} - S_x S_y\right)}{n(n-2)}} \ . \tag{5.18}$$

The ordinary least square (OLS) estimates $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\sigma}$ are:

$$\hat{\alpha} = -\frac{\ln\left(\frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\right)}{\Delta t} \qquad \hat{\mu} = \frac{S_y - \left(\frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\right) S_x}{n\left[1 - \left(\frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\right)\right]} \qquad \hat{\sigma} = \hat{\varepsilon}_{sd} \sqrt{\frac{-2\ln\left(\frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\right)}{\Delta t\left(1 - \left(\frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\right)^2\right)}}$$

5.6 Example 5.2 – Ordinary Least Squares Estimation

By using the following table along with equations (5.17) and equations (5.18), the estimators in equations above are calculated.

Table 5.2: Calculations of Equations (5.17)

i	t	Z_i	x_{t_i}	$x_{t_i}^2$	$x_{t_{i-1}} x_{t_i}$
0	0		15	225	
1	0.25	-0.021674	7.144863	51.04906	107.1729
2	0.5	0.910547	6.037468	36.45102	43.13688
3	0.75	2.316137	6.233474	38.85619	37.6344
4	1	-0.537386	5.478949	30.01889	34.15289
5	1.25	-0.459722	5.369295	28.82933	29.4181
⋮	⋮	⋮	⋮	⋮	⋮
154	38.5	-0.11734	5.45406	29.74677	29.57323
155	38.75	-1.93103	4.958368	24.58541	27.04323

From Table 5.2 the solutions for equations (5.17) are:

$$S_x = \sum_{i=1}^{n} x_{t_{i-1}} = 865.6887 \qquad S_y = \sum_{i=1}^{n} x_{t_i} = 855.6471$$

$$S_{xx} = \sum_{i=1}^{n} x_{t_{i-1}}^2 = 4939.401 \qquad S_{yy} = \sum_{i=1}^{n} x_{t_i}^2 = 4738.98603$$

$$S_{xy} = \sum_{i=1}^{n} x_{t_{i-1}} x_{t_i} = 24796.626046$$

Using these results, the solutions of equations (5.18) yield:

$$\hat{a} = \frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2} = 0.17004140\text{5} \qquad \hat{b} = \frac{S_y - \hat{a}S_x}{n} = 4.570607537$$

$$\hat{\varepsilon}_{sd} = \sqrt{\frac{nS_{yy} - S_y^2 - \hat{a}\left(nS_{xy} - S_x S_y\right)}{n(n-2)}} = 0.286218753.$$

Finally the estimators obtained from the OLS method are:

$$\hat{\alpha} = -\frac{\ln(\hat{a})}{\Delta t} = 7.08685326 \qquad \hat{\mu} = \frac{\hat{b}}{1-\hat{a}} = 5.507030788$$

$$\hat{\sigma} = \hat{\varepsilon}_{sd}\sqrt{\frac{-2\ln(\hat{a})}{\Delta t\left(1-\hat{a}^2\right)}} = 1.093480328.$$

5.7 Maximum Likelihood Estimate (Method 2) - Vasicek Model

Alternative calculations for the maximum likelihood estimates are made according to M.A. van den Berg (2007)[21] as follows. The conditional density function for x_{t_i} given $x_{t_{i-1}}$ is:

$$f\left(x_{t_i} \mid x_{t_{i-1}}; \mu, \alpha, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\left(x_{t_i} - x_{t_{i-1}} e^{-\alpha\Delta t} - \mu\left(1 - e^{-\alpha\Delta t}\right)\right)^2}{2\sigma^2}\right]$$

where

$$\sigma^2 = \hat{\sigma}^2 \frac{\left(1 - e^{-2\alpha\Delta t}\right)}{2\alpha}. \qquad (5.19)$$

[21] Van den Berg "Calibrating the Ornstein-Uhlenbeck model".

The log-likelihood function is given by:

$$\ln L(\mu,\alpha,\sigma) = \sum_{i=1}^{n} \ln f\left(x_{t_i} \mid x_{t_{i-1}};\mu,\alpha,\sigma\right)$$
$$= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_{t_i} - x_{t_{i-1}}e^{-\alpha\Delta t} - \mu\left(1-e^{-\alpha\Delta t}\right)\right)^2 . \quad (5.20)$$

The log-likelihood function needs to be maximised by taking partial derivatives of equation (5.20) with respect to μ, α and σ which yield three equations all equal to zero:

$$\frac{\partial \ln L(\mu,\alpha,\sigma)}{\partial \mu}|_{\hat{\mu}} = 0$$
$$\frac{\partial \ln L(\mu,\alpha,\sigma)}{\partial \alpha}|_{\hat{\alpha}} = 0$$
$$\frac{\partial \ln L(\mu,\alpha,\sigma)}{\partial \sigma}|_{\bar{\sigma}} = 0$$

The estimators will be:

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\left(x_{t_i} - x_{t_{i-1}}e^{-\hat{\alpha}\Delta t}\right)}{n\left(1-e^{-\hat{\alpha}\Delta t}\right)} \qquad \hat{\alpha} = -\frac{1}{\Delta t}\ln\left[\frac{\sum_{i=1}^{n}\left(x_{t_i}-\hat{\mu}\right)\left(x_{t_{i-1}}-\hat{\mu}\right)}{\sum_{i=1}^{n}\left(x_{t_{i-1}}-\hat{\mu}\right)^2}\right]$$
$$\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_{t_i} - \hat{\mu} - e^{-\hat{\alpha}\Delta t}\left(x_{t_{i-1}}-\hat{\mu}\right)\right)^2 . \quad (5.21)$$

By using equations (5.17), the MLE's are:

$$\hat{\mu} = \frac{S_y - e^{-\hat{\alpha}\Delta t}S_x}{n\left(1-e^{-\hat{\alpha}\Delta t}\right)} \quad (5.22)$$

$$\hat{\alpha} = -\frac{1}{\Delta t}\ln\left(\frac{\left[S_{xy} - \hat{\mu}S_x - \hat{\mu}S_y + n\hat{\mu}^2\right]}{S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2}\right) \quad (5.23)$$

$$\bar{\sigma}^2 = \frac{1}{n}\left[S_{yy} - 2e^{-\hat{\alpha}\Delta t}S_{xy} + e^{-2\hat{\alpha}\Delta t}S_{xx} - 2\hat{\mu}\left(1-e^{-\hat{\alpha}\Delta t}\right)\left(S_y - e^{-\hat{\alpha}\Delta t}S_x\right) + n\hat{\mu}^2\left(1-e^{-\hat{\alpha}\Delta t}\right)^2\right]. \quad (5.24)$$

If equation (5.23), $\hat{\alpha}$, is substituted into (5.22), $\hat{\mu}$, it yields:

$$
\begin{aligned}
n\hat{\mu} &= \frac{S_y - \left(\frac{S_{xy} - \hat{\mu}S_x - \hat{\mu}S_y + n\hat{\mu}^2}{S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2}\right)S_x}{1 - \left(\frac{S_{xy} - \hat{\mu}S_x - \hat{\mu}S_y + n\hat{\mu}^2}{S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2}\right)} \\
&= \frac{S_y\left(S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2\right) - \left(S_{xy} - \hat{\mu}S_x - \hat{\mu}S_y + n\hat{\mu}^2\right)S_x}{\left(S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2\right) - \left(S_{xy} - \hat{\mu}S_x - \hat{\mu}S_y + n\hat{\mu}^2\right)} \\
&= \frac{S_yS_{xx} - S_xS_{xy} + \hat{\mu}\left(S_x^2 - S_xS_y\right) + n\hat{\mu}^2\left(S_y - S_x\right)}{S_{xx} - S_{xy} + \hat{\mu}\left(S_y - S_x\right)}.
\end{aligned}
$$

The optimal estimate according to the alternative maximum likelihood estimation method is:

$$
\hat{\mu} = \frac{S_yS_{xx} - S_xS_{xy}}{n\left(S_{xx} - S_{xy}\right) - \left(S_x^2 - S_xS_y\right)}. \tag{5.25}
$$

Substitute (5.25), $\hat{\mu}$, into (5.23) gives:

$$
\hat{\alpha} = -\frac{1}{\Delta t}\ln\left[\frac{S_{xy} - \hat{\mu}\left(S_x + S_y\right) + n\hat{\mu}^2}{S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2}\right] \tag{5.26}
$$

and using (5.25) and (5.26), the third estimate $\hat{\sigma}^2$ is:

$$
\hat{\sigma}^2 = \frac{2\hat{\alpha}}{n\left(1 - e^{-2\hat{\alpha}\Delta t}\right)}\left[S_{yy} - 2e^{-\hat{\alpha}\Delta t}S_{xy} + e^{-2\hat{\alpha}\Delta t}S_{xx} - 2\hat{\mu}_t\left(1 - e^{-\hat{\alpha}\Delta t}\right)\left(S_y - e^{-\hat{\alpha}\Delta t}S_x\right) + n\hat{\mu}_t^2\left(1 - e^{-\hat{\alpha}\Delta t}\right)^2\right]
$$

(5.27)

5.8 Example 5.3 –Maximum Likelihood Estimation (Method 2)

By using the results from the previous section, the following results are obtained.

$$
S_x = \sum_{i=1}^{n} x_{t_{i-1}} = 865.688728 \qquad S_y = \sum_{i=1}^{n} x_{t_i} = 855.6470955
$$

$$
S_{xx} = \sum_{i=1}^{n} x_{t_{i-1}}^2 = 4939.400626 \qquad S_{yy} = \sum_{i=1}^{n} x_{t_i}^2 = 4738.98603
$$

$$
S_{xy} = \sum_{i=1}^{n} x_{t_{i-1}}x_{t_i} = 4796.626046.
$$

Substituting these results into equations (5.22), (5.23) and (5.24) give:

$$\hat{\mu} = \frac{S_y S_{xx} - S_x S_{xy}}{n\left(S_{xx} - S_{xy}\right) - \left(S_x^2 - S_x S_y\right)} = 5.507030788$$

$$\hat{\alpha} = -\frac{1}{\Delta t}\ln\left[\frac{S_{xy} - \hat{\mu}\left(S_x + S_y\right) + n\hat{\mu}^2}{S_{xx} - 2\hat{\mu}S_x + n\hat{\mu}^2}\right] = 7.08685326$$

$$\bar{\sigma}^2 = \frac{1}{n}\left[S_{yy} - 2e^{-\hat{\alpha}\Delta t}S_{xy} + e^{-2\hat{\alpha}\Delta t}S_{xx} - 2\hat{\mu}\left(1-e^{-\hat{\alpha}\Delta t}\right)\left(S_y - e^{-\hat{\alpha}\Delta t}S_x\right) + n\hat{\mu}^2\left(1-e^{-\hat{\alpha}\Delta t}\right)^2\right] = 0.080864127$$

so that:

$$\hat{\sigma} = \bar{\sigma}\sqrt{\frac{2\hat{\alpha}}{1-e^{-2\hat{\alpha}\Delta t}}} = 1.086402711.$$

5.9 Maximum Likelihood Estimate (Method 3) – Vasicek Model

Yet another method of determining the estimators is by using Microsoft Excel's solver add-inn. Microsoft Excel Solver uses the Generalized Reduced Gradient (GRG2). [22]

5.10 Example 5.4 –Maximum Likelihood Estimation (Method 3)

In this example Microsoft Excel's solver add-inn is used to maximise equation (5.20). The log-likelihood function given in (5.20) is determined in the last column of Table 5.3 indicated by $\ln L$ for different values of $i = 1,...,155$. The cell for $\ln L$ if $i = 155$ will be maximised by changing the cells selected for the estimators $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\sigma}$.

Table 5.3: Optimisation of Equation (5.20)

i	t	Z_i	x_{t_i}	$\ln L$
0	0		15	
1	0.25	-0.021674	7.14486264	0.335101
2	0.5	0.910547	6.037468392	0.281184
3	0.75	2.316137	6.233473608	-1.8833
4	1	-0.537386	5.478949333	-1.68686
5	1.25	-0.459722	5.369295498	-1.45761
6	1.5	2.956149	6.294232775	-5.18212
7	1.75	1.364187	6.015016672	-5.70906
⋮	⋮	⋮	⋮	⋮
154	38.5	-0.11734	5.454060205	-23.562
155	38.75	-1.93103	4.958367504	**-25.0241**

The maximum value of the log-likelihood function is:

[22] N Herrala "Vasicek Interest Rate Model: Parameter Estimation, Evolution of the Short-Term Interest Rate and Term Structure" https://oa.doria.fi/bitstream/handle/10024?43257/nbnfi-fe200901141021.pdf?.sequence=3 (Accessed 09 September 2009).

$$\ln L = -25.0241$$

for estimators:

$$\hat{\mu} = 5.507031376, \qquad \hat{\alpha} = 7.086842792 \qquad \text{and} \qquad \bar{\sigma} = 0.28436607$$

so that:

$$\hat{\sigma} = \bar{\sigma}\sqrt{\frac{2\hat{\alpha}}{1-e^{-2\hat{\alpha}\Delta t}}} = 1.086401554.$$

Table 5.4: Summary of Vasicek Optimal Estimators

	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\sigma}$
OLS	5.507030788	7.08685326	1.093480328
MLE - Method 2	5.507030788	7.08685326	1.086402711
MLE - Method 3	5.507031376	7.08684279	1.086401554

5.11 Testing for Mean Reversion and Fat Tails for Simulated Model

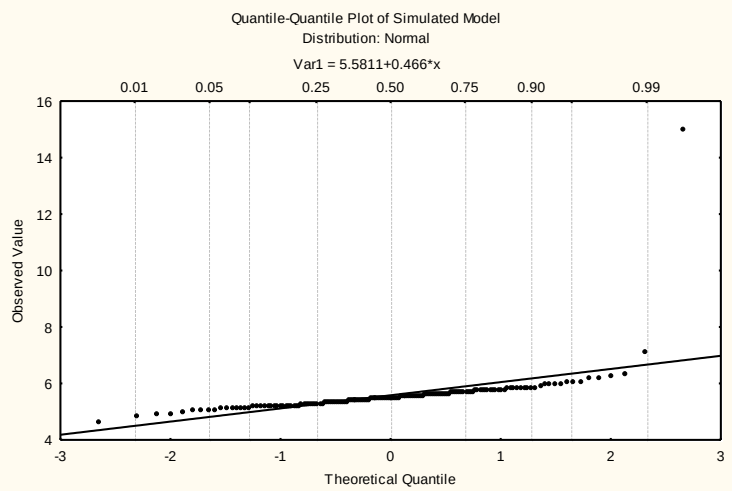

Figure 5.3: QQ-plot for Simulated Vasicek Model

The simulated Vasicek model has a QQ-plot that indicates that the process has normal tails and one need to test if there is any presence of mean reversion. The OLS estimates are very close to the MLE estimates and the original parameters used for the simulated model. The estimators for the OLS and MLE are close due to the fact that there are no fat fails present in the QQ-plot for the generated dataset as well as the simulated model and consequently the biasness in the OLE will not be majorly influenced.

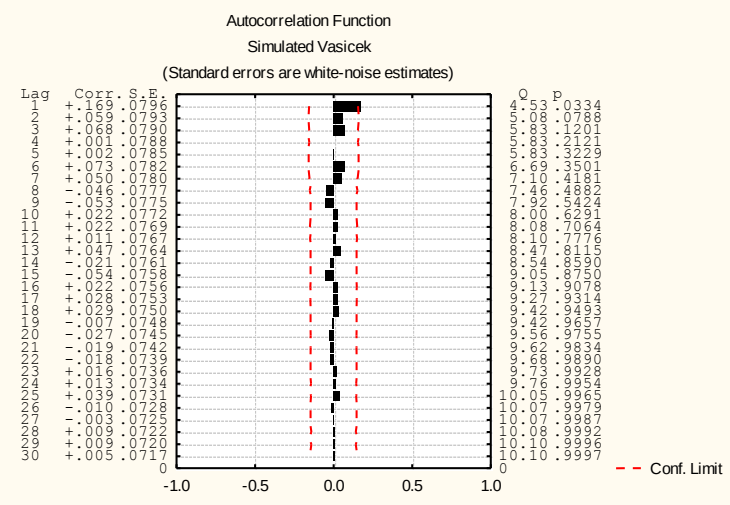

Figure 5.4: Estimated Autocorrelations for Simulated Vasicek Model

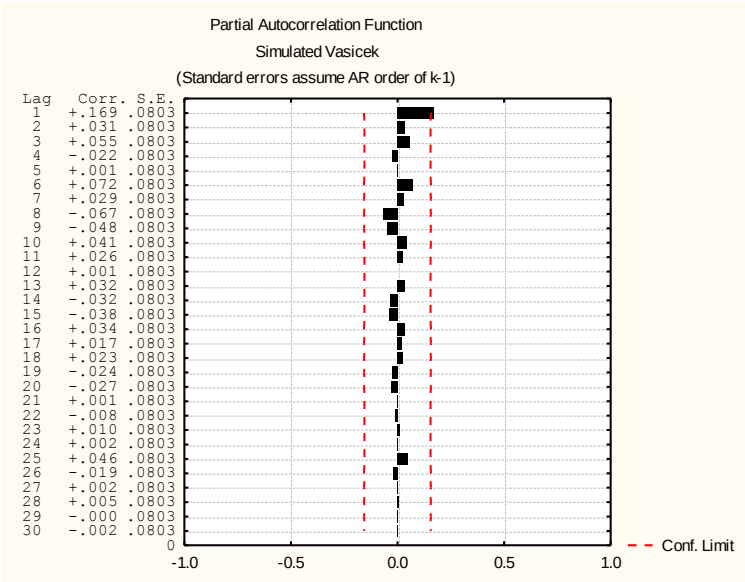

Figure 5.5: Estimated Partial Autocorrelations for Simulated Vasicek Model

5.12 Notes and Research

My aims in this chapter are first to generate the original dataset. One needs to find the risk properties of this generated dataset and obtain a simulated model from Table 1.1 which has the same risk properties. The Vasicek model will be an appropriate simulation model to choose based on the risk properties of the generated dataset. The OLS and two methods for the MLE of a randomly generated time series that consist of 155 observations are calculated.

After calculating the OLS and the two MLE's I showed via the QQ-plot that normal tails exist in the simulated models. Due to the fact that normal tails are present in the simulated models small differences will occur between the OLS and the different MLE methods. The OLS and MLE are approximately the same if the tails are not fat, but if the tails get fatter the more inaccurate the OLS procedure become due to the fact that the OLS method is biased. The mean reversion property for the simulated model is also determined by using the Ljung-Box-Pierce-test to determine the significance of the ACF and PACF.

According to Table 3.1 and the results found in Figure 5.5, the estimated PACF cuts off after lag 1 and the estimated autocorrelations produce damped sine waves which indicates that this simulated model is similar to the $AR(1)$ stationary mean reversion process. According to the Ljung-Box-Pierce-test based on the simulated dataset $Q = 10.10 < \chi^2_{30,0.05} = 43.7729$ indicates H_0 will not be rejected, meaning the simulated Vasicek model does follow a white noise process and significant autocorrelations are not present thus mean reversion is present. These results coincide with Table 1.1 for the risk properties of Vasicek Model. For this reason the Vasicek model will be a good fitted model to choose for this generated dataset according to the generated dataset's risk properties.

Chapter 6

6 THE TWO-FACTOR VASICEK MODEL

The main purpose for this chapter is:

- using the MLE method of finding the optimal estimator to be used for this simulated model.

6.1 Basic Properties of the Two-Factor Vasicek Model

In this study, the two-factor Vasicek model is used to find the optimal multi-period allocation between long and short bonds for an investor.

With reference to J. J. Kung (2008)[23], consider the following two-factor Vasicek interest rate model for the interest rates, which is divided into short rate for the one factor and long rate for the other factor. Let $r(t)$ be the short rate and $l(t)$ be the long rate. The two-factor mean-reverting Vasicek model is defined as:

$$dr(t) = \alpha^r\left(\beta^r - r(t)\right)dt + \sigma^r dZ^r(t) \tag{6.1}$$
$$dl(t) = \alpha^l\left(\beta^l - l(t)\right)dt + \sigma^l dZ^l(t) \tag{6.2}$$

where $dZ^r(t)$ and $dZ^l(t)$ are two Wiener processes such that $dZ^r(t)dZ^l(t) = \rho dt$ and ρ the correlation between the short and long rate. Let α^r and α^l be the measurements of the strength of reversion to their respective mean levels; β^r and β^l, the instantaneous volatilities of changes in the short rate and finally σ^r and σ^l, the instantaneous volatilities of changes in the long rate. Let the price of the short bond and long bond be assigned as $P^r(t)$ and $P^l(t)$ respectively. The price dynamics of the short and long bonds are:

$$\frac{dP^r(t)}{P^r(t)} = r(t)dt \tag{6.3}$$
$$\frac{dP^l(t)}{P^l(t)} = \left[r(t) + v^l\lambda^l\right]dt + v^l dZ^l(t). \tag{6.4}$$

The volatility of $P^l(t)$ is assigned as v^l, while λ^l is the market price for interest rate risk. Let $w(t)$ be a fraction of wealth invested in the short rate and $1 - w(t)$ the remaining wealth invested in the long bond. The dynamics of wealth $V(t)$ are:

$$dV(t) = V(t)w(t)\frac{dP^r(t)}{P^r(t)} + V(t)(1 - w(t))\frac{dP^l(t)}{P^l(t)}. \tag{6.5}$$

[23] JJ Kung "Multi-period asset Allocation by Stochastic Dynamic Programming" (2008) *Applied Mathematics and Computation* 199 341-348.

By substituting equations (6.3) and (6.4) into (6.5) one obtains:

$$dV(t) = \left(V(t)r(t) + V(t)(1-w(t))v^l\lambda^l\right)dt + V(t)(1-w(t))v^l dZ^l(t). \tag{6.6}$$

The long bond is calculated by the following formula:

$$P^l(t) = \frac{C}{(1+l(t))^1} + \frac{C}{(1+l(t))^2} + \cdots + \frac{C}{(1+l(t))^n} + \frac{M}{(1+l(t))^n}. \tag{6.7}$$

where C is the coupon payment, n is the number payments, $l(t)$ is the interest rate of the long bond at time t and M is the value at maturity. [24]

6.2 Maximum Likelihood Estimation for the Two-Factor Vasicek Model

The datasets used for the application of the two-factor Vasicek model are the daily market yield on U.S. Treasury security rates at 6-month constant maturity, and 10-year constant maturity from the 4th of January 1982 up to the 26th of June 2009. There are a total number of 6872 observations for both datasets.

If a coupon of \$10 is paid every six months and the value of coupon at maturity is \$1000 then the price of the long bond in 10-year constant maturity is:

$$P^l(t) = \frac{\$10}{\left(1+\frac{l(t)}{2}\right)^1} + \frac{\$10}{\left(1+\frac{l(t)}{2}\right)^2} + \cdots + \frac{\$10}{\left(1+\frac{l(t)}{2}\right)^{20}} + \frac{\$1000}{\left(1+\frac{l(t)}{2}\right)^{20}}. \tag{6.8}$$

The MLE is based on large sample asymptotic estimators and should be appropriate even if the datasets are not normal. The discrete form of equations (6.1) and (6.2) are:

$$r(t+\Delta t) - r(t) = \alpha^r\left(\beta^r - r(t)\right)\Delta t + \sigma^r\varepsilon^r\sqrt{\Delta t} \tag{6.9}$$

$$l(t+\Delta t) - l(t) = \alpha^l\left(\beta^r - l(t)\right)\Delta t + \sigma^l\varepsilon^l\sqrt{\Delta t} \tag{6.10}$$

where ε^r and ε^l are independent standard normal variables. The natural logarithm of the likelihood functions must be maximised. The first logarithmic likelihood function derived from equation (6.9), using the fact that $r(t+\Delta t) - r(t) - \alpha^r\left(\beta^r - r(t)\right)\Delta t = \sigma^r\varepsilon^r\sqrt{\Delta t} \sim N\left(0, \left(\sigma^r\right)^2\Delta t\right)$, is defined as:

$$\ln L\left(\alpha^r, \beta^r, \sigma^r\right) = -\frac{n}{2}\ln\left(\sigma^r\right)^2\Delta t - \frac{n}{2}\ln(2\pi) - \frac{\sum_{i=1}^{n}\left[r(t+\Delta t) - r(t) - \alpha^r\Delta t\left(\beta^r - r(t)\right)\right]^2}{2\left(\sigma^r\right)^2\Delta t}. \tag{6.11}$$

[24] Investopedia "Advanced Bond Concepts: Bond Pricing" http://www.investopedia.com/university/advancedbond/advancedbond2.asp?viewed=1 (Accessed 30 June 2009).

where the optimal values of α^r, β^r and σ^r indicated respectively as $\hat{\alpha}^r$, $\hat{\beta}^r$ and $\hat{\sigma}^r$ are chosen to maximise (6.11). Let $r(t)$ be assigned as the daily market yield on U.S. Treasury security rates at 6-month constant maturity. The same method applies for the second logarithmic likelihood function derived from equation (6.10), using the fact that $l(t+\Delta t)-l(t)-\alpha^l\left(\beta^r-l(t)\right)\Delta t=\sigma^l\varepsilon^l\sqrt{\Delta t}\sim N\left(0,\left(\sigma^l\right)^2\Delta t\right)$, is given by:

$$\ln L\left(\alpha^l,\beta^l,\sigma^l\right)=-\frac{n}{2}\ln\left(\sigma^l\right)^2\Delta t-\frac{n}{2}\ln(2\pi)-\frac{\sum_{i=1}^{n}\left[l(t+\Delta t)-l(t)-\alpha^l\Delta t\left(\beta^l-l(t)\right)\right]^2}{2\left(\sigma^l\right)^2\Delta t} \tag{6.12}$$

where the optimal values of α^l, β^l and σ^l indicated respectively as $\hat{\alpha}^l$, $\hat{\beta}^l$ and $\hat{\sigma}^l$ are chosen to maximise (6.12). Let $l(t)$ be assigned as the daily market yield on U.S. Treasury security rates at 10-year constant maturity. The discrete form for equation (6.4) is:

$$\frac{\Delta P^l(t)}{P^l(t)}=\left[r(t)+v^l\lambda^l\right]\Delta t+v^l\varepsilon^l\sqrt{\Delta t} \tag{6.13}$$

where ε^l is a standard normal random variable. In order to derive the logarithm of the likelihood function from (6.13) one will obtain:

$$\frac{\Delta P^l(t)}{P^l(t)}-\left(r(t)+v^l\lambda^l\right)\Delta t=v^l\varepsilon^l\sqrt{\Delta t}\sim N\left(0,\left(v^l\right)^2\Delta t\right).$$

and by using $x(t)=\ln\left(\frac{P^l(t)}{P^l(t-1)}\right)$, one can show that

$$x(t)-\left(r(t)+v^l\lambda^l-\frac{\left(v^l\right)^2}{2}\right)\Delta t=v^l\varepsilon^l\sqrt{\Delta t}\sim N\left(0,\left(v^l\right)^2\Delta t\right).$$

The logarithm of the likelihood function using coefficients v^l and λ^l is given as:

$$\ln L\left(v^l,\lambda^l\right)=-\frac{n}{2}\ln\left(v^l\right)^2\Delta t-\frac{n}{2}\ln(2\pi)-\frac{\sum_{i=1}^{n}\left[x(t)-\left(r(t)+v^l\lambda^l-\frac{\left(v^l\right)^2}{2}\right)\Delta t\right]^2}{2\left(v^l\right)^2\Delta t} \tag{6.14}$$

where $x(t)=\ln\left(\frac{P^l(t)}{P^l(t-1)}\right)$ is derived from modifying equation (6.8) for different values of t. The optimal values of v^l and λ^l indicated respectively as $\hat{v}^l$ and $\hat{\lambda}^l$ are chosen to maximise (6.14).

Table 6.1: Estimators $\hat{\alpha}^r$, $\hat{\beta}^r$ and $\hat{\sigma}^r$

Date	t	$r(t)$ - 6 Month	$l(t)$ - 10 years	$\left[r(t+\Delta t)-r(t)-\hat{\alpha}^r\Delta t\left(\hat{\beta}^r-r(t)\right)\right]^2$
01/04/1982	1	0.1316	0.1419	
01/05/1982	2	0.1341	0.1444	6.2500000E-06
01/06/1982	3	0.1346	0.1459	2.5000000E-07
⋮	⋮	⋮	⋮	⋮
06/26/2009	6872	0.0031	0.0352	3.1354652E-10

By assigning the daily market yield on U.S. Treasury security rates at 6-month constant maturity as $r(t)$ and the 10-year constant maturity as $l(t)$ from the 4th of January 1982 up to the 26th of June 2009 one may use Table 6.1 to determine the optimal estimators $\hat{\alpha}^r$, $\hat{\beta}^r$ and $\hat{\sigma}^r$ which will maximise equation 6.11. I used the Microsoft Excel solver add-inn function to maximise equation 6.11 to obtain:

$$\ln L\left(\hat{\alpha}^r,\hat{\beta}^r,\hat{\sigma}^r\right)=-\frac{6872}{2}\ln\left(\hat{\sigma}^r\right)^2\Delta t-\frac{6872}{2}\ln(2\pi)-\frac{\sum_{i=1}^{6872}\left[r(t+\Delta t)-r(t)-\hat{\alpha}^r\Delta t\left(\hat{\beta}^r-r(t)\right)\right]^2}{2\left(\hat{\sigma}^r\right)^2\Delta t}$$
$$=3941491$$

The same method was applied to obtain the maximum logarithmic likelihood functions:

$$\ln L\left(\hat{\alpha}^l,\hat{\beta}^l,\hat{\sigma}^l\right)=40116.62 \qquad \text{and} \qquad \ln L\left(\hat{v}^l,\hat{\lambda}^l\right)=14720.51.$$

Table 6.2: Summary of the Estimators for Two-Factor Vasicek Model

Parameter	MLE
$\hat{\alpha}^r$	0.000714234
$\hat{\beta}^r$	0.027891921
$\hat{\sigma}^r$	0.000781326
$\hat{\alpha}^l$	0.000817419
$\hat{\beta}^l$	0.050127054
$\hat{\sigma}^l$	0.000705433
$\hat{v}^l$	0.028409325
$\hat{\lambda}^l$	-1.884035964

The optimal estimators in Table 6.2 are used as well as backward recursion to optimise proportion $w(t)$ expressed in equation (6.5).

6.3 Notes and Research

Only the estimators for the simulated Two-Factor Vasicek Model using the MLE are illustrated in this chapter.

Chapter 7

7 MEAN REVERSION – THE CIR MODEL

The square root process (CIR process) is a widely used for modelling interest rates. This process is an underlying process for the Cox-Ingersoll-Ross term structure model (1985).

The main purposes for this chapter are:

- finding the risk properties for the original dataset (the historical model involve 30-day AA financial interest rates)
- choosing the simulated model which should have the same risk properties as the original dataset. Table 1.1 is used to find the most appropriate simulated model with the same risk properties as the original dataset
- finding different methods of calculating the estimators for this simulated model. These methods used to find the estimators are called OLS estimation and MLE
- comparing the accuracy of the OLS estimates and the MLE by observing whether the model have fat or normal tails. This is done by using the QQ-plots
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

7.1 Basic Properties – CIR Model

The dynamic representation of the CIR model according to K. Kladivko [25] is:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{7.1}$$

where W_t is a standard Brownian motion for $t \geq 0$. The functions $\mu(\cdot)$ and $\sigma(\cdot)$ are respectively the drift and the diffusion functions of the process. The square root process (CIR process) given in the following stochastic differential equation is the fundamental process in interest rate modelling.

$$dx_t = \alpha(\mu - x_t)dt + \sigma\sqrt{x_t}dW_t \tag{7.2}$$

where x_t is the interest rate and $\theta = (\alpha, \mu, \sigma)$ are the model parameters. It has the mean reverting property, i.e. interest rate x_t moves in the direction of its mean μ at speed α. The drift function is known as $\mu(x_t, \theta) = \alpha(\mu - x_t)$ and is linear. The diffusion function $\sigma^2(x_t, \theta) = x_t\sigma^2$ is proportional to the interest rate x_t.

[25] K Kladivko "Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Process: The Matlab Implementation" http://dsp.vscht.cz/konference_matlab/MATLAB07/prispevky/kladivko_k/kladivko_k.pdf (Accessed 9 September 2009).

7.2 Maximum Likelihood Estimate - CIR Model

Assume a discrete time version on the time grid $t_1, t_2, ..., t_n$ with time step $\Delta t = t_i - t_{i-1}$ will be used in this section.

If x_{t_i} is given, the conditional density function g of $x_{t_{i+1}}$ is:

$$g\left(x_{t_{i+1}} \mid x_{t_i}; \theta\right) = ce^{-u-v}\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right)^{\frac{q}{2}} I_q\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right)$$

where

$$\begin{aligned} c &= \frac{2\alpha}{\sigma^2\left(1-e^{-\alpha\Delta t}\right)} \\ u_{t_i} &= cx_t e^{-\alpha\Delta t} \\ v_{t_{i+1}} &= cx_{t_{i+1}} \\ q &= \frac{2\alpha\mu}{\sigma^2} - 1 \end{aligned}$$

(as indicated by A.S. Hurn *et al* (1999)[26]) and $I_q\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right)$ is a modified Bessel function of the first kind and of order q. The likelihood function for the interest rate x_t with n observations is:

$$L(\theta) = \prod_{i=1}^{n-1} g\left(x_{t_{i+1}} \mid x_{t_i}; \theta\right). \tag{7.3}$$

The log-likelihood function is given by:

$$\begin{aligned} \ln L(\mu, \alpha, \sigma) &= \sum_{i=1}^{n-1} \ln\left(g\left(x_{t_{i+1}} \mid x_{t_i}; \theta\right)\right) \\ &= (n-1)\ln(c) + \sum_{i=1}^{n-1}\left(-u_{t_i} - v_{t_{i+1}} + 0.5q\ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + \ln\left[I_q\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right)\right]\right) \end{aligned} \tag{7.4}$$

[26] AS Hurn *et al* "On the Efficacy of Simulated Maximum Likelihood For Estimating the Parameters of Stochastic Differential Equations" (1999) *Journal of Time Series Analysis* Vol 24 No 1.

where $u_{t_i} = cx_{t_i} e^{-\alpha\Delta t}$ and $v = cx_{t_{i+1}}$. The log-likelihood function needs to be maximised by taking partial derivatives of equation (7.4) with respect to μ, α and σ, putting them equal to zero yield three equations:

$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \mu}|_{\hat{\mu}} = 0.$$
$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \alpha}|_{\hat{\alpha}} = 0.$$
$$\frac{\partial \ln L(x_t;\mu,\alpha,\sigma)}{\partial \sigma}|_{\hat{\sigma}} = 0.$$

Solving these equations will yield the maximum likelihood estimates:

$$\hat{\theta} = (\hat{\alpha},\hat{\mu},\hat{\sigma}) = \arg\max_{\theta} \ln L(\theta).$$

7.3 The Ordinary Least Squares Estimation Procedure for CIR-Model

The starting points for optimisation are vital and the Ordinary Least Squares method is used to determine these points.

The simulation of equation (7.2) is illustrated as:

$$x_{t_{i+1}} - x_{t_i} = \alpha(\mu - x_{t_i})\Delta t + \sigma\sqrt{x_{t_i}}\,\varepsilon_{t_i} \tag{7.5}$$

where $\varepsilon_{t_i} \sim N(0,\Delta t)$ and also as:

$$x_{t_{i+1}} = \alpha\mu\Delta t + (1-\alpha\Delta t)x_{t_i} + \sigma\sqrt{x_{t_i}\Delta t}\,\varepsilon_{t_i} \tag{7.6}$$

where $\varepsilon_{t_i} \sim N(0,1)$. In order to use the OLS, equation (7.5) must be transformed to:

$$\frac{(x_{t_{i+1}} - x_{t_i})}{\sqrt{x_{t_i}}} = \frac{\alpha\mu\Delta t}{\sqrt{x_{t_i}}} - \alpha\sqrt{x_{t_i}}\,\Delta t + \sigma\varepsilon_{t_i}.$$

The sum square of the error $\sum_{i=1}^{n-1}(\sigma\varepsilon_{t_i})^2$ must be minimised in terms of α and μ to obtain $\hat{\alpha}$ and $\hat{\mu}$ such that:

$$(\hat{\alpha},\hat{\mu}) = \arg\min_{\alpha,\mu}\sum_{i=1}^{n-1}(\sigma\varepsilon_{t_i})^2 = \arg\min_{\alpha,\mu}\sum_{i=1}^{n-1}\left(\frac{(x_{t_{i+1}} - x_{t_i})}{\sqrt{x_{t_i}}} - \frac{\alpha\mu\Delta t}{\sqrt{x_{t_i}}} + \alpha\sqrt{x_{t_i}}\,\Delta t\right)^2$$

which will be:

$$\hat{\alpha} = \frac{n^2 - 2n + 1 + \sum_{i=1}^{n-1} x_{t_{i+1}} \sum_{i=1}^{n-1} \frac{1}{x_{t_i}} - \sum_{i=1}^{n-1} x_{t_i} \sum_{i=1}^{n-1} \frac{1}{x_{t_i}} - (n-1) \sum_{i=1}^{n-1} \frac{x_{t_{i+1}}}{x_{t_i}}}{\left(n^2 - 2n + 1 - \sum_{i=1}^{n-1} x_{t_i} \sum_{i=1}^{n-1} \frac{1}{x_{t_i}} \right) \Delta t}$$

$$\hat{\mu} = \frac{(n-1) \sum_{i=1}^{n-1} x_{t_{i+1}} - \sum_{i=1}^{n-1} \frac{x_{t_{i+1}}}{x_{t_i}} \sum_{i=1}^{n-1} x_{t_i}}{\left(n^2 - 2n + 1 + \sum_{i=1}^{n-1} x_{t_{i+1}} \sum_{i=1}^{n-1} \frac{1}{x_{t_i}} - \sum_{i=1}^{n-1} x_{t_i} \sum_{i=1}^{n-1} \frac{1}{x_{t_i}} - (n-1) \sum_{i=1}^{n-1} \frac{x_{t_{i+1}}}{x_{t_i}} \right)}.$$

The standard deviation, $\hat{\sigma}$, of the errors is the estimated diffusion parameter,

$$\hat{\sigma} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-1} \left(\frac{x_{t_{i+1}} - x_{t_i}}{\sqrt{x_{t_i}}} - \frac{\hat{\mu}}{\sqrt{x_{t_i}}} + \hat{\alpha} \sqrt{x_{t_i}} \right)^2}.$$

7.4 The MLE for CIR-Model Using MATLAB Command - *besseli*

In MATLAB the modified Bessel function of the first kind is available under the command $besseli\left(q, 2\sqrt{u_{t_i} v_{t_{i+1}}}\right)$ but calculations results in an estimation failure. The scaled version of the Bessel function in MATLAB is denoted as $I_q^1\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right)$ and its command in MATLAB is $besseli\left(q, 2\sqrt{u_{t_i} v_{t_{i+1}}}, 1\right)$.

The modified Bessel function can be defined as:

$$I_q\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right) = \exp\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right) I_q^1\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right). \tag{7.7}$$

By substituting equation (7.7) in (7.4) will yield a log-likelihood function given by:

$$\ln L(\mu, \alpha, \sigma) = (n-1)\ln(c) + \sum_{i=1}^{n-1} \left(-u_{t_i} - v_{t_{i+1}} + 0.5q \ln\left(\frac{v_{t_{i+1}}}{u_{t_i}} \right) + 2\sqrt{u_{t_i} v_{t_{i+1}}} + \ln\left[I_q^1\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right) \right] \right) \tag{7.8}$$

The Bessel function of the first kind for order q according to R. Haberman, Fourth Edition (2004)[27] is defined as:

$$\tilde{I}_q(z) = \begin{cases} 1 & q = 0 \\ \dfrac{1}{2^q q!} z^q & q > 0. \end{cases}$$

and $\tilde{I}_{-q}(z) = (-1)^q \tilde{I}_q(z)$ and according to www.mathworks.com[28] the modified Bessel's equation for noninteger q is defined as:

$$I_q(z) = \left(\frac{z}{2}\right)^q \sum_{k=0}^{\infty} \left(\left(z^2 / 4\right)^k \Big/ k!\Gamma(q+k+1) \right)$$

where $\Gamma(\alpha)$ is a gamma function. The negative logarithmic likelihood function given by:

$$\ln L(\mu, \alpha, \sigma) = -(n-1)\ln(c) - \sum_{i=1}^{n-1} \left(-u_{t_i} - v_{t_{i+1}} + 0.5q \ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + 2\sqrt{u_{t_i} v_{t_{i+1}}} + \ln\left[I_q^1\left(2\sqrt{u_{t_i} v_{t_{i+1}}}\right)\right]\right)$$

is minimised by using the 'fminsearch' functionality provided by Matlab.

7.5 The MLE for CIR-Model Using MATLAB Command - *ncx2pdf*

The optimal estimators can be derived from an alternative maximum likelihood estimation procedure. The MATLAB program uses the non-central χ^2 probability density function available in the Statistics Toolbox under *ncx2pdf* command. The modified Bessel function of the first kind will not be used but rather the non-central χ^2 distribution, which is based on the central χ^2 distribution weighted by a Poisson distribution.

[27] R Haberman *Applied Partial Differential Equations: with Fourier Series and Boundary Value Problems* 4 ed (2004) 308.

[28] Mathworks http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/ref/besseli.html&http://www.mathworks.com/cgi-bin/texis/webinator/search/?db=MSS&prox=page&rorder=750&rprox=750&rdfreq=500&rwfreq=500&rlead=250&sufs=0&order=r&is_summary_on=1&ResultCount=10&query=bessel

The density function for the interest process according to S. Zeytun (2007)[29] is:

$$f_{x_t}(z) = f_{\chi^2(v,\lambda_t)/c_t}(z) = c_t f_{\chi^2(v,\lambda_t)}(c_t z)$$

where

$$c_t = \frac{4\alpha}{\sigma^2\left(1-e^{-4\alpha}\right)}$$
$$v = \frac{4\alpha\mu}{\sigma^2}$$
$$\lambda_t = c_t x_0 \exp(-\alpha t)$$

and $\chi^2(v,\lambda_t)$ is the non-central chi-squared, χ^2, distribution function with v degrees of freedom and non-central parameter λ_t.

[29] S Zeytun and A Gupta "A Comparative Study of the Vasicek and the CIR Model of the Short Rate" (2007) *ITWM* 124.

7.6 Example 7.1 –Estimators for the CIR Model

Consider the data, which consists of the 30-day AA financial interest rates. The data was annualised using a 360-day year or bank interest from the 1st of January 1997 to the 26th of June 2009. The dataset consists of 3131 observations. The Ljung-Box-Pierce test for the 30-day AA financial interest rates is:

$$Q = 572972.4 > \chi^2_{780,0.05} = 846.083$$

The corresponding $p-value$ is 2.2e-16 which is less than 0.05. The null hypothesis (H_0) will be rejected indicating that the distribution of the original dataset is not mean reverting.

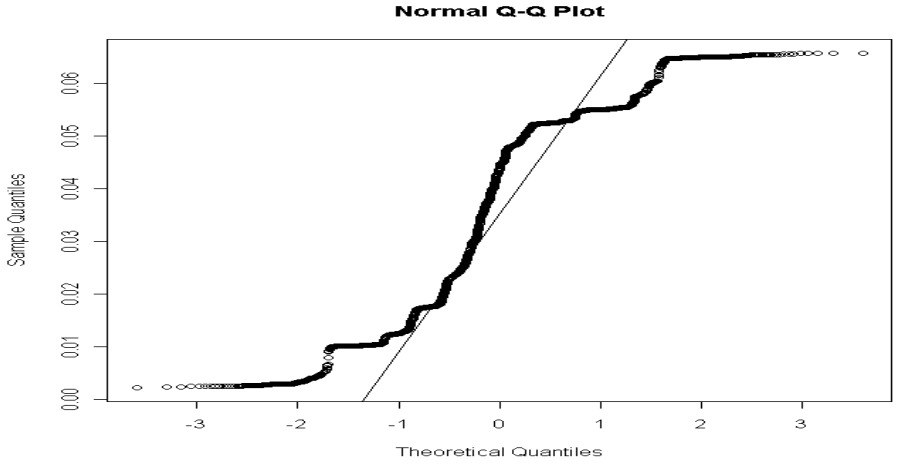

Figure 7.1: QQ-Plot for 30-day AA financial interest rates

According to the QQ-plot for the 30-day AA financial interest rates fat tails are present. The CIR simulation model will not be the most appropriate model for this dataset. According to Table 1.1 one should rather choose the GBM + Jumps, One-Factor Logarithmic Vasicek Model + Jumps. I choose the CIR model only for example purposes. In order to use the CIR simulation model, one needs to determine the optimal estimators. The ordinary least squares, optimal maximum likelihood estimation using the *besseli* command and the optimal maximum likelihood estimation using the *ncx2pdf* command are given in Table 7.1. The MATLAB program used to determine the MLE, using the *besseli* command is given in appendix B1 and the MATLAB program for *ncx2pdf* command in appendix B2. The OLS estimation program is included at the beginning of each of the *besseli* and *ncx2pdf* command programs, which serves as the initial values for the estimators.

Table 7.1: Estimators for CIR Model

	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\sigma}$	$\ln L$
OLS (initial)	0.184091278	0.037013893	0.139535942	
ML - *besseli*	0.283542293	0.016454519	0.107825364	-1.705689364245570e+004
ML – *ncx2pdf*	0.169209129	0.016931142	0.107884752	-1.705689364056405e+004

The time step used from 1st of January 1997 to the 26th of June 2009 for the 30-Day AA financial interest rates is $\Delta t = 1/360$. Using equation (7.6) the CIR-Model simulation, together with the 30-day AA financial interest rates, are plotted in

Figure 7.2. The solid line indicates the original interest rates and the dotted line is the simulated CIR-Model using *besseli*'s maximum likelihood estimates. No mean reversion is clearly visible from the solid black line in Figure 7.2. Choose an initial starting value $x_0 = 0.0535$, for the simulated CIR-Model. I also attempt the MLE procedure in Microsoft Excel which resulted in inconclusive results.

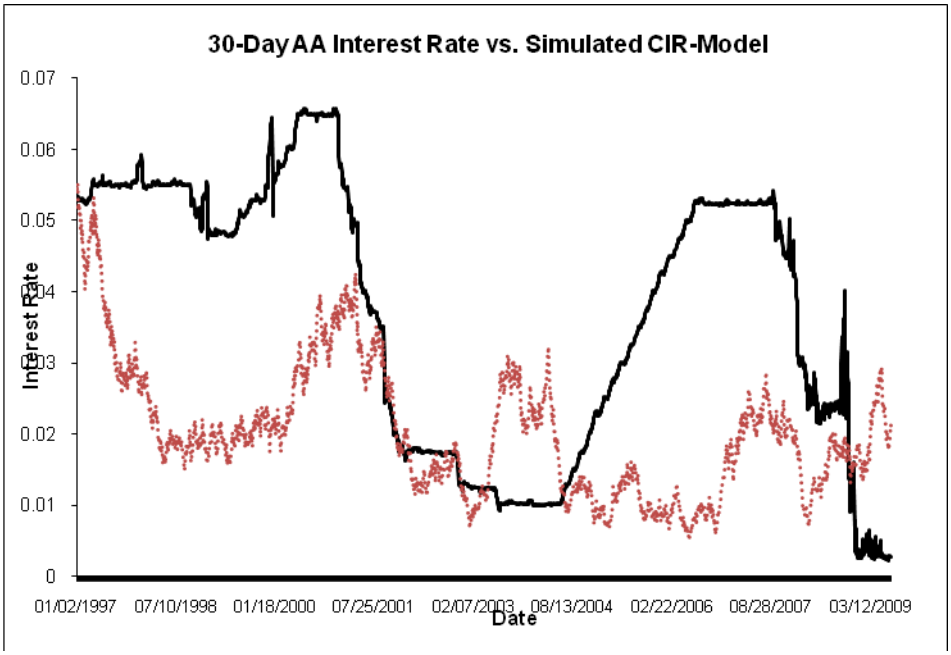

Figure 7.2: A Simulated CIR-Model, $x_{t_{i+1}} = \hat{\alpha}\hat{\mu}\Delta t + (1 - \hat{\alpha}\Delta t)x_{t_i} + \hat{\sigma}\sqrt{x_{t_i}\Delta t}\varepsilon_{t_i}$

7.7 Testing for Mean Reversion and Fat Tails for Simulation Model

According to Figure 7.3 fat fails are present in the QQ-plot for the simulated CIR model.

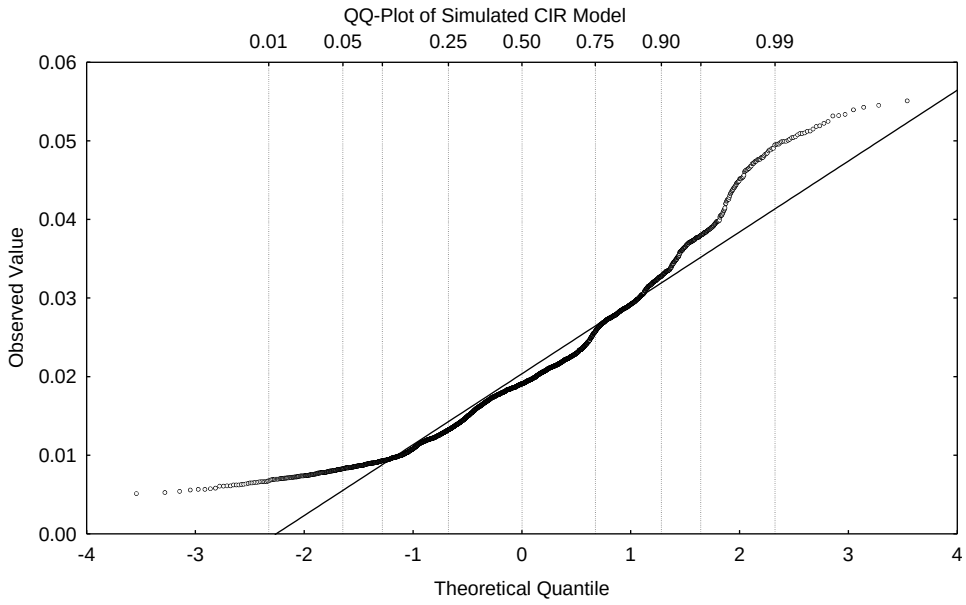

Figure 7.3: QQ-plot for the Simulated CIR Model

7.8 Notes and Research

The null hypothesis, H_0, is rejected for the Ljung-Box-Pierce test on the original dataset (30-day AA financial interest rates). The test statistic on the simulated model is rejected for H_0 meaning the the simulated CIR model (using *besseli's* maximum likelihood estimates in simulated model) do not follow a white noise process and significant autocorrelations are present thus mean reversion is not present (no stationarity present). Fat tails are observed from the QQ-plots of the original dataset (Figure 7.1) and the simulated CIR model (Figure 7.3).

The risk properties of the original dataset do not coincide with Table 1.1 for the risk properties of the simulated CIR model. This means that the fitted CIR model on the 30-day AA financial interest rates is a rough estimated model and one might try to fit a GBM with jumps on the dataset or one could choose a different time window on the dataset and try a fitted CIR model again.

From observing Figure 7.1 and Figure 7.3, fat tails are present for the original dataset and the simulated model. Due to the presence of fat tails the differences for the estimators for the OLS and MLE are large indicated in Table 7.1.

Chapter 8

8 NO MEAN REVERSION – AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (*ARCH*) MODELS

The *ARCH* models were first introduced by Engle (1982) and are used to model conditional variances. Suppose a time series exhibit periods of unusually large volatility, followed by periods of tranquillity, we cannot assume a constant variance. One might want to forecast the conditional variance of a series.

The main purposes for this chapter are:

- finding the risk properties for the volatility of the original dataset (the historical model involve S&P 500 closing prices)
- construct several simulated volatility models (which include different ARCH and GARCH models) that should have the same risk properties as the volatility of the original dataset
- finding the estimators for this simulated volatility model. The method used to find the estimators are called Berndt-Hall-Hall-Hausman Iteration
- finding the risk properties of the simulated model
- comparing the original and simulated risk properties and analysing the appropriateness of the simulated model.

8.1 General Properties for the *ARCH(q)* Process

Consider the following time series, $x_1, x_2, ..., x_n$ and assume that the returns of this time series is defined by $y_t = \ln(x_t / x_{t-1})$. Let F_t denote the information set at time t, which include y_t and all the past realisations of the process y_t. The *ARCH* model will be applied on the returns.

The $ARCH(q)$ model for the returns are:

$$y_t = E_{t-1}[y_t] + \varepsilon_t \tag{8.1}$$

$$\varepsilon_t = z_t \hat{\sigma}_t \tag{8.2}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{8.3}$$

The *ARCH* model produced a revolution in time series because it has properties that the Autoregressive Moving Average Model (*ARMA*) is not able to produce. These properties are:

- The return series show little serial correlation.
- The serial correlation of the absolute or squared returns are significant where $\rho(\varepsilon_t^2, \varepsilon_{t-1}^2) = \alpha \neq 0$.
- The expected returns are close to zero, such that $E[\varepsilon_t \mid F_{t-1}] = 0$.
- The volatility vary over time.

- The return series are leptokurtic or heavy-tailed (fat tails). This check is done by the kurtosis, which must always be positive and infinite if $3\alpha^2 \geq 1$.
- Extreme returns appear in clusters.

The unconditional variance, $\bar{\sigma}^2$, for the $ARCH(q)$ model is defined by:

$$\begin{aligned}\bar{\sigma}^2 &= Var(\hat{\sigma}_t^2) = E(\sigma_t^2) - [E(\sigma_t)]^2 \\ &= E(\sigma_t^2) \\ &= E(\alpha_0) + \alpha_1 E(\varepsilon_{t-1}^2) + ... + \alpha_q E(\varepsilon_{t-q}^2) \\ &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 + ... + \alpha_q \sigma_{t-q}^2 \\ &= \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i}.\end{aligned}$$

The unconditional variance will be finite (converge to a constant value) and non-negative if the coefficients α_i have conditions, $\alpha_0 \geq 0$, $\alpha_i \geq 0$, for $i = 1,2,...,q$ and $0 \leq \sum_{i=1}^{q} \alpha_i < 1$. Equations (8.3) can be written as an $AR(q)$ for ε_t^2 if $u_t = \varepsilon_t^2 - \hat{\sigma}_t^2$, such that:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + u_t$$

where $E_{t-1}[u_t] = 0$ and $E(\varepsilon_t^2) < \infty$.

8.2 General Properties for the *ARCH(1)* Process

One can easily derive $ARCH(1)$ from $ARCH(q)$ if $q = 1$ in (8.3):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \tag{8.4}$$

The conditional distribution for ε_t given ε_{t-1} is:

$$(\varepsilon_t \mid \varepsilon_{t-1}) \sim N(0, \sigma_t^2).$$

This model (8.4) is known as the $ARCH(1)$ model, since the conditional variance depends only on one lagged squared error. The $ARCH(1)$ model has stationary moments of the second and fourth order, if $3\alpha_0^2 < 1$. These moments are:

$$\begin{aligned}E(\varepsilon_t^2) &= \frac{\alpha_0}{(1 - \alpha_1)} \\ E(\varepsilon_t^4) &= \frac{3\alpha_0^2}{(1 - \alpha_1)^2} \frac{(1 - \alpha_1^2)}{(1 - 3\alpha_1^2)}.\end{aligned}$$

The proof of the forth moment, $E(\varepsilon_t^4)$ according to P. Cizck et al (2005)[30] is:

$$c = E[\varepsilon_t^4] = E[E(\varepsilon_t^4 \mid \varepsilon_{t-1}^4)] = E[\sigma_t^4 E(Z_t^4 \mid \varepsilon_{t-1}^4)] = E[Z_t^4]E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2]$$
$$= 3(\alpha_0^2 + 2\alpha_0\alpha_1 E[\varepsilon_{t-1}^2] + \alpha_1^2 E[\varepsilon_{t-1}^4]) \quad (8.5)$$

and let $Z_t = \frac{\varepsilon_t}{\hat{\sigma}_t} \sim N(0,1)$, using $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2) = \frac{\alpha_0}{(1-\alpha_1)}$ and $c = E[\varepsilon_{t-1}^4]$, equation (8.5) becomes:

$$c = 3(\alpha_0^2 + 2\alpha_0\alpha_1(\alpha_0/(1-\alpha_1)) + \alpha_1^2 c).$$

Solving c, one will obtain the fourth moment. The kurtosis coefficient κ for $ARCH(1)$ model, is then defined as:

$$Kurt(\varepsilon_t) = \kappa = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} = \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} \geq 3.$$

If the kurtosis coefficient $\kappa = 3$ is normal and if $\kappa > 3$, it then indicates that fat tails are present. This is called leptokurtosis.

8.3 Generalised Autoregressive Conditional Heteroscedasticity (*GARCH*) Models

For the *GARCH* models the conditional variance is dependent upon its previous lags. The $GARCH(1,1)$ model is:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (8.6)$$

According to P. Hanson and A. Lunde (2004)[31], there is compelling evidence that it is difficult to find a volatility model that outperforms the simple $GARCH(1,1)$ model. This model can be expanded to become the $GARCH(q, p)$ model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_p \sigma_{t-p}^2 \quad (8.7)$$
$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 .$$

where the current conditional variance depends upon q lags of the squared error ($q-order$ moving average *ARCH* term) and p lags of the conditional variance ($p-order$ autoregressive *GARCH* term).

---

[30] P Cizck (2005) "*ARCH* and *GARCH* Models" http://mars.wiwi.hu-berlin.de/tutorials/sfehtm1node66.html (Accessed 25 February 2009).

[31] P Hansen *et al* "A Forecast Comparison of Volatility Models: Does Anything Beat a *GARCH*(1,1) Model?" (2004) 20 *Journal of Applied Econometrics* 873-889.

For a stationary $GARCH(q,p)$, the coefficients α_i and β_j for the unconditional variance must have the following restrictions:

$$\alpha_i \geq 0, \text{ for } i = 0,1,2,...,q,$$
$$\beta_j \geq 0, \text{ for } i = 1,2,...,p$$
$$0 \leq \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

The unconditional variance, $\bar{\sigma}_t^2$, for $GARCH(q,p)$, under these conditions will be finite and non-negative. The unconditional variance, $\bar{\sigma}_t^2$, for $GARCH(q,p)$ is:

$$\bar{\sigma}_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

for $0 \leq \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$

The fourth moment for the $GARCH(1,1)$ process according P. Cizck et al (2005)[32] is given as:

$$Kurt(\varepsilon_t) = \kappa = \frac{E(\varepsilon_t^4)}{\left[E(\varepsilon_t^2)\right]^2} = 3 + \frac{6\alpha_1^2}{(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)}.$$

It can be proven that $E_{t-1}(\varepsilon_t^2) = \hat{\sigma}_t^2$ and with $u_t = \varepsilon_t^2 - E_{t-1}(\varepsilon_t^2)$, equation (8.6) can be written as:

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 - \beta_1)\varepsilon_{t-1}^2 + u_t - \beta_1 u_{t-1} \tag{8.8}$$

which constitutes an $ARMA(1,1)$ model.

8.4 Estimation of *ARCH-GARCH* Models

Assume the error, ε_t, obtained from the return, y_t, needs to be modelled with an $ARCH(1)$ process for $t = 0,1,2,...,n$. Two parameters, α_0, α_1 according to equation (8.4), needs to be maximised through the maximum likelihood estimator. The joint conditional density function for the first given value ε_0, is:

$$f(\varepsilon_n, \varepsilon_{n-1}, ..., \varepsilon_1 \mid \varepsilon_0) = f(\varepsilon_n \mid \varepsilon_{n-1}) f(\varepsilon_{n-1} \mid \varepsilon_{n-2}) \times \cdots \times f(\varepsilon_1 \mid \varepsilon_0). \tag{8.9}$$

By substituting equation (8.4) into (8.2), one will obtain:

[32] P Cizck "*ARCH* and *GARCH* Models".

$$\varepsilon_t = \sigma_t z_t = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}\, z_t$$

where $(\varepsilon_t \mid \varepsilon_{t-1}) \sim N(0, \sigma_t^2)$, $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$. This leads to the conclusion that:

$$f(\varepsilon_n \mid \varepsilon_{n-1}) = \frac{1}{\sigma_t} \Phi\left(\frac{\varepsilon_t}{\sigma_t}\right). \tag{8.10}$$

Let $\Phi(x)$ denote the standard normal distribution function such that:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Using equation (8.10), equation (8.9) becomes the conditional likelihood function:

$$\begin{aligned} L(\alpha_0, \alpha_1) &= f(\varepsilon_n, \varepsilon_{n-1}, ..., \varepsilon_1 \mid \varepsilon_0) = f(\varepsilon_n \mid \varepsilon_{n-1}) f(\varepsilon_{n-1} \mid \varepsilon_{n-2}) \times \cdots \times f(\varepsilon_1 \mid \varepsilon_0) \\ &= \frac{1}{\sigma_n} \Phi\left(\frac{\varepsilon_n}{\sigma_n}\right) \times \frac{1}{\sigma_{n-1}} \Phi\left(\frac{\varepsilon_{n-1}}{\sigma_{n-1}}\right) \times \cdots \times \frac{1}{\sigma_1} \Phi\left(\frac{\varepsilon_1}{\sigma_1}\right) \\ &= \frac{1}{(2\pi)^{n/2} \sigma_n \sigma_{n-1} \times \cdots \times \sigma_1} \exp\left(-\frac{1}{2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{\sigma_t^2}\right) \\ &= \frac{1}{(2\pi)^{n/2} \left(\prod_{t=1}^{n} \sigma_t\right)} \exp\left(-\frac{1}{2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{\sigma_t^2}\right). \end{aligned} \tag{8.11}$$

Taking the natural logarithm of the conditional likelihood function, equation (8.11) becomes:

$$\begin{aligned} \ln L(\alpha_0, \alpha_1) &= -\frac{1}{2} \sum_{t=1}^{n} \left(\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2} \right) \\ &= -\frac{1}{2} \sum_{t=1}^{n} \left(\ln(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) + \frac{\varepsilon_t^2}{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \right) \end{aligned} \tag{8.12}$$

where the constant term $(n/2)\ln(2\pi)$ is discarded.

The optimal values for the parameters (α_0, α_1) are determined by equating the derivatives of these parameters to zero:

$$\begin{aligned} \frac{\partial \ln L(\alpha_0, \alpha_1)}{\partial \alpha_0} &= -\frac{1}{2} \sum_{t=1}^{n} \left(\frac{1}{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} - \frac{\varepsilon_t^2}{(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2} \right) \\ &= -\frac{1}{2} \sum_{t=1}^{n} \left(\frac{1}{\sigma_t^2} - \frac{\varepsilon_t^2}{\sigma_t^4} \right) = 0 \end{aligned} \tag{8.13}$$

and

$$\frac{\partial \ln L(\alpha_0,\alpha_1)}{\partial \alpha_1} = -\frac{1}{2}\sum_{t=1}^{n}\varepsilon_{t-1}^2\left(\frac{1}{\sigma_t^2}-\frac{\varepsilon_t^2}{\sigma_t^4}\right)=0. \tag{8.14}$$

One cannot find explicit solutions for equations (8.13) and (8.14), to obtain the optimal estimators $(\hat{\alpha}_0,\hat{\alpha}_1)$. A recursive method, called the Berndt-Hall-Hall-Hausman (BHHH) method proposed by Engle (1982)[33] is used. Let

$$\theta=(\alpha_0,\alpha_1)' \text{ and } \frac{\partial \ln L}{\partial \theta}=\left(\frac{\partial \ln L}{\partial \alpha_0},\frac{\partial \ln L}{\partial \alpha_1}\right)'$$

where (8.12) becomes:

$$\ln L_t(\theta)=-\frac{1}{2}\left(\ln(\sigma_t^2)+\frac{\varepsilon_t^2}{\sigma_t^2}\right).$$

One may say that

$$\ln L=\sum_{t=1}^{n}\ln L_t$$
$$\frac{\partial \ln L}{\partial \theta}=\sum_{t=1}^{n}\frac{\partial \ln L_t}{\partial \theta}.$$

8.5 The Berndt-Hall-Hall-Hausman Iteration Method

The following steps are needed to determine the optimal estimators:

- Firstly, one needs to choose the initial values as $\theta_0=(\alpha_{00},\alpha_{01})$.
  Secondly, let $v_k=\sum_{t=1}^{n}\frac{\partial \ln L_t}{\partial \theta}(\theta_k)$ be a vector after k steps.
- Consequently, a matrix $B=\sum_{t=1}^{n}\frac{\partial \ln L_t}{\partial \theta}\left(\frac{\partial \ln L_t}{\partial \theta}\right)'$ needs to be computed.
- Calculate a matrix defined as: $\theta_{k+1}=\theta_k+B^{-1}v_k$.
- The iteration will stop when the values stabilises.

After one applied the Berndt-Hall-Hall-Hausman Iteration Method in minimising the likelihood function $\ln L_t(\theta)$ in terms of σ. The optimal σ assigned as $\hat{\sigma}$ is now substituted into the GARCH volatility models.

[33] RF Engle "The Use of *ARCH/GARCH* Models in Applied Econometrics" (2006) 15 *Journal of Economic Perspectives* 157-168.

8.6 Test for Normality of Dataset

Before and after the analysis on the dataset is done, a normality test called the Jarqu-Bera (JB) is performed. The test statistic is:

$$Jarque-Beru = \frac{n}{6}\left(S^2 + \frac{(\kappa - 3)}{4} \right) \sim \chi_2^2$$

where S is the skewness, and κ is the kurtosis. The JB test statistic follows a chi-square distribution with 2 degrees of freedom. The skewness is a measure of asymmetry of the distribution of the series around its mean. It may be expressed as:

$$S = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{\hat{\hat{\sigma}}} \right)^3 .$$

The kurtosis will chance depending on the *ARCH / GARCH* model used. The estimator for the standard deviation, that is on the biased estimator for the variance, is defined as:

$$\hat{\hat{\sigma}} = s\sqrt{\frac{n-1}{n}}$$

where "s" is the standard deviation of the returns, indicated as:

$$s = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}} .$$

The hypothesis test with Jarqu-Bera is:

H_0 : the data is normally distributed versus
H_1 : the data is not normally distributed.

One will reject H_0, if $JB > \chi^2_{2;\alpha}$, where 2 is the degrees of freedom and α is the level of significance. One may also reject H_0 if the $p-value$ is less than 0.05.

8.7 Application on the *ARCH* and *GARCH* Models

Let x_t denote the closing price of the S&P500 at the end of trading day t. The closing prices for the S&P500 index start at the 1st of January 1980 and end at the 19th of March 1988, with a total of 3000 observations. The returns of this time series are defined as $y_t = \ln(x_t / x_{t-1})$. Based on these returns one can determine the sample variances "s^2" of the returns, indicated as:

$$s^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1} .$$

By keeping the first return $y_1 = \ln(x_1 / x_0)$ fixed one can determine the sample variance of for any time $t = 2,3,4,...3000$. The Ljung-Box-Pierce based on paragraph 2.8 for the volatility of the S&P500 dataset is:

$$Q = 477347.3 > \chi^2_{750,0.05} = 814.8215$$

The corresponding $p-value$ is 2.2e-16 which is less than 0.05. The null hypothesis (H_0) will be rejected indicating that the volatility distribution of the original dataset is not mean reverting.

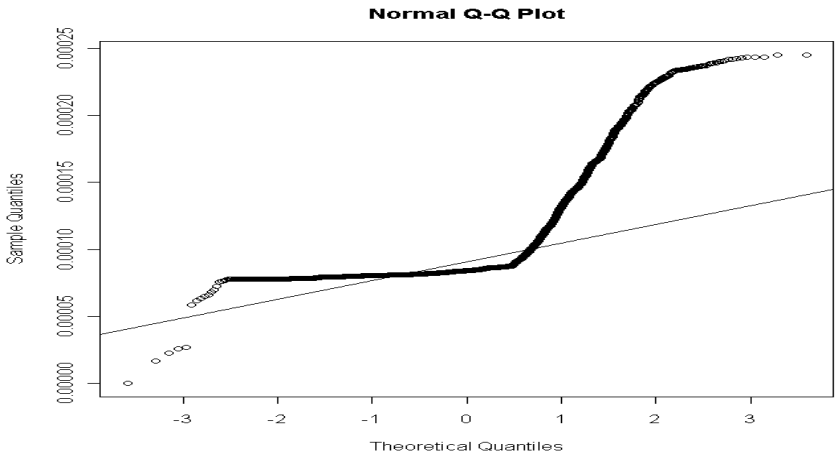

Figure 8.1: The QQ-Plot of the sample variances of S&P500 Returns - $y_t = \ln(x_t / x_{t-1})$

According to the QQ-Plot for the volatility of the S&P500 returns, the tails are fat. The risk properties of the volatilities of the returns of the original dataset is not mean reverting and the tails are fat. These risk properties correspond with Table 1.1 risk properties for the GARCH simulation model for the volatilities of the returns. I will fit different GARCH volatility models on the sample variances, s^2.

Figures 8.2, 8.3 and 8.4 plot the returns, square returns and the absolute returns of the S&P500.

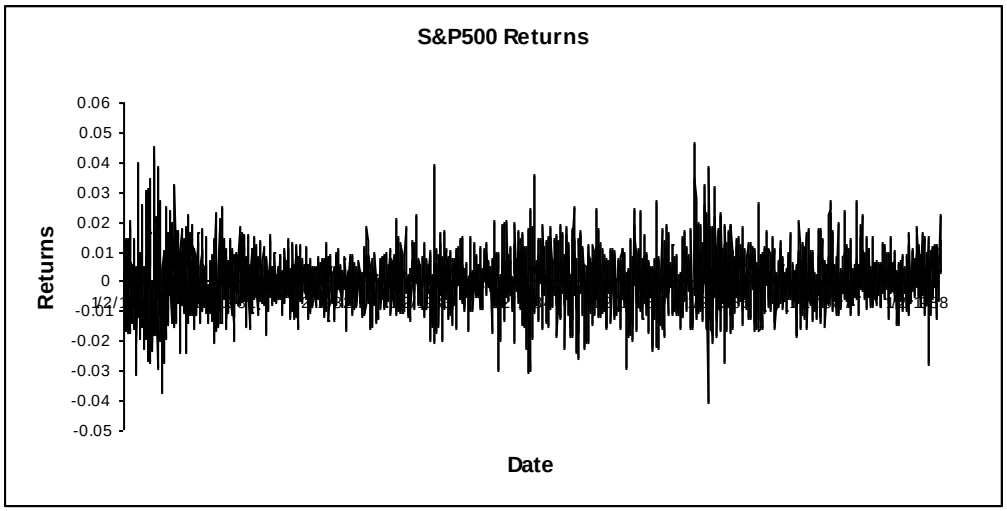

Figure 8.2: Plot of S&P500 Returns - $y_t = \ln(x_t / x_{t-1})$

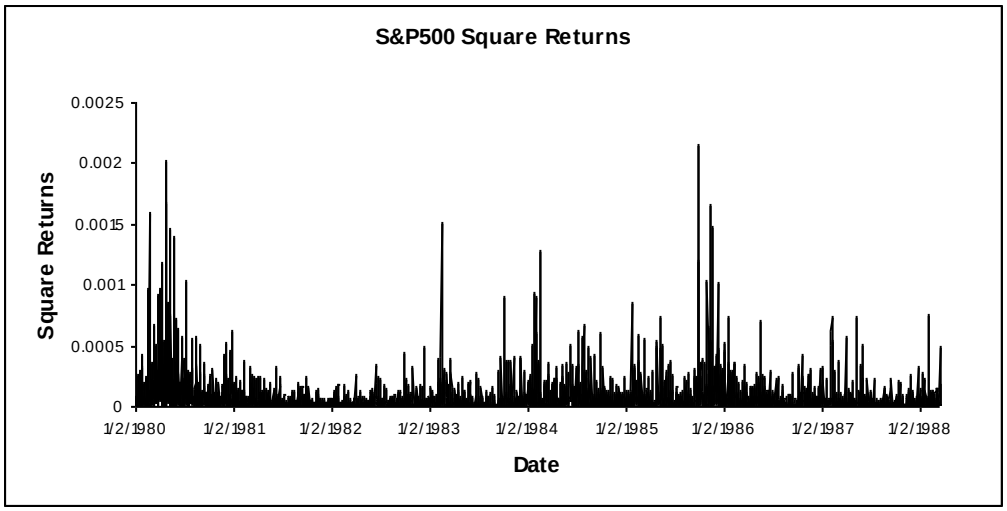

Figure 8.3: Plot of S&P500 Square Returns - $y_t^2 = (\ln(x_t / x_{t-1}))^2$

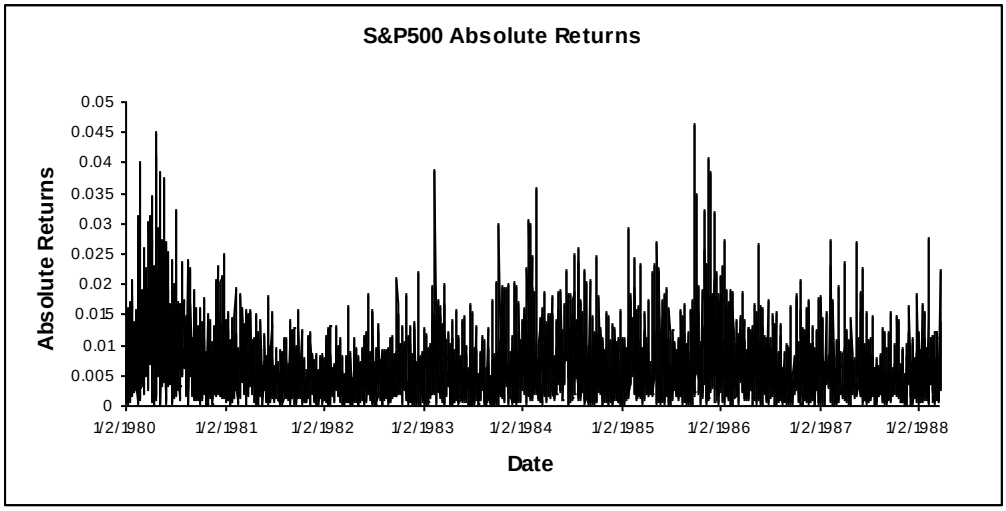

Figure 8.4: Plot of S&P500 Absolute Returns - $|y_t| = |\ln(x_t / x_{t-1})|$

From observing Figure 8.2, there are patterns, although not as clear for the returns of the S&P500, but still distinctive, emerge from the square and absolute returns. This indicates the volatility of returns. The plots of the square and absolute returns indicate volatility clustering. Low values of volatility are followed by low values, while high values of volatility are followed by high values. This observation is very important for forecasting purposes. Patterns usually emerge when serial correlation is present. In the following Figures 8.4, 8.5 and 8.6 the sample autocorrelations for the first 20 lags of the returns, square returns and absolute returns for the S&P500 are displayed.

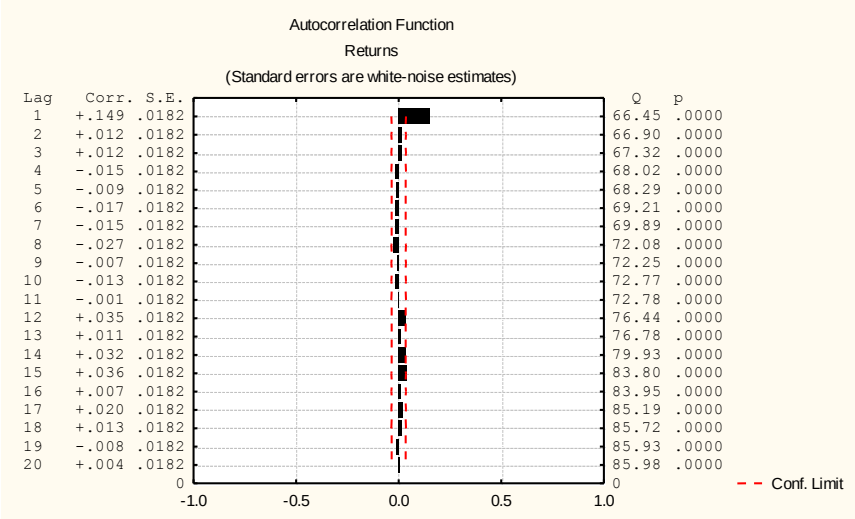

Figure 8.5: Estimated Autocorrelation of $y_t = \ln(x_t / x_{t-1})$

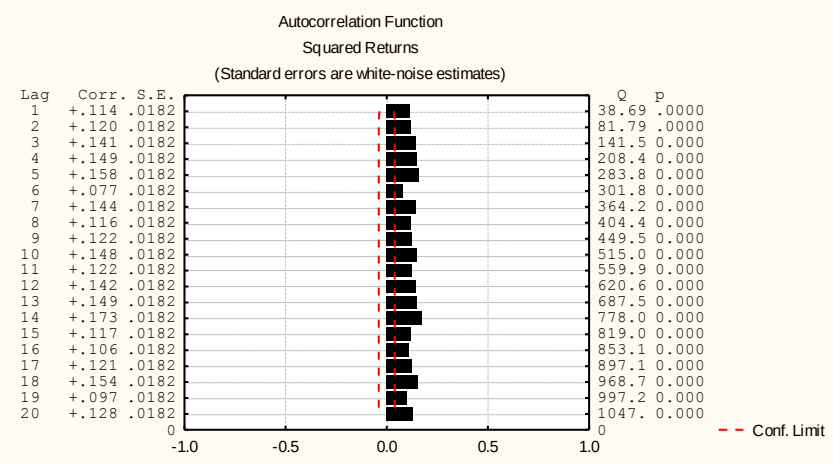

Figure 8.6: Estimated Autocorrelation $y_t^2 = \left(\ln\left(x_t / x_{t-1}\right)\right)^2$

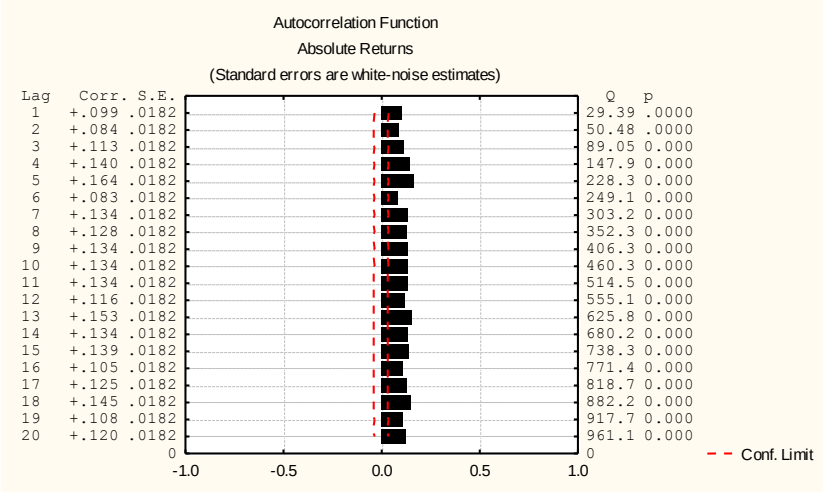

Figure 8.7: Estimated Autocorrelation $\left|y_t\right| = \left|\ln\left(x_t / x_{t-1}\right)\right|$

The sample autocorrelations for the returns, squared returns and absolute returns indicate significance in their first 20 lags according to the Ljung-Box-Pierce-test procedure, explained in chapter 2 of this thesis. The sample autocorrelations for the squared and absolute returns indicate positive serial autocorrelation and their decay are much slower than compared with the sample autocorrelations of the returns.

8.8 Testing the *ARCH/GARCH* Effects

The hypothesis test applicable to the Ljung-Box-Pierce-test statistic, is defined as:

$$Q(q) = n(n-2)\sum_{k=1}^{q} \frac{r_k^2}{(n-k)}$$

where r_k denotes the estimated sample autocorrelation of the squared and absolute returns, which will be used to determine evidence for time varying conditional volatility.

The estimated sample autocorrelation of the squared returns is denoted as:

$$r_k = \frac{\sum_{t=1}^{n-|k|}\left(y_t^2 - \hat{y}^2\right)\left(y_{t+k}^2 - \hat{y}^2\right)}{\sum_{t=1}^{n}\left(y_t^2 - \hat{y}^2\right)^2}$$

where $\hat{y}^2 = \sum_{t=1}^{n}\frac{y_t^2}{n}$. Testing whether the true autocorrelations, ρ_k, of the squared or absolute returns are significant, involve the hypothesis test:

$H_0: \rho_k = 0$, for all $k \le q$ (no evidence of time varying volatility), versus
$H_1: \rho_k \neq 0$, for some value of $k \le q$ (evidence of time varying volatility).

We reject H_0 if $Q(q) > \chi^2_{q;\alpha}$, q, which indicates the degree of freedom, while α represents the level of significance. We also reject H_0 if the $p-value$ is less than 0.05. The *p-value* for the two-tail test, according to the alternative hypothesis, is:

$$p-value = 2P\left(\chi^2 < Q(q)\right) \text{ or,}$$
$$p-value = 2P\left(\chi^2 > Q(q)\right).$$

8.9 Lagrange Multiplier (LM) Test

The LM test is used to determine if any *ARCH* effects are present. The null hypothesis is denoted as:

$H_0: \alpha_0 = \alpha_1 = \cdots = \alpha_q = 0$, (no *ARCH* effects are present), versus
$H_{a1}: \alpha_0 \neq 0$ or $\alpha_1 \neq 0$ or … or $\alpha_q \neq 0$ (*ARCH* effects are present, $ARCH(q+r)$)
or
$H_{a2}: \alpha_0 \neq 0$ or $\alpha_1 \neq 0$ or … or $\alpha_q \neq 0$ (*GARCH* effects are present, $GARCH((q+r),\beta)$).

The test statistic is:

$$LM(q) = nR^2 \sim \chi^2_{q,\alpha}$$

where R^2 is calculated from the regression of $\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + u_t$ mentioned at the end of paragraph 8.1. One will reject H_0 if $LM(q) > \chi^2_{q;\alpha}$. The LM test can also be used for general classification of the *GARCH* models. The LM test can however be rejected, not because *ARCH* effects are present, but because of misspecification of the *ARCH* / *GARCH* model. Therefore caution must be applied.

8.10 Testing *ARCH/GARCH* Effects on S&P500

Table 8.1 illustrates the Ljung-Box-Pierce-test procedure, while Table 8.2 illustrates the Lagrange Multiplier (LM) test for $ARCH(q)$, such that $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$ on 3000 daily return values of the S&P500 dataset.

Table 8.1: $Q(q)$ Using the Estimated Sample Autocorrelation of the Squared Returns

q	$Q(q)$	$\chi^2_{q;\alpha}$	$p-value$
1	38.6863	3.841459	0.000
3	141.4671	7.814728	0.000
5	283.7676	11.0705	0.000
7	364.1553	14.06714	0.000
9	449.451	16.91898	0.000

Some illustrations of the calculations in Table 8.1 are:

$$Q(1)=3000(3000+2)\frac{r_1^2}{(3000-1)}=38.6863>\chi^2_{1;0.05}=3.841459.$$
$$\vdots$$
$$Q(9)=Q(8)+3000(3000+2)\frac{r_9^2}{(3000-9)}=449.451>\chi^2_{9;0.05}=16.91898.$$

The results obtained in Table 8.1 illustrate that the null hypothesis will be rejected, indicating that *ARCH* effects are present for $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$. This is in accordance to the Ljung-Box-Pierce-test procedure. This indicates that the dataset will have to be transformed into $ARCH(q)$ or $GARCH(q,p)$, such that $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$ are used for more accurate forecasting.

Table 8.2: Calculation for $LM(q)$ Before $ARCH(q)$ or $GARCH(q,p)$ Transformation

q	$LM(q)$	$\chi^2_{q;\alpha}$	$p-value$
1	82.55084	3.841459	0.000
3	66.52530	7.814728	0.000
5	62.86956	11.0705	0.000
7	57.57094	14.06714	0.000
9	60.01441	16.91898	0.000

Some illustrations of the calculations in Table 8.2 are:

$$LM(1)=3000(0.027517)=82.55084>\chi^2_{1;0.05}=3.841459$$
$$\vdots$$
$$LM(9)=3000(0.020005)=60.01441>\chi^2_{9;0.05}=16.91898$$

The results obtained in Table 8.2 illustrate that the null hypothesis will be rejected. This indicates that the $ARCH$ effects are present for $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$, according to the Lagrange Multiplier (LM) test. The dataset will then have to be transformed into $ARCH(q)$ or the $GARCH(q,p)$, for $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$, for more accurate forecasting. The results in both Tables 8.1 and 8.2 mean that volatility was successfully removed.

After the $ARCH(q)$ models are applied on the dataset, the Lagrange Multiplier test is used again to determine whether the transformation removed the volatility. The results in Table 8.3 indicate that all the tests will be rejected, indicating that all the $ARCH(q)$ models are adequate in removing the volatility.

Table 8.3: Calculation for $LM(q)$ After $ARCH(q)$ or $GARCH(q,p)$ Transformation

q	$LM(q)$	$\chi^2_{q;\alpha}$	$p-value$
1	0.056904	3.841459	0.8115
3	2.585095	7.814728	0.4601
5	4.063720	11.0705	0.5403
7	6.955938	14.06714	0.4335
9	12.41581	16.91898	0.1909

8.11 Model Selection

The determination for the orders of q and p in the $ARCH(q)$ and $GARCH(q,p)$ models are of great significance. E-Views uses the Akaike info criterion (AIC), Schwarz criterion and the Hannan-Quinn criterion for the model selection. By using the lowest values of the AIC, Schwarz or Hannan-Quinn procedures, one will obtain the best orders for the models. According to E. Zivot (2008)[34], the AIC will choose values of $q \le 2$ and $p \le 2$ for the $GARCH(q,p)$ models. For the $ARCH(q)$ models, the AIC will choose large values of q. Low orders for the $GARCH(q,p)$ models are preferred above larger orders for the $ARCH(q)$ models. This is due to numerical estimation stability and over arching. Table 8.4 provides the AIC, Schwarz criterion, Hannan-Quinn criterion, as well as the logarithmic likelihood for $ARCH(q)$ and $GARCH(q,p)$. The orders used for the $ARCH(q)$ model are $q=1,\ q=3,\ q=5,\ q=7$ and $q=9$, while $1 \le q \le 2$ and $1 \le p \le 2$ is used for the $GARCH(q,p)$ model.

[34] E Zivot "Practical Issues in the Analysis of Univariate GARCH Models" (2008) 15 *Handbook of Financial Time Series* http://faculty.washington.edu/ezivot/research/practicalgarchfinal.pdf. (Accessed 09 September 2009).

Table 8.4: Model Selection Test

(q, p)	AIC	Schwarz	Hannan	Log Likelihood
(1,0)	-6.61796	-6.61396	-6.61652	9928.939
(3,0)	-6.64338	-6.63538	-6.64050	9969.075
(5,0)	-6.67713	-6.66512	-6.67281	10021.690
(7,0)	-6.68499	-6.66897	-6.67923	10035.480
(9,0)	-6.69847	-6.67844	-6.69126	10057.700
(1,1)	-6.73013	-6.72413	-6.72797	10098.200
(1,2)	-6.73144	-6.72343	-6.72856	10101.150
(2,1)	-6.72947	-6.72146	-6.72659	10098.200
(2,2)	-6.73194	-6.72193	-6.72834	10102.92000

For a pure $ARCH(q)$ model, the $ARCH(9)$ model will be chosen for all selection criterion, as well as a maximum logarithmic likelihood (at $q = 9$). Due to the possibility of over aching the $ARCH(9)$ model is not the best model to choose. For the $GARCH(q, p)$ model, a $GARCH(2,2)$ will be selected by AIC, a $GARCH(1,1)$ will be selected by Schwartz and a $GARCH(1,2)$ by the Hannan-Quinn test. For many applications it is hard to beat the simple $GARCH(1,1)$.

8.12 Volatility Estimation of *ARCH/GARCH* Models

The estimated volatility models for $ARCH(q)$ (if $q = 1,\ q = 3$) and $GARCH(q, p)$ (for $1 \le q \le 2$ and $1 \le p \le 2$) are given in Table 8.5.

Table 8.5: Volatility Estimation of *ARCH/GARCH* Models

(q, p)	*ARCH / GARCH* Model	Kurtosis
(1,0)	$\hat{\sigma}_t^2 = 0.0000705 + 0.111566\varepsilon_{t-1}^2$	4.781474
(3,0)	$\hat{\sigma}_t^2 = 0.0000564 + 0.0.096435\varepsilon_{t-1}^2 + 0.085644\varepsilon_{t-2}^2 + 0.106815\varepsilon_{t-3}^2$	4.348387
(1,1)	$\hat{\sigma}_t^2 = 0.000000683 + 0.042050\varepsilon_{t-1}^2 + 0.949273\hat{\sigma}_{t-1}^2$	3.712614
(1,2)	$\hat{\sigma}_t^2 = 0.00000116 + 0.074204\varepsilon_{t-1}^2 + 0.040348\hat{\sigma}_{t-1}^2 + 0.871010\hat{\sigma}_{t-2}^2$	3.678737
(2,1)	$\hat{\sigma}_t^2 = 0.00000068 + 0.0428894\varepsilon_{t-1}^2 - 0.000926\varepsilon_{t-2}^2 + 0.949394\hat{\sigma}_{t-1}^2$	3.712253
(2,2)	$\hat{\sigma}_t^2 = 0.00000125 + 0.063448\varepsilon_{t-1}^2 + 0.018710\varepsilon_{t-2}^2 + 0.006581\hat{\sigma}_{t-1}^2 + 0.895692\hat{\sigma}_{t-2}^2$	3.659480

8.13 Notes and Research

Firstly I investigated the risk properties of the volatility of the returns of the S&P500 closing prices. According to paragraph 8.7, there is no mean reversion and fat tails are present in the volatility of the returns. According to Figure 8.1 fat tails are present. These risk properties correspond to the risk properties of the GARCH simulation model for the volatilities indicated in Table 1.1.

All the simulated volatility models will have very large Jarque-Bera values. These results confirm non-normality. The coefficients of the sum of all the models add up to a number less than one. If the sum is close to one, as in $GARCH(q, p)$ for $1 \leq q \leq 2$ and $1 \leq p \leq 2$, it indicates a slow mean reversion for the simulated volatility models. On the other hand, if it is much less than one, as in the $ARCH(1)$ and the $ARCH(3)$ models, it indicates a much faster mean reversion but no significant mean reversion is present on the these simulated volatility models.

Due to the fact that the kurtosis of all the simulated volatility models in Table 8.5 exceeds 3 these models will have fat tails. The simulated volatility models have the same risk properties compared to the risk properties of the original sample volatilities of the returns.

Chapter 9

9 GENERAL CONCLUSION

I. Main Purpose of Script

The main purpose of this script is to apply techniques for determining the risk properties of the original dataset. After these risk properties are identified one may choose an appropriate stochastic model in Table 1.1 to simulate the original dataset. One need to calculate the estimated parameters for the simulation models and *GARCH* simulated volatility models. Risk properties should then be similar in the simulated models compared to the risk properties of the origin dataset. In chapter 2 a MLE procedure was used to determine the optimal parameters on the original dataset (FTSE100) and then applied for the simulated GBM.

In chapter 3 the conditional sum of squares were minimised to find the optimal parameter we used in the simulated *AR* model.

In chapter 4 the optimal parameters were calculated by using the OLS and MLE procedures on the original dataset (Sweden's Electricity Prices). According to S.M. Sheikh (2007)[11], the OLS and MLE for his electricity prices were approximately the same. I found these estimated parameters were not the same for my annual electricity price dataset. Due to the fact that fat tails in the QQ-plot for Sweden's Electricity prices were found the MLE procedure will be more accurate compared to the OLS. I found that the greater the observation deviates from normality indicated as fat tails in the QQ-plot, the larger the difference between the parameters for the OLS and MLE procedure. The reason for this difference is due to the fact that the Ordinary Least Squares method relies on the structure of the dataset to be approximately linear. The Ordinary Least Squares method is useful to determine the initial values for the parameters in iteration of optimisation.

In chapter 6 I only determined the estimated parameters of the two-factor Vasicek model.

For chapter 8, I used a very large S&P500 dataset of 4382, but because of extreme fluctuation in the data at some point in time, I could not manage to sufficiently get rid of the volatility by using the *ARCH/GARCH* models. For this reason I chose a smaller S&P500 dataset of 3000 observation, which worked better. Due to the fact that the kurtosis of all the volatility models in Table 8.5 exceeds 3 these models will have fat tails.

[11] SM Sheikh "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed 9 September 2009).

II. Examples Used in the Thesis

This thesis contain examples such as FTSE100 monthly closing prices, the daily electricity prices for Sweden [in Euros per megawatt], daily average number of truck manufacturing defects, S&P500 closing index's, the 30-day AA financial interest rates and the daily market yield on U.S. Treasury security rates at 6-month constant maturity, and 10-year constant maturity.

III. Programs Used in the Thesis

I used a Microsoft Excel Macro to calculate the estimated sample autocorrelation. I initially tried to determine the parameters for the CIR model by using Microsoft Excel, but estimation failure occurred and thus I successfully determined these parameters by means of the scaled version of the Bessel function in MATLAB. I also used Statistica for the determination of the sample autocorrelation functions and sample partial autocorrelation functions. The adjusted chi-squared distribution in Statistica was used to obtain Figure 1.8. Statistica and R version 2.8.1 were used to determine the Ljung-Box-Pierce tests as well as the QQ-plots. The Microsoft Excel solver add-inn was used to determine the maximum for the likelihood function, as well as the minimum of the least squares of the errors, by changing the parameters. E-views was used to determine the parameter for the *ARCH/GARCH* volatility models, as well as the Lagrange Multiplier Test and kurtosis of these models.

Appendix A: Selected Datasets

A.1: FTSE100 Monthly Closing Prices (Read across).

1138.3	1016.6	1039.2	1009.4	1103.3	1139.3	1152.1	1181.1	1231.2	1280.2	1260.8	1277
1291	1313	1234.9	1261.7	1341.1	1290	1377.2	1439.1	1412.6	1435	1543.9	1668.8
1660.5	1602.8	1649.8	1558.1	1661.2	1555.8	1632.2	1636.6	1679	1808.2	1979.2	1997.5
2050.5	2203	2284.1	2360.9	2249.7	2366	1749.8	1579.9	1712.7	1790.8	1768.8	1742.5
1802.2	1784.4	1857.6	1853.6	1753.6	1826.5	1852.4	1792.4	1793.1	2052.1	2002.4	2075
2118	2114.4	2151	2297	2387.9	2299.4	2142.6	2276.8	2422.7	2337.3	2255.4	2247.9
2103.4	2345.1	2374.6	2326.2	2162.8	1990.2	2050.3	2149.4	2143.5	2170.3	2380.9	2456.5
2486.2	2499.5	2414.8	2588.8	2645.7	2621.7	2566	2420.2	2493.1	2571.2	2562.1	2440.1
2654.1	2707.6	2521.2	2399.6	2312.6	2553	2658.3	2778.8	2846.5	2807.2	2868	2878.7
2813.1	2840.7	2900	2926.5	3100	3037.5	3171	3166.9	3418.4	3491.8	3328.1	3086.4
3125.3	2970.5	2919.2	3082.6	3251.3	3026.3	3097.4	3081.4	3065.5	2991.6	3009.3	3137.9
3216.7	3319.4	3314.6	3463.3	3477.8	3508.2	3529.1	3664.3	3689.3	3759.3	3727.6	3699.7
3817.9	3747.8	3711	3703.2	3867.6	3953.7	3979.1	4058	4118.5	4275.8	4308.3	4312.9
4436	4621.3	4604.6	4907.5	4817.5	5244.2	4842.3	4831.8	5135.5	5458.5	5767.3	5932.2
5928.3	5870.7	5832.5	5837	5249.4	5064.4	5438.4	5743.9	5882.6	5896	6175.1	6295.3
6552.2	6226.2	6318.5	6231.9	6246.4	6029.8	6255.7	6597.2	6930.2	6268.5	6232.6	6540.2
6327.4	6359.3	6312.7	6365.3	6672.7	6294.2	6438.4	6142.2	6222.5	6297.5		

A.2: Daily Average Number of Truck Manufacturing Defects (Read across).

1.2	1.5	1.54	2.7	1.95	2.4	3.44	2.83	1.76	2	2.09	1.89	1.8
1.25	1.58	2.25	2.5	2.05	1.46	1.54	1.42	1.57	1.4	1.51	1.08	1.27
1.18	1.39	1.42	2.08	1.85	1.82	2.07	2.32	1.23	2.91	1.77	1.61	1.25
1.15	1.37	1.79	1.68	1.78	1.84							

A.3: The Daily Electricity Prices for Sweden in Euros per Megawatt (Read across).

45.4	54.8	52.5	48	47	47.6	57.2	49.1	47.2	48.5	47.1	46.4	46.1	47.4
45.7	45.6	45.6	44.7	41.1	40.9	48.4	49.4	51.2	43.6	41.8	39.8	39.8	44.3
41.7	41.6	40.7	39.3	39.6	39.8	49.3	46.9	41.4	43.2	40	38.3	38.1	40.8
40.7	41.1	47.7	52.8	44.8	38.9	44.6	44	41.3	38.5	34.9	32.8	31	35.3
32.8	29.4	32.3	32.7	28.6	27.7	33.3	36.7	38.1	32.5	36.2	28.4	26.3	29.4
29.7	28.4	27.9	32	26.4	24.8	27.8	29.1	31	30.4	27.6	27.6	29.9	30
44	37.2	52.4	41.8	29.3	26.6	37	39.1	38.5	40.1	39.4	37.6	35.2	42.6
46.4	50.4	49.3	48.6	43.3	38.6	53.5	46.3	47.6	47.7	51.3	44.8	43.2	51.1
47.7	46.9	46.7	45.8	42.4	30.4	37.4	40.2	35.1	14.2	32.7	22.2	9.18	32.7
36.3	35	31.6	28	8.06	13.7	29.4	30.9	35.6	36.7	40.9	43.3	41.2	49.4
53.9	60.1	52.5	58.3	45.7	27.9	48.5	59.9	53.9	54.1	52.4	51	42.8	66.8
76.7	58.9	58.7	52.6	59.5	52.4	67.6	59	58.3	64.5	56.5	49	49.2	74.5
68.1	67.5	63.6	53.7	50.1	46.6	48.5	55.3	56.9	55.8	54.6	54.7	47.8	58.8
62.2	62.6	60.7	58.2	59.3	52.6	60.2	64.3	63.3	62.9	61.3	58	54.9	60.4
61	58.3	61.3	66.3	58	33.8	26.1	52.4	60.8	57.7	57.7	59.6	55.6	47.5
64	64.6	65.9	65.8	65.1	57.1	47.7	62.8	57	63.5	61.2	60.4	59.8	55.8
68.1	71.5	71.6	69.6	67.4	56.9	49.4	70.5	69.5	67.9	74.6	74.4	66.5	58.8
68.2	70.5	70.7	71	75.6	72.3	68.8	76.4	71.1	71.4	71.1	70.3	68.1	62.1
69.3	71.3	71.7	71.3	70.2	66.2	64.6	78.4	76.6	75	81.5	77.6	73	68.2
84	77.2	79.2	87.4	79.2	68.1	66.5	81.1	77.3	71.8	72.2	71.8	66.1	62.4
74.5	73.2	73.2	71.7	65.2	58.9	57	59.8	58.8	59.1	57.4	56.5	54.9	51.6
53.2	53.9	59.9	57.9	54.5	51.5	43.2	53.2	56.9	57.3	60	59.1	53.6	52.6
55.3	57.4	56.9	55.7	56.4	55.8	49.6	51.2	52.2	53.8	55.2	54.4	48.9	48.6
55.8	52.6	52	51.8	54.8	48.1	48.3	63.3	60.5	50.6	48.3	47.6	48.6	47.3
54.9	49.1	49.9	48.6	49.2	44.4	43.8	46.8	54.4	54	52.9	48.1	44.8	42.8
49.5	45.6	43.5	42.1	40.3	37.6	38.6	41.2	41.6	38.6	37.7	38.8	41.5	41.3
43.6	42.9	41.8											

Appendix B: Programs for Implementing Estimation

B.1: The MATLAB program used to determine the MLE for the CIR Model using *besseli* command.

```
%-------------------------------------------------------------------------------------------------------------------------%
%        Main Program: CIR.m
%-------------------------------------------------------------------------------------------------------------------------%

load C.txt –ascii                                          % Load the text file which include 30-Day AA interest rates

Int_Rates = C;
dt=1/360;                                                  % The Time Step delta t
params = [0.8 .2 .9];                                      % Set initial values for parameters
format long                                                % Set precision of parameters to 16 decimal places

CIR = CIR_calibr_b(Int_Rates, dt)                          % Calling Function CIR_calibr_b

%-------------------------------------------------------------------------------------------------------------------------%
%        Function Called: CIR_calibr_b
%-------------------------------------------------------------------------------------------------------------------------%

function CIR = CIR_calibr_b(Int_Rates, dt)

%-------------------------------------------------------------------------------------------------------------------------%
%        Ordinary Least Square Estimation Procedure
%-------------------------------------------------------------------------------------------------------------------------%

   N=length(Int_Rates);                                    % Find the length of interest rate
   x = [ones(N-1,1)  Int_Rates(1:N-1)];                    % Construct (N-1 X 2)-Matrix
   ols = (x'*x)^(-1)*(x'* Int_Rates(2:N));                 % OLS calculation
    m = mean(Int_Rates);                                   % Determine the average of interest rate

   v = var(Int_Rates);                                     % Determine the variance of interest rate

   params = [-log(ols(2))/dt,m,sqrt(2*ols(2)*v/m)];        % OLS calculation

%-------------------------------------------------------------------------------------------------------------------------%
%        Main Program For Estimators
%-------------------------------------------------------------------------------------------------------------------------%
options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
CIR = fminsearch(@Function_CIR, params, options);          % Find optimal  parameter
maxL = Function_CIR(params)                                % Set Likelihood function as a function

 function maxL = Function_CIR(params)                      %Assign parameters to function

                    alpha = params(1);                     % Set initial value, 0.8, to parameter 1
                    theta = params(2);                     % Set initial value, 0.2, to parameter 2
                    sigma = params(3);                     % Set initial value, 0.9, to parameter 3
                    c = (2*alpha)/((sigma^2)*(1-exp(-alpha*dt)));
                    q = ((2*alpha*theta)/(sigma^2))-1;
                    u = c*exp(-alpha*dt)* Int_Rates(1:N-1);
                    v = c* Int_Rates(2:N);
                    maxL = -(N-1)*log(c)+sum(u+v-log(v./u)*q/2-...
                     log(besseli(q,2*sqrt(u.*v),1))-abs(real(2*sqrt(u.*v))));
 end
end
```

B.2: The MATLAB program used to determine the MLE for the CIR model using *ncx2pdf* command.

```
%-------------------------------------------------------------------------------------------------------------------%
%        Main Program: CIR2.m
%-------------------------------------------------------------------------------------------------------------------%

load C.txt –ascii                                   % Load the text file which include 30-Day AA interest rates

Int_Rates = C;
dt=1/360;                                           % The Time Step delta t
params = [0.8 .2 .9];
format long                                         Set precision of parameters to 16 decimal places

CIR2 = CIR_calibr_b2(Int_Rates, dt)                 % Calling Function CIR_calibr_b2

%-------------------------------------------------------------------------------------------------------------------%
%        Function Called: CIR_calibr_b2
%-------------------------------------------------------------------------------------------------------------------%

function CIR2 = CIR_calibr_b2(Int_Rates, dt)

%-------------------------------------------------------------------------------------------------------------------%
%        Ordinary Least Square Estimation Procedure
%-------------------------------------------------------------------------------------------------------------------%

   N=length(Int_Rates);
   x = [ones(N-1,1)  Int_Rates(1:N-1)];             % Construct a (N-1 X 2)-Matrix
   ols = (x'*x)^(-1)*(x'* Int_Rates(2:N));
   m = mean(Int_Rates);
   v = var(Int_Rates);
   params = [-log(ols(2))/dt,m,sqrt(2*ols(2)*v/m)];
%-------------------------------------------------------------------------------------------------------------------%
%        Main Program For Estimators
%-------------------------------------------------------------------------------------------------------------------%
options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
CIR2 = fminsearch(@Function_CIR2, params, options);
maxL = Function_CIR2(params)

   function maxL2 = Function_CIR2(params)
   alpha = params(1);                               % Set initial value, 0.8, to parameter 1
   theta = params(2);                               % Set initial value, 0.2, to parameter 2
   sigma = params(3);                               % Set initial value, 0.9, to parameter 3
   c = (2*alpha)/((sigma^2)*(1-exp(-alpha*dt)));
   q = ((2*alpha*theta)/(sigma^2))-1;
   u = c*exp(-alpha*dt)* Int_Rates(1:N-1);
   v = c* Int_Rates(2:N);
   s = 2*c* Int_Rates(2:N);
   nc = 2*u;
   df = 2*q+2;
   gpdf = ncx2pdf(s, df, nc);
   ppdf = 2*c*gpdf;
   maxL2 = sum(-log(ppdf));
   end
end
```

B.3: The Microsoft Excel Macro used to determine the ACF.

```
‘-------------------------------------------------------------------------------------------------------------------------------‘
‘          Sub MM() ‘ Subroutine name is MM ‘
‘-------------------------------------------------------------------------------------------------------------------------------‘

Dim colcount(5500) As Integer                    ‘Define Integer variable for column count
Dim ggg, Name As Variant
Dim i As Integer, k As Integer
Dim neq As Integer
Dim Sum_data(5500), Average(5500), As Double
Dim Square(5500), Autocorr(5500), As Double

‘Define Sum_data, Average, Square, Autocorr as Double vectors

Dim Sum_Square(5500), r As Double
Dim Max(5500), Min(5500) As Double

Name = "Total"

ggg = "notempty"
k = 4
While ggg <> ""
  ggg = Worksheets("DATA").Cells(k, 3).Value
  k = k + 1
Wend

rowcount = k - 5                                 'Amount of last row
n = rowcount

For i = 1 To rowcount                            'Assign t values to
Cells(3 + i, 2) = i

Next i

For k = 0 To rowcount                            'Assign k values to
Cells(3 + k, 1) = k

Next k

Average(1) = 0
Sum_data(1) = 0
Sum_Square(1) = 0

‘-------------------------------------------------------------------------------------------------------------------------------‘
'Determine Average
‘-------------------------------------------------------------------------------------------------------------------------------‘

For k = 1 To rowcount

Sum_data(i) = Sum_data(i) + Worksheets("DATA").Cells(k + 3, 3).Value

Next k

Worksheets("DATA").Cells(rowcount + 6, 8).Value = Sum_data(i)
```

```
Worksheets("DATA").Cells(rowcount + 7, 8).Value = Sum_data(i) / rowcount
Average(i) = Sum_data(i) / rowcount

‘-----------------------------------------------------------------------------------------------------------------------------------‘
'Assign Names to Cell in Excel Spreadsheet
‘-----------------------------------------------------------------------------------------------------------------------------------‘

Worksheets("DATA").Cells(rowcount + 6, 7).Value = "Sum"
Worksheets("DATA").Cells(rowcount + 7, 7).Value = "Average"

‘-----------------------------------------------------------------------------------------------------------------------------------‘
'Determine Variance and Average
‘-----------------------------------------------------------------------------------------------------------------------------------‘

Sum_Square(1) = 0

For k = 1 To rowcount
Sum_Square(i) = Sum_Square(i) + (Worksheets("DATA").Cells(k + 3, 3).Value - Average(i)) ^ 2
Next k

Variance = Sum_Square(i) / (rowcount)
Worksheets("DATA").Cells(3, 5).Value = Variance

Average(i) = Sum_data(i) / rowcount

‘-----------------------------------------------------------------------------------------------------------------------------------‘
'Determine Autocovariance and Autocorrelation
‘-----------------------------------------------------------------------------------------------------------------------------------‘
Autocorr(lagk) = 0

For lagk = 1 To rowcount - 1

For k = 1 To rowcount - lagk

Autocorr(lagk) = (Worksheets("DATA").Cells(k + 3, 3).Value - Average(i)) *
(Worksheets("DATA").Cells(k + 3 + lagk, 3) - Average(i)) + Autocorr(lagk)

Next k

' Autocorrealtion
Worksheets("DATA").Cells(3 + lagk, 4).Value = Autocorr(lagk) / Sum_Square(i)

' Autocovariance
Worksheets("DATA").Cells(3 + lagk, 5).Value = (Autocorr(lagk) / Sum_Square(i)) * Variance

‘-----------------------------------------------------------------------------------------------------------------------------------‘
'Assign a Name to Cell in Excel Spreadsheet
‘-----------------------------------------------------------------------------------------------------------------------------------‘

Worksheets("DATA").Cells(2, 4).Value = "Autocorrelation"

Next lagk

End Sub
```

B.4: The MATLAB program used to determine the MLE for the one-factor logarithmic Vasicek Model using (4.42).

```
%-------------------------------------------------------------------------------------------------------------------%
%        Main Program: vas.m
%-------------------------------------------------------------------------------------------------------------------%

load swe.txt -ascii
load time.txt -ascii
V_data=swe; %x1 - x365
t=time; % 1 - 365

ML_log_vasicek_params = log_vasicek(V_data, t)   % Calling Function
                                                 log_vasicek

%-------------------------------------------------------------------------------------------------------------------%
%        Function Called: log_vasicek
%-------------------------------------------------------------------------------------------------------------------%

function ML_log_vasicek_params = log_vasicek(V_data, t)

n=length(V_data);                                % 365 time points
%params = [65.31102 3.98896 12.92752];           % OLS already determined
params = [65.31102 3.98896];

options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
ML_log_vasicek_params = fminsearch(@FT_VAS_LL_ExactFull, params, options);

function mll = FT_VAS_LL_ExactFull(params)

alpha = params(1);
mu = params(2);
%sigma_square = params(3);
sigma_square = 12.92752;

xo=3.815732;

v= xo*exp(-alpha*t)+(mu-(sigma_square)/(2*alpha))*(1-exp(-alpha*t));
w_2 = ((sigma_square)/(2*alpha))*(1-exp(-2*alpha*t));

mll = (n/2)*log(2*pi)+sum(log(w_2/2)+(1/(2*w_2))*(V_data-v).^2);

end
end
```

References

TEXTBOOKS

[1] Enders, W *Applied Econometric Time Series* 2 ed (2004) John Wiley and Sons, Incorporated: United States of America.

[2] Haberman, R *Applied Partial Differential Equations: with Fourier Series and Boundary Value Problem* 4 ed (2004) Pearson Prentice Hall: New Jersey.

[3] Hull, JC *Options Futures & Other Derivatives* 4 ed (2000) Prentice-Hall International, Incorporated: United States of America.

[4] Luenberger, DG *Investment Science* (1998) Oxford University Press: New York.

[5] Milton, JS and Arnold, JC *Introduction to Probability and Statistics: Principles and Applications for Engineers and the Computing Sciences* 3 ed (1995) McGraw-Hill, Incorporated: New York.

[6] Øksendal, B *Stochastic Differential Equation: An Introduction with Applications* 5 ed (2000) Springer-Verlag: New York.

[7] Rice, JA *Mathematical Statistics and Data Analysis* (1988) Duxbury Press: California.

[8] Walpole, RE *Introduction to Statistics* 3 ed (1990) Macmillan Publishing Company, Incorporated: New York.

[9] Wei, WWS *Time Series Analysis: Univariate and Multivariate Methods* (1990) Addison-Wesley Publishing Company: United States of America.

[10] Brzezniak, Z and Zastawniak, T, *Basic Stochastic Processes* (1999) Springer-Verlag London Limited.

JOURNAL ARTICLES

[1] Engle, RF "The Use of ARCH/GARCH Models in Applied Econometrics" (2006) 15 *Journal of Economic Perspectives* 157.

[2] Hansen, P and Lunde, A "A Forecast Comparison of Volatility Models: Does Anything Beat a *GARCH*(1,1) Model?" (2004) 20 *Journal of Applied Econometrics* 873.

[3] Hurn, AS, Lindsay, KA and Martin, VL "On the Efficacy of Simulated Maximum Likelihood for Estimating the Parameters of Stochastic Differential Equations" (1999) 24 *Journal of Time Series Analysis* 1.

[4] Kung, JJ "Multi-period asset Allocation by Stochastic Dynamic Programming" (2008) 199 *Applied Mathematics and Computation* 341.

[5] Lari-Lavassani, A, Sadeghi, AA and Ware, A "Mean reverting models for energy option pricing" (2001) *University of Calgary*.

[6] Ljung, GM and Box, GEP "The likelihood function of stationary autoregressive-moving average models" (1979) 66 *Biometrika* 265.

[7] Skantze, P, Ilic, M and Chapman J "Stochastic Modeling of Electric Power Prices in a Multi-market Environment" (2000) *IEEE (Cambridge University)* 1109.

[8] Zeytun, S and Gupta, A "A Comparative Study of the Vasicek and the CIR Model of the Short Rate" (2007) 124 *ITWM* 1.

INTERNET SOURCES

[1] Bjerksund, P, Stensland, G and Vagstad F (2008) "Gas Storage Valuation: Price Modelling v. Optimization Methods" *Journal of Econometric Literature* http://www.nhh.no/Admin/Public/Download.aspx?file=/Files/Filer/institutter/for/dp/2008/2008.pdf (Accessed 09 September 2009).

[2] Brigo, D, Dalessandro A, Neugebauer M and Triki F "A Stochastic Processes Toolkit for Risk Management" (2007) *SSRN* http://ssrn.com/abstract=1109160 (Accessed 09 September 2009).

[3] Cizck, P "*ARCH* and *GARCH* Models" (2005) http://mars.wiwi.hu-berlin.de/tutorials/sfehtm1node66.html (Accessed 25 February 2009).

[4] Franco JCG "Maximum Likelihood Estimation of Mean Reverting Processes" (2008) http://www.investmentscience.com/Content/howtoArticles/MLE_for_ORmean_reverting.pdf (Accessed 09 September 2009).

[5] Herrala, N "Vasicek Interest Rate Model: Parameter Estimation, Evolution of the Short-Term Interest Rate and Term Structure" https://oa.doria.fi/bitstream/handle/10024/43257/nbnfi-fe200901141021.pdf?sequence=3 (Accessed 9 September 2009)

[6] Investopedia "Advanced Bond Concepts: Bond Pricing" http://www.investopedia.com/university/advancedbond/advancedbond2.asp?viewed=1 (Accessed 30 June 2009).

[7] Kladivko, K "Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Process: The Matlab Implementation" http://dsp.vscht.cz/konference_matlab/MATLAB07/prispevky/kladivko_k/kladivko_k.pdf (Accessed 09 September 2009).

[8] Mathworks http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/ref/besseli.html&http://www.mathworks.com/cgi-bin/texis/webinator/search/?db=MSS&prox=page&rorder=750&rprox=750&rdfreq=500&rwfreq=500&rlead=250&sufs=0&order=r&is_summary_on=1&ResultCount=10&query=bessel (Accessed 09 September 2009).

[9] Sheikh, SM. "Modeling Energy Spot Market and Pricing Energy Derivatives: A Technical Analysis" (2007) http://etd.library.pitt.edu/ETD/available/etd-04262007-152450/unrestricted/Final-Thesis.pdf (Accessed, 09 September 2009).

[10] Van den Berg, MA "Calibrating the Ornstein-Uhlenbeck model" (2009) http://www.sitmo.com/doc/Calibrating_the_Omstein-Uhlenbeck_model (Accessed 02 June 2009).

[11] Zivot, E "Practical Issues in the Analysis of Univariate *GARCH* Models" (2008) 15 *Handbook of Financial Time Series* http://faculty.washington.edu/ezivot/research/practicalgarchfinal.pdf (Accessed 09 September 2009).

[12] Carnegie Mellon University "Manual for the Sphinx-III recognition system" Technical version 1" http://www.speech.cs.cmu.edu/sphinxman/HMM.pdf (Accessed 21 September 2010).